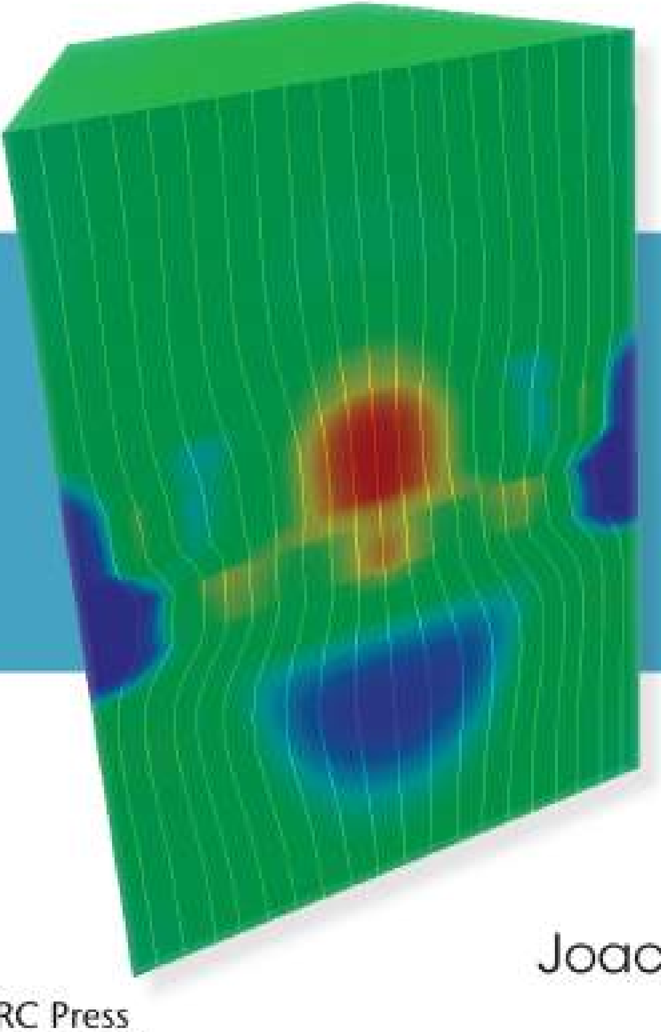
SERIES IN OPTICS AND OPTOELECTRONICS

Handbook of Optoelectronic Device Modeling & Simulation

Fundamentals, Materials, Nanostructures, LEDs, and Amplifiers

Volume I



Edited by

Joachim Piprek

CRC Press
Taylor & Francis Group

Handbook of Optoelectronic Device Modeling and Simulation

Series in Optics and Optoelectronics

Series Editors: **Robert G W Brown**
University of California, Irvine, USA

E Roy Pike
Kings College, London, UK

Recent titles in the series

Optical MEMS, Nanophotonics, and Their Applications
Guangya Zhou and Vincent C. Lee (Eds.)

Optical Microring Resonators: Theory, Techniques, and Applications
Vien Van

Thin-Film Optical Filters, Fifth Edition
H. Angus Macleod

Handbook of Optoelectronics, Second Edition: Concepts, Devices, and Techniques – Volume One
John P. Dakin and Robert Brown (Eds.)

Handbook of Optoelectronics, Second Edition: Enabling Technologies – Volume Two
John P. Dakin and Robert Brown (Eds.)

Handbook of Optoelectronics, Second Edition: Applied Optical Electronics – Volume Three
John P. Dakin and Robert Brown (Eds.)

Handbook of GaN Semiconductor Materials and Devices
Wengang (Wayne) Bi, Hao-chung (Henry) Kuo, Pei-Cheng Ku, and Bo Shen (Eds.)

Handbook of Optoelectronic Device Modeling and Simulation: Fundamentals, Materials, Nanostructures, LEDs, and Amplifiers – Volume One
Joachim Piprek (Ed.)

Handbook of Optoelectronic Device Modeling and Simulation: Lasers, Modulators, Photodetectors, Solar Cells, and Numerical Methods – Volume Two
Joachim Piprek (Ed.)

Nanophotonics and Plasmonics: An Integrated View
Dr. Ching Eng (Jason) Png and Dr. Yuriy Akimov

Handbook of Solid-State Lighting and LEDs
Zhe Chuan Feng (Ed.)

Optical Compressive Imaging
Adrian Stern

Singular Optics
Gregory J. Gbur

Handbook of Optoelectronic Device Modeling and Simulation

Fundamentals, Materials, Nanostructures, LEDs, and Amplifiers

VOLUME ONE

Edited by
Joachim Piprek

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB software.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2018 by Taylor & Francis Group, LLC

CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed on acid-free paper

International Standard Book Number-13: 978-1-4987-4946-6 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Names: Piprek, Joachim, editor.
Title: Handbook of optoelectronic device modeling and simulation / edited by Joachim Piprek.
Other titles: Series in optics and optoelectronics (CRC Press) ; 27.
Description: Boca Raton, FL : CRC Press, Taylor & Francis Group, [2017] |
Series: Series in optics and optoelectronics ; 27 | Includes bibliographical references and index.
Contents: volume 1. Fundamentals, materials, nanostructures, LEDS, and amplifiers – volume 2. Lasers, modulators, photodetectors, solar cells, and numerical methods.
Identifiers: LCCN 2016058063 | ISBN 9781498749466 (v. 1; hardback ; alk. paper) |
ISBN 1498749461 (v. 1; hardback ; alk. paper) | ISBN 9781498749565 (v. 2 ; hardback ; alk. paper) |
ISBN 1498749569 (v. 2 ; hardback ; alk. paper)
Subjects: LCSH: Optoelectronic devices–Mathematical models–Handbooks, manuals, etc. |
Optoelectronic devices–Simulation methods–Handbooks, manuals, etc. |
Semiconductors–Handbooks, manuals, etc. | Nanostructures–Handbooks, manuals, etc.
Classification: LCC TK8304 .H343 2017 | DDC 621.381/045011–dc23
LC record available at https://lccn.loc.gov/2016058063

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Series Preface ix

Preface xi

Editor xiii

Contributors xv

PART I Fundamentals

1 Electronic Band Structure 3
Stefan Schulz and Eoin P. O'Reilly

2 Electron Transport 35
Francesco Bertazzi, Michele Goano, Giovanni Ghione, Alberto Tibaldi, Pierluigi Debernardi, and Enrico Bellotti

3 Electron–Photon Interaction 81
Angela Thränhardt

4 Optical Waveguiding 103
Slawomir Sujecki

5 Heat Generation and Dissipation 129
Giovanni Mascali and Vittorio Romano

6 Process Simulation 167
Simon Z. M. Li, Changsheng Xia, and Yue Fu

PART II Novel Materials

7 Organic Semiconductors 191
Roderick C. I. MacKenzie

8 Polarization in III-N Semiconductors 219
Max A. Migliorato, Joydeep Pal, Xin Huang, Weiguo Hu, Morten Willatzen, and Yousong Gu

9 Dilute Nitride Alloys 251
Christopher A. Broderick, Masoud Seifikar, Eoin P. O'Reilly, and Judy M. Rorison

10 Dilute Bismide Alloys 313
Christopher A. Broderick, Igor P. Marko, Eoin P. O'Reilly, and Stephen J. Sweeney

PART III Nanostructures

11 Quantum Wells 365
Seoung-Hwan Park and Doyeol Ahn

12 Nanowires 397
Oliver Marquardt, Vladimir M. Kaganer, and Pierre Corfdir

13 Quantum Dots 419
Stanko Tomić and Nenad Vukmirović

PART IV Light-Emitting Diodes (LEDs)

14 Light-Emitting Diode Fundamentals 451
Sergey Yu. Karpov

15 Organic Light-Emitting Diodes 473
Pascal Kordt, Peter Bobbert, Reinder Coehoorn, Falk May, Christian Lennartz, and Denis Andrienko

16 Tunnel-Junction Light-Emitting Diodes 523
Yen-Kuang Kuo, Jih-Yuan Chang, Ya-Hsuan Shih, Fang-Ming Chen, and Miao-Chan Tsai

17 Quantum Disk Nanowire Light-Emitting Diodes 541
Fabio Sacconi

18 Influence of Random InGaN Alloy Fluctuations on GaN-Based Light-Emitting Diodes 559
Chen-Kuo Wu, Tsung-Jui Yang, and Yuh-Renn Wu

19 Superluminescent Light-Emitting Diodes 589
Nicolai Matuschek and Marcus Duelk

PART V Semiconductor Optical Amplifiers (SOAs)

20 Semiconductor Optical Amplifier Fundamentals 611
Michael J. Connelly

21 Traveling-Wave and Reflective Semiconductor Optical Amplifiers 631
Angelina Totović and Dejan Gvozdić

22 Tapered Semiconductor Optical Amplifiers 697
José-Manuel G. Tijero, Antonio Pérez-Serrano, Gonzalo del Pozo, and Ignacio Esquivias

23 Quantum-Dot Semiconductor Optical Amplifiers 715
Benjamin Lingnau and Kathy Lüdge

24 Wave Mixing Effects in Semiconductor Optical Amplifiers 747
Simeon N. Kaunga-Nyirenda, Michal Dlubek, Jun Jun Lim, Steve Bull, Andrew Phillips, Slawomir Sujecki, and Eric Larkins

25 Semiconductor Optical Amplifier Dynamics and Pattern Effects 771
Zoe V. Rizou and Kyriakos E. Zoiros

Index 797

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Series Preface

This international series covers all aspects of theoretical and applied optics and optoelectronics. Active since 1986, eminent authors have long been choosing to publish with this series, and it is now established as a premier forum for high-impact monographs and textbooks. The editors are proud of the breadth and depth showcased by the published works, with levels ranging from advanced undergraduate and graduate student texts to professional references. Topics addressed are both cutting edge and fundamental, basic science and applications-oriented, on subject matter that includes lasers, photonic devices, nonlinear optics, interferometry, waves, crystals, optical materials, biomedical optics, optical tweezers, optical metrology, solid-state lighting, nanophotonics, and silicon photonics. Readers of the series are students, scientists, and engineers working in optics, optoelectronics, and related fields in the industry.

Proposals for new volumes in the series may be directed to Lu Han, senior publishing editor at CRC Press/Taylor & Francis Group (lu.han@taylorandfrancis.com).

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Preface

Optoelectronic devices have become ubiquitous in our daily lives. For example, light-emitting diodes (LEDs) are used in almost all household appliances, in traffic and streetlights, and in full-color displays. Laser diodes, optical modulators, and photodetectors are key components of the Internet. Solar cells are core elements of energy supply systems. Optoelectronic devices are typically based on nanoscale semiconductor structures that utilize the interaction of electrons and photons. The underlying and highly complex physical processes require mathematical models and numerical simulation for device design, analysis, and performance optimization. This handbook gives an introduction to modern optoelectronic devices, models, and simulation methods.

Driven by the expanding diversity of available and envisioned practical applications, mathematical models and numerical simulation software for optoelectronic devices have experienced a rapid development in recent years. In the past, advanced modeling and simulation was the domain of a few specialists using proprietary software in computational research groups. The increasing user-friendliness of commercial software now also opens the door for nontheoreticians and experimentalists to perform sophisticated modeling and simulation tasks. However, the ever-growing variety and complexity of devices, materials, physical mechanisms, theoretical models, and numerical techniques make it often difficult to identify the best approach to a given project or problem. This book presents an up-to-date review of optoelectronic device models and numerical techniques. The handbook format is ideal for beginners but also gives experienced researchers an opportunity to renew and broaden their knowledge in this expanding field.

Semiconductors are the key material of optoelectronic devices, as they enable propagation and interaction of electrons and photons. The handbook starts with an overview of fundamental semiconductor device models, which apply to almost all device types, followed by sections on novel materials and nanostructures. The main part of the handbook is ordered by device type (LED, amplifier, laser diode, photodetector, and solar cell). For each device type, an introductory chapter is followed by chapters on specialized device designs and applications, describing characteristic effects and models. Finally, novel device concepts and applications are reviewed. At the end of the handbook, an overview of numerical techniques is provided, both for electronic and photonic simulations.

I would like to thank the publisher for initiating this important handbook project and for giving me the opportunity to serve as editor. Many years of organizing the annual international conference on *Numerical Simulation of Optoelectronic Devices (NUSOD)* enabled me to attract a large number of experts from all over the world to write handbook chapters on their research area. I sincerely thank all authors for their valuable contributions.

Joachim Piprek
Newark, Delaware, USA

MATLAB® is a registered trademark of The MathWorks, Inc. For product information, please contact:

The MathWorks, Inc.
3 Apple Hill Drive
Natick, MA 01760-2098 USA
Tel: 508-647-7000
Fax: 508-647-7001
Email: info@mathworks.com
Web: www.mathworks.com

Editor

Joachim Piprek received his diploma and doctoral degrees in physics from the Humboldt University in Berlin, Germany. For more than two decades, he worked in industry and academia on modeling, simulation, and analysis of various semiconductor devices used in optoelectronics. Currently, he serves as president of the NUSOD Institute, Newark, Delaware (see http://www.nusod.org). During his previous career in higher education, Dr. Piprek taught various graduate courses at universities in Germany, Sweden, and the United States. Since 2001, he has been organizing the annual international conference on *Numerical Simulation of Optoelectronic Devices (NUSOD)*. Thus far, Dr. Piprek has published three books, six book chapters, and about 250 research papers, which have received more than 6000 citations. He was an invited guest editor for several journal issues on optoelectronic device simulation and currently serves as an executive/associate editor of two research journals in this field.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Contributors

Doyeol Ahn
Department of Electrical and Computer Engineering
University of Seoul
Seoul, Republic of Korea

Denis Andrienko
Max Planck Institute for Polymer Research
Mainz, Germany

Enrico Bellotti
Photonics Center
Boston University
Boston, Massachusetts

Francesco Bertazzi
Department of Electronics and Telecommunications
Politecnico di Torino
Torino, Italy

and

Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni
Consiglio Nazionale delle Ricerche
Torino, Italy

Peter Bobbert
Department of Applied Physics
Eindhoven University of Technology
Eindhoven, the Netherlands

Christopher A. Broderick
Department of Electrical and Electronic Engineering
University of Bristol
Bristol, United Kingdom

Steve Bull
Department of Electrical and Electronic Engineering
University of Nottingham
Nottingham, United Kingdom

Jih-Yuan Chang
Center for Teacher Education
National Changhua University of Education
Changhua, Taiwan

Fang-Ming Chen
Institute of Photonics
National Changhua University of Education
Changhua, Taiwan

Reinder Coehoorn
Department of Applied Physics
Eindhoven University of Technology
Eindhoven, the Netherlands

Michael J. Connelly
Department of Electronic and Computer Engineering
University of Limerick
Limerick, Ireland

Pierre Corfdir
Paul Drude Institute
Berlin, Germany

Pierluigi Debernardi
Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni
Consiglio Nazionale delle Ricerche
Torino Italy

Gonzalo del Pozo
Center of Advanced Materials and Devices for ICT Applications
Universidad Politécnica de Madrid
Madrid, Spain

Michal Dlubek
ELMAT Group
Podkarpackie, Poland

Marcus Duelk
Exalos AG
Schlieren, Switzerland

Ignacio Esquivias
Department of Photonics Technology and Bioengineering
Center of Advanced Materials and Devices for ICT Applications
Universidad Politécnica de Madrid
Madrid, Spain

Yue Fu
Crosslight Software, Inc.
Vancouver, Canada

Giovanni Ghione
Department of Electronics and Telecommunication
Politecnico di Torino
Torino, Italy

Michele Goano
Department of Electronics and Telecommunications
Politecnico di Torino
Torino, Italy

and

Istituto di Elettronica e di Ingegneria dell'Informazione e delle Telecomunicazioni
Consiglio Nazionale delle Ricerche
Torino, Italy

Yousong Gu
School of Materials Science and Engineering
University of Science and Technology
Beijing, China

Dejan Gvozdić
School of Electrical Engineering
University of Belgrade
Belgrade, Serbia

Weiguo Hu
Beijing Institute of Nanoenergy and Nanosystems
Chinese Academy of Sciences
Beijing, China

Xin Huang
Beijing Institute of Nanoenergy and Nanosystems
Chinese Academy of Sciences
Beijing, China

Vladimir M. Kaganer
Paul Drude Institute
Berlin, Germany

Sergey Yu. Karpov
Device Modelling Department
STR Group – Soft-Impact, Ltd.
Saint-Petersburg, Russia

Simeon N. Kaunga-Nyirenda
Department of Electrical and Electronic Engineering
University of Nottingham
Nottingham, United Kingdom

Pascal Kordt
Max Planck Institute for Polymer Research
Mainz, Germany

Yen-Kuang Kuo
Department of Physics
National Changhua University of Education
Changhua, Taiwan

Eric Larkins
Department of Electrical and Electronic Engineering
University of Nottingham
Nottingham, United Kingdom

Christian Lennartz
TrinamiX GmbH
Ludwigshafen, Germany

Simon Z. M. Li
Crosslight Software, Inc.
Shanghai, China

Jun-Jun Lim
OSRAM Opto Semiconductors
Penang, Malaysia

Benjamin Lingnau
Institute of Theoretical Physics
Technical University Berlin
Berlin, Germany

Kathy Lüdge
Institute of Theoretical Physics
Technical University Berlin
Berlin, Germany

Roderick C. I. MacKenzie
Faculty of Engineering
University of Nottingham
Nottingham, United Kingdom

Igor P. Marko
Advanced Technology Institute and Department of Physics
University of Surrey
Guildford, United Kingdom

Oliver Marquardt
Paul Drude Institute
Berlin, Germany

Giovanni Mascali
Department of Mathematics
University of Calabria and INFN-Gruppo c
Cosenza, Italy

Nicolai Matuschek
Exalos AG
Schlieren, Switzerland

Falk May
BASF SE, GME/MC
Ludwigshafen, Germany

Max A. Migliorato
School of Electrical and Electronic Engineering
University of Manchester
Manchester, United Kingdom

Eoin P. O'Reilly
Photonics Theory Group
Tyndall National Institute
Cork, Ireland

and

Department of Physics
University College Cork
Cork, Ireland

Joydeep Pal
School of Electrical and Electronic Engineering
University of Manchester
Manchester, United Kingdom

Seoung-Hwan Park
Department of Electronics Engineering
Catholic University of Daegu
Kyeongbuk, Republic of Korea

Antonio Pérez-Serrano
Center of Advanced Materials and Devices for ICT Applications
Universidad Politécnica de Madrid
Madrid, Spain

Andrew Phillips
Department of Electrical and Electronic Engineering
University of Nottingham
Nottingham, United Kingdom

Zoe V. Rizou
Department of Electrical and Computer Engineering
Democritus University of Thrace
Xanthi, Greece

Vittorio Romano
Department of Mathematics and Computer Science
University of Catania
Catania, Italy

Judy M. Rorison
Department of Electrical and
Electronic Engineering
University of Bristol
Bristol, United Kingdom

Fabio Sacconi
Tiberlab Srl
Rome, Italy

Stefan Schulz
Photonics Theory Group
Tyndall National Institute
Cork, Ireland

Masoud Seifikar
Integrated Photonics Group
Tyndall National Institute
Cork, Ireland

Ya-Hsuan Shih
Department of Photonics
National Cheng Kung University
Tainan, Taiwan

Slawomir Sujecki
Department of Electrical and Electronic
Engineering
University of Nottingham
Nottingham, United Kingdom

and

Department of Telecommunications and
Teleinformatics
Faculty of Electronics
Wroclaw University of Science and Technology
Wroclaw, Poland

Stephen J. Sweeney
Advanced Technology Institute and Department
of Physics
University of Surrey
Guildford, United Kingdom

Angela Thränhardt
Institute of Physics
Chemnitz University of Technology
Chemnitz, Germany

Alberto Tibaldi
Istituto di Elettronica e
di Ingegneria dell'Informazione e
delle Telecomunicazioni
Consiglio Nazionale delle Ricerche
Torino, Italy

José-Manuel G. Tijero
Center of Advanced Materials and Devices for
ICT Applications
Universidad Politécnica de Madrid
Madrid, Spain

Stanko Tomić
Joule Physics Laboratory
University of Salford
Manchester, United Kingdom

Angelina Totović
School of Electrical Engineering
University of Belgrade
Belgrade, Serbia

Miao-Chan Tsai
Institute of Photonics
National Changhua University of
Education
Changhua, Taiwan

Nenad Vukmirović
Scientific Computing Laboratory
Institute of Physics Belgrade
University of Belgrade
Belgrade, Serbia

Morten Willatzen
Department of Photonics Engineering
Technical University of Denmark
Lyngby, Denmark

Chen-Kuo Wu
Department of Electrical Engineering
National Taiwan University
Taipei, Taiwan

Yuh-Renn Wu
Department of Electrical Engineering
National Taiwan University
Taipei, Taiwan

Changsheng Xia
Crosslight Software, Inc.
Shanghai, China

Tsung-Jui Yang
Department of Electrical Engineering
National Taiwan University
Taipei, Taiwan

Kyriakos E. Zoiros
Department of Electrical and Computer Engineering
Democritus University of Thrace
Xanthi, Greece

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

I

Fundamentals

1 Electronic Band Structure *Stefan Schulz and Eoin P. O'Reilly* 3
Introduction and Overview of Importance of Semiconductor Materials • Bloch's Theorem and Band Structure • Semiconductor Band Structure: Direct and Indirect Gap Materials (GaAs versus Si) • Schrödinger Equation and "Exact" Solution Using LDA–DFT • Beyond LDA: GGA, Hybrid Functional, GW, and Related Approaches • **k·p** and Envelope-Function Approaches • Need for Semiempirical Atomistic Approaches—Tight-Binding and Pseudopotential Methods • Summary

2 Electron Transport *Francesco Bertazzi, Michele Goano, Giovanni Ghione, Alberto Tibaldi, Pierluigi Debernardi, and Enrico Bellotti* 35
Introduction • From the Boltzmann Transport Equation to the Drift-Diffusion Model • Quantum-Corrected Drift-Diffusion • Towards a Genuine Quantum Approach

3 Electron–Photon Interaction *Angela Thränhardt* 81
Introduction • Two-Level System • The Solid-State System • Quantization of the Light Field • Role of the Electromagnetic Environment

4 Optical Waveguiding *Slawomir Sujecki* 103
Introduction to Optical Waveguide Theory • Modes of Optical Waveguides • Beam Propagation Method • Tapered and Bent Waveguides

5 Heat Generation and Dissipation *Giovanni Mascali and Vittorio Romano* 129
Crystal Lattices • Crystal Vibrations and Phonons • Phonon Transport Equation • Boltzmann Equations for Carrier Transport • Macroscopic Balance Equations for Heat Transport in Semiconductors • Energy-Transport Model and Its Numerical Discretization • Thermal Properties of Graphene

6 Process Simulation *Simon Z. M. Li, Changsheng Xia, and Yue Fu* 167
Introduction • MOCVD Reactor Simulation • Implant Models • Dopant Diffusion • Oxidation • Summary

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

1

Electronic Band Structure

Stefan Schulz
and
Eoin P. O′Reilly

1.1 Introduction and Overview of Importance of Semiconductor Materials 3
1.2 Bloch's Theorem and Band Structure 4
Electronic Band Structure
1.3 Semiconductor Band Structure: Direct and Indirect Gap Materials (GaAs versus Si) 7
1.4 Schrödinger Equation and "Exact" Solution Using LDA–DFT 8
Preliminaries • Hohenberg–Kohn Theorem for Nondegenerate Ground States • Kohn–Sham Theory and LDA
1.5 Beyond LDA: GGA, Hybrid Functional, GW, and Related Approaches 13
1.6 **k·p** and Envelope-Function Approaches 17
1.7 Need for Semiempirical Atomistic Approaches—Tight-Binding and Pseudopotential Methods 20
The Empirical Tight-Binding Method • Empirical Pseudopotential Method
1.8 Summary 30

1.1 Introduction and Overview of Importance of Semiconductor Materials

Semiconductors are probably the most technologically important materials in widespread use today. The "information age" in which we now live is entirely dependent on the modern computer, which has at its heart a microprocessor constituted mainly of silicon (Si). Another example is the reading and recording of data onto CDs, DVDs, and Blu-Rays that rely on using light, in the form of laser emission, produced by the semiconductor compounds based on gallium arsenide (GaAs) and gallium nitride (GaN), for example. These advances have undoubtedly had a huge impact on human society and are a testament to the power of scientific research.

These transformational developments are primarily due to both our fundamental understanding of the electronic band structure of semiconductor materials and the precision with which semiconductor structures can now be formed using epitaxial growth techniques—methods which allow a crystal lattice to "grow" layer-by-layer [1,2]. This fundamental understanding combined with advanced experimental techniques has driven worldwide research efforts to engineer the electronic band structure of semiconductor materials, including the use of alloys, heterostructures, quantum confinement, and strain [1,2].

A wide range of different semiconductor materials can now be used in heterostructures, increasing their versatility and enabling tunability of their optical and electrical properties. Overall, a very important class of semiconductor heterostructure, the “quantum well” (QW), is the mainstay of modern optoelectronics [1,2]. QWs consist of a thin layer of one semiconductor compound sandwiched between thicker layers of another. Typical QW thicknesses are very small (only a few nanometers), with their physical properties depending very sensitively on the layer thickness. By confining charge carriers on these length scales, one enters the quantum mechanical regime. This link between quantum mechanics and real-world “observable” properties, for example, emission wavelength, provides a fascinating and important topic not only for basic scientific research but also for designing novel and energy-efficient devices [1,2].

In this chapter, we discuss the fundamental properties of semiconductor materials and review different theoretical methods that provide insight into the electronic structure of these systems to understand and predict their properties. The methods addressed here range from first-principles calculations up to continuum-based models. This chapter is organized as follows. In Section 1.2, we focus on the consequences of the fact that solid-state materials are crystalline with a periodic lattice. This leads to the Bloch theorem and first insights into the basic band structure of solids. Subsequently, in Section 1.3, we turn our attention to the specifics of semiconductor band structures and discuss in detail similarities and differences in the band structures of two of the most often used materials for device applications, namely Si and GaAs. Having discussed the general details and aspects of the electronic structure, we introduce in Section 1.4 the basics of first-principles techniques for calculating the band structure of solid-state materials. This is followed, in Section 1.5, by an overview of current state-of-the-art techniques to address the electronic and optical properties of semiconductor materials. Since these techniques are computationally very expensive, they are typically restricted to structures containing at most several hundreds of atoms. Therefore, to describe the electronic and optical properties of semiconductor heterostructures, (semi)empirical methods are required. We start with $\mathbf{k}\cdot\mathbf{p}$ and envelope-function models in Section 1.6 and turn then to introduce the basic ideas of atomistic approaches such as tight-binding (TB) and empirical pseudopotential theory in Section 1.7. A summary of the chapter content is given in Section 1.8.

1.2 Bloch's Theorem and Band Structure

In a typical solid, around 10^{23} valence electrons contribute to the bonding in each cubic centimeter. Therefore, when one is interested in calculating and studying the electronic structure of a solid, in general, a complex many-body problem has to be solved, since the exact wave function and energy of each electron depends on those of all the others. However, a factor that considerably simplifies the calculation of the energy spectrum is the fact that most solid-state materials are crystalline with a periodic lattice. It can be shown that this symmetry property must also be reflected in the individual electron wave functions, satisfying therefore a condition referred to as *Bloch's theorem.*

We start by considering a solid with a periodic potential $V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r})$, where $\mathbf{R}$ is a *lattice vector*. When studying the energy levels of a single electron in this periodic potential, the system can be described by the one-electron Hamiltonian:

$$H = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}), \tag{1.1}$$

with the kinetic energy part $T = -\frac{\hbar^2}{2m}\nabla^2$ and the potential energy contribution $V(\mathbf{r})$. *Bloch's theorem* states that the corresponding wave functions can be chosen to have the form of a plane wave times a function with the periodicity of the lattice:

$$\psi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r}), \tag{1.2}$$

with

$$u_{n\mathbf{k}}(\mathbf{r} + \mathbf{R}) = u_{n\mathbf{k}}(\mathbf{r}). \tag{1.3}$$

The subscript n denotes the nth state with wave vector $\mathbf{k}$. Combining Equations 1.2 and 1.3, Bloch's theorem can also be written as

$$\psi_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}) = e^{i\mathbf{k}\cdot\mathbf{R}}\psi_{n\mathbf{k}}(\mathbf{r}). \tag{1.4}$$

The proof of Bloch's theorem can be found in many textbooks [3,4]. Instead of proving Bloch's theorem, here we rather make its proof plausible by noting two consequences of Equations 1.2 and 1.4:

1. *Periodic electron density:* In a periodic solid, one expects that the probability density, $|\psi_{n\mathbf{k}}(\mathbf{r})|^2$, normally varies between different points within the unit cell under consideration. Equation 1.2 allows for this since

$$|\psi_{n\mathbf{k}}(\mathbf{r})|^2 = |e^{i\mathbf{k}\cdot\mathbf{r}}|^2|u_{n\mathbf{k}}(r)|^2 = |u_{n\mathbf{k}}(\mathbf{r})|^2. \tag{1.5}$$

The periodic function $u_{n\mathbf{k}}(\mathbf{r})$ does not have to be constant, consequently it can vary within a given unit cell. At a given point $\mathbf{r}$ within one unit cell, the charge density should be equivalent to the charge density at the point $\mathbf{r}+\mathbf{R}$ within another unit cell. This is consistent with Bloch's theorem, since from Equation 1.4 it follows that the probability density $\rho_{n\mathbf{k}}(\mathbf{r}+\mathbf{R})$ can be written as:

$$\rho_{n\mathbf{k}}(\mathbf{r}+\mathbf{R}) = |\psi_{n\mathbf{k}}(\mathbf{r}+\mathbf{R})|^2 = |e^{i\mathbf{k}\cdot\mathbf{R}}|^2|\psi_{n\mathbf{k}}(\mathbf{r})|^2 = \rho_{n\mathbf{k}}(\mathbf{r}). \tag{1.6}$$

This shows that one is left with an equal probability of finding an electron at $\mathbf{r}$ or at $\mathbf{r}+\mathbf{R}$. Obviously, this implies equal charge density at the two points, as one would expect.

2. *Empty lattice model:* In free space ($V(\mathbf{r}) \equiv 0$), the electron wave functions can be chosen to take the form of plane waves. The unnormalized wave function $\psi_k(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}$ describes a state with energy $E = \hbar^2k^2/2m$. If we partition free space into a periodic array of identical boxes of volume L^3, which, which is referred to as the "empty lattice," the free space wave functions can be expressed as a product of a plane wave times a constant function, which is obviously periodic:

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}} \cdot \frac{1}{\sqrt{L^3}}. \tag{1.7}$$

Consequently, Bloch's theorem describes wave functions that reduce to the correct form in the case where the periodic potential $V(\mathbf{r}) \equiv 0$.

Having discussed Bloch's theorem, we consider in the next section its consequences for the electronic structure of a crystalline solid and introduce the concept of the *electronic band structure* of the system under consideration.

1.2.1 Electronic Band Structure

From Bloch's theorem, it follows that in a periodic solid, we can associate a wave vector $\mathbf{k}$ with each energy state, $E_n(\mathbf{k})$. Thus, it is useful to display the energies $E_n(\mathbf{k})$ as a function of the wave vector $\mathbf{k}$. The resulting diagram is referred to as the *electronic band structure* of the given solid. Figure 1.1a shows the electronic band structure for an electron in an empty lattice (free space). As described above, the wave function can be expressed as a plane wave and the resulting $E(\mathbf{k})$ is described by the parabola $E = \hbar^2k^2/2m$.

In a periodic solid, due to the presence of the potential $V(\mathbf{r})$, the free electron band structure is modified in several ways. In addition, the wave vector $\mathbf{k}$ associated with a given energy state is no longer uniquely defined. To illustrate this in more detail, we assume in the following a one-dimensional (1-D) periodic structure. A unit cell of length L shall be considered. In this case, we can write the wave function for the nth state with wave number k as follows:

$$\psi_{nk}(x) = e^{ikx}u_{nk}(x). \tag{1.8}$$

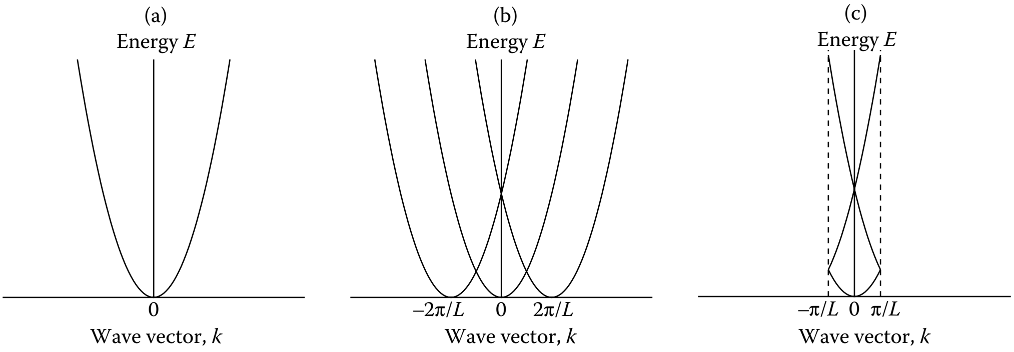


FIGURE 1.1 (a) Free electron "band structure". Here the energy E depends quadratically on the wave vector k. (b) Band structure of a one-dimensional (1-D) lattice with period L. Here wave numbers k and $k + 2\pi m/L$, where m is an integer, are equivalent. In the repeated zone scheme, all wave numbers $k + 2\pi m/L$ associated with each energy state are plotted. (c) In the reduced zone scheme, the wave number for each energy state is chosen in such a way that the magnitude of k is minimized. For the 1-D lattice this implies $-\pi/L < k < \pi/L$.

Here we have a plane wave e^{ikx} of wave number k and $u_{nk}(x)$ is a periodic function, with $u_{nk}(x) = u_{nk}(x + L)$. We wish to show that the wave number k in this case is not uniquely defined. For this we multiply Equation 1.8 by a plane wave with the periodicity of the lattice, $e^{i2\pi mx/L}$, and by its complex conjugate, $e^{-i2\pi mx/L}$, where m is an integer. This yields

$$\begin{aligned}\psi_{nk}(x) &= e^{ikx} e^{\frac{i2\pi m}{L}x} e^{\frac{-i2\pi m}{L}x} u_{nk}(x) \\ &= e^{i(k+2\pi m/L)x} \left(e^{-\frac{i2\pi m}{L}x} u_{nk}(x) \right). \end{aligned} \tag{1.9}$$

Here $e^{i(k+2\pi m/L)x}$ is a plane wave but a different one compared to the original choice. The term $e^{-i2\pi mx/L} u_{nk}(x)$ is still a periodic function with period L. Introducing the notation $G_m = 2\pi m/L$ that we refer to as a *reciprocal lattice vector*, we observe that the wave number k is then equivalent to the wave vector $k + G_m$ in the given 1-D periodic system.

If we now consider dividing 1-D free space into unit cells of length L to create an empty lattice, we have several choices of how to plot the free electron band structure:

- In the *extended zone scheme*, one tries to associate a single, "correct" wave number k with each state, as displayed in Figure 1.1a. For the empty lattice, this is probably the best approach. However, in a periodic crystal it becomes very difficult to assign a unique, "correct" k to each state.
- In the *repeated zone scheme*, for a given energy state, one can include several (in principle, all) wave numbers k associated with this state. In doing so one can circumvent the problem of choosing the "correct" wave number k for each state. This repeated zone scheme is illustrated in Figure 1.1b. However, it is clear that this approach contains a lot of redundant information.
- In the *reduced zone scheme*, for each energy state we select the wave number k in such a way that the magnitude of the wave number k is minimized. In doing so one provides a simple rule for assigning a preferred k value to each state. This gives a simple prescription for displaying the band structure, as illustrated in Figure 1.1c. The reduced zone is also widely referred to as the *first Brillouin zone* for the given crystal structure. Thus, when plotting the electronic band structure of a periodic crystal, we can restrict our choice of wave vectors to those in the first Brillouin zone.

In summary, we see how Bloch's theorem significantly simplifies the calculation of the electronic band structure, since it allows that in a periodic solid, a wave vector **k** can be associated to each energy state, $E_n(\mathbf{k})$. Furthermore, the wave vector can be restricted to the first Brillouin zone. As an example we discussed the "band structure" of a free electron. However, the presence of the potential arising from the crystal lattice has not been addressed. In the next section, we do so by considering as an example the band structure of the two technologically very important semiconductor materials, Si and GaAs. Differences and similarities between the band structures of the two materials will be discussed before coming back in Section 1.4 to the question of how to calculate these band structures.

1.3 Semiconductor Band Structure: Direct and Indirect Gap Materials (GaAs versus Si)

As discussed in Section 1.1, the electronic band structure of semiconductor materials and their specific features are of central importance for determining their electronic and optical properties. This is not only of interest for a fundamental understanding of material properties; it also relates directly to their usefulness for devices such as transistors, lasers, or light-emitting diodes to name only a few. This has resulted in intense research efforts around the world to modify and tailor the electronic band structure of semiconductor materials, including the use of alloys, heterostructures, quantum confinement, and strain. Some of the most advanced devices now employ all of these techniques. Here, we focus our attention on some of the central features of the electronic band structure of bulk semiconductors.

When plotting the band structure of a bulk semiconductor, it is not possible to plot the energy E for each three-dimensional wave vector **k**. Instead, we plot how the energy varies with wave vector **k** along different high-symmetry directions. Different high-symmetry directions and points are given specific names. In particular, the [001] direction between the Γ point ($\mathbf{k} = \mathbf{0}$) and the Brillouin zone edge at the X point is referred to as the Δ direction, while the [111] direction is denoted as the Λ-direction, joins Γ and what is referred to as the L point at the Brillouin zone edge.

Figure 1.2 displays the band structures of bulk GaAs and Si, which are two of the most technologically important semiconductors. For most applications, the bands near the energy gap, E_g, are of particular interest. The energy gap E_g is defined as the energetic separation between the highest filled valence (bonding) band state and the lowest empty conduction (antibonding) band state. From Figure 1.2 one can infer that the lowest energy state in the GaAs conduction band, E_c, is at the Γ-point ($\mathbf{k} = \mathbf{0}$). E_c is directly above the highest energy state in the valence band, E_v. For this reason, GaAs is called a *direct gap semiconductor*. This result is key to the optical properties of GaAs-based systems. If an electron relaxes to the lowest possible state in the conduction band it can recombine directly with a hole at the top of the valence band by emitting a photon—note that on the scale of Figure 1.2 a photon has negligible momentum. To realize efficient light-emitting diodes or laser structures, semiconductors with a direct band gap are utilized [1].

The band structure of GaAs can be contrasted with the band structure of Si (cf. Figure 1.2b). Si is referred to as an *indirect band gap semiconductor*, since the bottom of the conduction band is not directly above the valence band maximum at the Γ point. In Si, the conduction band minima are along the six equivalent Δ (cubic axis) directions near to the X points. For the electron in the lowest conduction band state to recombine with a hole at the top of the valence band to emit a photon, phonon absorption or emission processes are required. This stems from the fact that the momentum must be conserved in such a process. A phonon is the name given to a quantized lattice vibration [3,6]. On the scale of Figure 1.2, phonons have a negligible energy but significant momentum. This originates from the fact that the mass of the nucleus is large compared to the electron mass. The light emission from an indirect gap semiconductor material is therefore less efficient when compared to direct gap materials since it requires the simultaneous occurrence of two processes (phonon absorption/emission plus photon emission/absorption) [1].

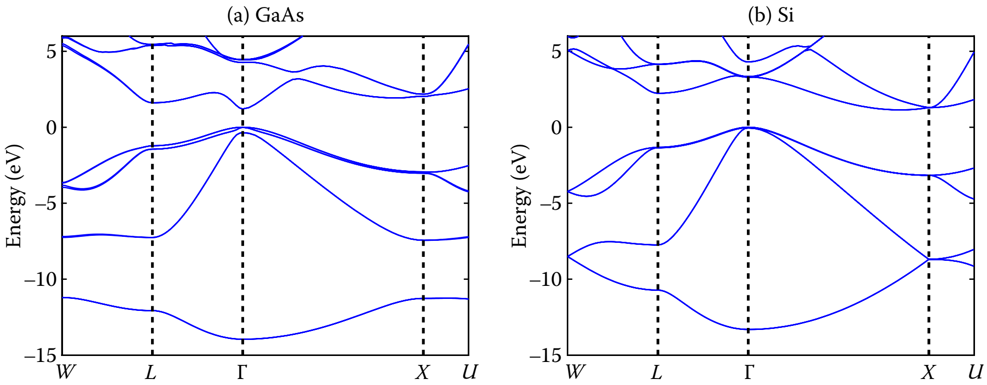


FIGURE 1.2 Electronic band structure of (a) GaAs and (b) Si calculated using the Heyd–Scuseria–Ernzerhof hybrid-functional approach [5] in the framework of density functional theory. For GaAs, the plane wave cutoff energy was set to 600 eV and the Ga semicore *d*-states are treated as valence electrons. A Γ-centered 6 × 6 × 6 *k*-point mesh was chosen. For Si similar settings have been used. In this case, the cutoff energy was 245 eV and the *k*-point mesh is 10 × 10 × 10. We have set the zero of energy in both cases to be at the valence band maximum, E_v. The calculated conduction band minimum is at energy $E_c = 1.2$ eV in GaAs and at $E_c = 1.1$ eV in Si, giving an energy gap of magnitude $E_g = 1.2$ eV and 1.1 eV, respectively.

In tetrahedrally bonded semiconductor materials, the valence band maximum is always at the Γ-point. When neglecting the electron spin, the valence band maximum consists of three degenerate *p*-like bonding states. However, in reality, we need to include the electron spin in the consideration. Starting from an isolated atom, the electron experiences an additional potential due to the interaction of the spin magnetic moment with its orbital angular momentum. This interaction is the so-called *spin–orbit interaction*, which splits the *p*-states in an isolated atom [7]. In solids, it splits the six states (including spin-up and spin-down components) of the otherwise degenerate valence band maximum, with two states shifted energetically below the other four states by an energy of Δ_{so}. The magnitude of Δ_{so} increases with atomic number [8]. It should be noted that the spin–orbit interaction not only affects the band structure of the valence band maximum, but can also lift other degeneracies in the band structure [9].

We have reviewed thus far only the general properties of bulk band structures. However, the question of how to calculate the band structure for real materials has not been addressed. To make predictions of the band structure of materials and also then to be able to tailor their properties, one needs tools that allow for the accurate calculation of the electronic structure. Different theoretical approaches that allow for a calculation of the electronic structure of semiconductor materials will be discussed and reviewed in the next sections. We introduce methods that range from first-principles calculations up to semiempirical models. We start in the next section with the "exact" solution of the Schrödinger equation to predict the band structure of a semiconductor material.

1.4 Schrödinger Equation and "Exact" Solution Using LDA–DFT

In general, for an accurate calculation of the electronic structure of materials, the Coulomb interaction between the carriers has to be considered. As already mentioned earlier, one is left with a highly complex many-body problem, even when making use of Bloch's theorem. However, the ground-breaking ideas and approaches of Hohenberg and Kohn as well as of Kohn and Sham paved the way for *density functional theory* (DFT), which has become the workhorse for first-principles electron structure calculations. Especially, the so-called *local density approximation* (LDA) has become a widely used approach to calculate the electronic structure of materials within DFT. In this section, we review the basic ideas of DFT, including how one can derive the Kohn–Sham equations and the approach taken to obtain the LDA.

1.4.1 Preliminaries

We focus in a first step on a nonrelativistic Hamiltonian for an N-electron system that accounts for the repulsive electron–electron interaction and that includes an external potential, which in solids can be assumed to be the periodic crystal lattice. The corresponding stationary many-body Hamiltonian $\hat{H}$ reads

$$\hat{H} = \underbrace{\sum_{i}^{N} \frac{-\hbar^2}{2m} \nabla_i^2}_{\hat{T}} + \underbrace{\sum_{i}^{N} v\left(\mathbf{r}_i\right)}_{\hat{V}} + \underbrace{\frac{1}{2} \sum_{i,j}^{N} \frac{e^2}{|\mathbf{r}_i - \mathbf{r}_j|}}_{\hat{W}} \,. \tag{1.10}$$

Here $\hat{T}$ is the operator for the kinetic energy, $\hat{V}$ for the potential energy, and $\hat{W}$ is the operator for the (repulsive) two-particle interaction. The kinetic energy operator $\hat{T}$ and the operator for the Coulomb interaction $\hat{W}$ are the same for all electronic systems, that is, atoms, molecules, or solids. The difference between these systems arises from differences in the form of $v(\mathbf{r})$ and also in the number of electrons in the system.

One quantity of interest for an electronic system is the ground state energy, which we denote here as E_{GS}. To calculate E_{GS} one would have to solve the full time-independent Schrödinger equation for the N particle system and the eigenvalue with the lowest energy is E_{GS}. Obviously, this is a highly nontrivial and complex problem. The basic idea of DFT is that, instead of calculating the antisymmetric ground state wave function $\psi_{GS}\left(\mathbf{r}_1, \ldots, \mathbf{r_N}\right)$, we study the corresponding electron density $\rho_{GS}(\mathbf{r})$

$$\begin{aligned} \rho_{GS}(\mathbf{r}) &= \langle \psi_{GS} | \hat{\rho} | \psi_{GS} \rangle \\ &= \int d^3 r_1 \cdots \int d^3 r_N \psi^*_{GS}\left(\mathbf{r}_1, \ldots, \mathbf{r_N}\right) \sum_{i}^{N} \delta\left(\mathbf{r} - \mathbf{r}_i\right) \psi_{GS}\left(\mathbf{r}_1, \ldots, \mathbf{r_N}\right), \end{aligned} \tag{1.11}$$

with the density operator $\hat{\rho}$ defined as

$$\hat{\rho} = \sum_{i}^{N} \delta\left(\mathbf{r} - \mathbf{r}_i\right). \tag{1.12}$$

The benefit of studying the ground state electronic density instead of the wave function is obvious. While the ground state wave function depends on three spatial coordinates per particle, $\rho_{GS}(\mathbf{r})$ depends only on three spatial variables, since it is a collective quantity.

Hohenberg and Kohn [10] were able to prove that the ground state energy E_{GS} of a many-particle system is an unambiguous functional of the ground state electron density ρ_{GS}. This statement, however, is not obvious. We present a proof of this result, which is given in the original paper by Hohenberg and Kohn [10], in the following section.

1.4.2 Hohenberg–Kohn Theorem for Nondegenerate Ground States

We know that a wave function ψ can be connected to an energy E via [3,7]:

$$E[\psi] = \frac{\langle \psi | \hat{H} | \psi \rangle}{\langle \psi | \psi \rangle}. \tag{1.13}$$

Thus, the energy E is a functional of ψ. On the other hand, from Equation 1.11, we know also that each wave function ψ is connected to an electron density ρ. The *Hohenberg–Kohn theorem* states that for the ground state there is an unambiguous connection between the ground state energy and the density ρ_{GS} [10]. One has to prove that the connection is unique, so that two different ground states cannot have the same ρ_{GS}. Let us assume we have different ground states $|\psi_{\mathrm{GS}}\rangle \neq |\psi'_{\mathrm{GS}}\rangle$ but the same density $\rho_{\mathrm{GS}} = \rho'_{\mathrm{GS}}$. We assume here that the ground state is always nondegenerate, thus $|\psi_{\mathrm{GS}}\rangle$ and $|\psi'_{\mathrm{GS}}\rangle$ must belong to different Hamiltonians $\hat{H}$ and $\hat{H}'$, respectively.[†] As discussed earlier, as long as the number of particles in the systems are identical, the only difference in the Hamiltonians can arise from the potential energy operator ($\hat{V} \neq \hat{V}'$). Using Equation 1.12 we can write $\hat{V}$ in general as

$$\hat{V} = \int v(\mathbf{r})\hat{\rho}(\mathbf{r})\,\mathrm{d}^3r. \tag{1.14}$$

With this we can write

$$\begin{aligned} E_{\mathrm{GS}} = \langle\psi_{\mathrm{GS}}|\hat{H}|\psi_{\mathrm{GS}}\rangle &= \langle\psi_{\mathrm{GS}}|\hat{T} + \hat{V}_1 + \hat{W}|\psi_{\mathrm{GS}}\rangle \\ &< \langle\psi'_{\mathrm{GS}}|\hat{H}|\psi'_{\mathrm{GS}}\rangle = \langle\psi'_{\mathrm{GS}}|\hat{H}' + \hat{V}_1 - \hat{V}_2|\psi'_{\mathrm{GS}}\rangle \\ &= E'_{\mathrm{GS}} + \int \left(v_1(\mathbf{r}) - v_2(\mathbf{r})\right)\rho'_{\mathrm{GS}}(\mathbf{r})\mathrm{d}^3r, \end{aligned} \tag{1.15}$$

where $H' = \hat{T} + \hat{V}_2 + \hat{W}$. Along the same line we obtain the inequality

$$E'_{\mathrm{GS}} < E_{\mathrm{GS}} + \int \left(v_2(\mathbf{r}) - v_1(\mathbf{r})\right)\rho_{\mathrm{GS}}(\mathbf{r})\mathrm{d}^3r. \tag{1.16}$$

Assuming that $\rho_{\mathrm{GS}} = \rho'_{\mathrm{GS}}$ and adding Equations 1.15 and 1.16 results in the contradiction

$$E_{\mathrm{GS}} + E'_{\mathrm{GS}} < E'_{\mathrm{GS}} + E_{\mathrm{GS}}. \tag{1.17}$$

Consequently, using/choosing a particular ground state density ρ determines $\hat{V}$ uniquely and therefore the Hamiltonian $\hat{H}$ and the connected ground state energy E_{GS}.

According to the Rayleigh–Ritz variation principle [3], we know that

$$E_{\mathrm{GS}} = E[\rho_{\mathrm{GS}}(\mathbf{r})] \leq \langle\psi|H|\psi\rangle = E[\rho(\mathbf{r})]\,. \tag{1.18}$$

Thus, the electron density ρ for which the energy functional $E[\rho(\mathbf{r})]$ has its absolute minimum corresponds to the ground state density ρ_{GS} and the value of the energy functional is then the ground state energy E_{GS} of the Hamiltonian $\hat{H}$. In general, we are therefore left with a minimization problem. We show in the next section that by starting from the variational problem one can derive the famous *Kohn–Sham equations*.

1.4.3 Kohn–Sham Theory and LDA

As discussed in the previous section, using the Hohenberg–Kohn theorem allowed us in principle to find the ground state density $\rho_{\mathrm{GS}}(\mathbf{r})$ and energy E_{GS} of a system of N interacting particles by minimizing the energy functional $E[\rho]$. However, even though this provides a general framework, it does not immediately give the form of the different functionals, for example, how the kinetic energy T depends on ρ. Kohn and Sham [12] introduced a scheme with an effective single-particle equation in which a particular part of the effective potential, namely the *exchange-correlation part*, requires approximations. The derived effective single-particle Schrödinger equation is known as the Kohn–Sham equation. Here, we discuss the simplest and most widely used approximation for the *exchange-correlation functional*, which gives the so-called *local density approximation* (LDA) [3,13].

[†] A discussion of how to extend the Hohenberg–Kohn theorem to a degenerate ground state can be found, for example, in Ref. [11].

As discussed in the previous section, in general, for Equation 1.18, we are left with the variational problem

$$\delta E[\rho(\mathbf{r})] = 0, \tag{1.19}$$

under the constraint that we have N particles in the system:

$$\int \rho(\mathbf{r})\,\mathrm{d}^3r = N. \tag{1.20}$$

Including this constraint into the variational problem via a Lagrange multiplier μ, one is left with [10]

$$\delta\left(E[\rho(\mathbf{r})] - \mu\left(\left(\int \rho(\mathbf{r})\,\mathrm{d}^3r - N\right)\right)\right) = 0. \tag{1.21}$$

Based on Equation 1.10, the energy functional $E[\rho(\mathbf{r})]$ is given by

$$E[\rho(\mathbf{r})] = T[\rho(\mathbf{r})] + V[\rho(\mathbf{r})] + W[\rho(\mathbf{r})]. \tag{1.22}$$

Here one encounters the problem that the functional dependence of the kinetic energy T and the interaction part W on the density ρ is unknown. Only for the external potential we know

$$V[\rho(\mathbf{r})] = \int \mathrm{d}^3r\, v(\mathbf{r})\rho(\mathbf{r})\ . \tag{1.23}$$

A first ansatz for $W[\rho(\mathbf{r})]$ could be

$$W[\rho(\mathbf{r})] = \frac{e^2}{2}\int \mathrm{d}^3r \int \mathrm{d}^3r' \frac{\rho(\mathbf{r})\,\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|}, \tag{1.24}$$

which would give the *Hartree contribution* to the interaction functional. However, from simple *Hartree–Fock theory* we know already that corrections due to the indistinguishability of the particles are required [6]. Therefore, a more general ansatz for $W[\rho(\mathbf{r})]$ can be made [3]

$$W[\rho(\mathbf{r})] = \frac{e^2}{2}\int \mathrm{d}^3r \int \mathrm{d}^3r' \frac{\rho(\mathbf{r})\,\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} + E_{\mathrm{xc}}[\rho(\mathbf{r})], \tag{1.25}$$

where $E_{\mathrm{xc}}[\rho(\mathbf{r})] = E_{\mathrm{x}}[\rho(\mathbf{r})] + E_{\mathrm{c}}[\rho(\mathbf{r})]$ is the *exchange-correlation functional.* The functional can be divided into an exchange part $E_{\mathrm{x}}[\rho(\mathbf{r})]$ and a correlation part $E_{\mathrm{c}}[\rho(\mathbf{r})]$. However, together with the kinetic energy functional, the dependence of $E_{\mathrm{xc}}[\rho(\mathbf{r})]$ on the density is still unknown.

Due to these obstacles Kohn and Sham introduced the idea that the density $\rho(\mathbf{r})$ of a system with interacting particles under the influence of an external potential $v(\mathbf{r})$ can be described by a non-interacting system with an effective external potential $\tilde{v}(\mathbf{r})$ that contains the contribution from the interaction part.

This concept now leads to the so-called *Kohn–Sham equations* [12], which can be solved once a choice is made for the exchange-correlation functional E_{xc}.

In the case of a noninteracting N-particle system, the corresponding Hamiltonian $\hat{H}_0$ reads

$$\hat{H}_0 = \sum_i^N \frac{-\hbar^2}{2m}\nabla_i^2 + \sum_i^N \tilde{v}\left(\mathbf{r}_i\right). \tag{1.26}$$

The ground state wave function of the non-interacting system can be written as a Slater determinant; thus it is an antisymmetrized product of single-particle orbitals $\phi_i(\mathrm{r})$. The particle density ρ_0 of the noninteracting system is given by

$$\rho_0\left(\mathbf{r}\right) = \sum_i^N |\phi_i\left(\mathbf{r}\right)|^2, \tag{1.27}$$

and the corresponding energy in this state is

$$\begin{aligned} E[\rho_0(\mathbf{r})] &= \sum_i^N \frac{-\hbar^2}{2m}\int \mathrm{d}^3r\,\phi_i^*\left(\mathbf{r}\right)\nabla^2\phi_i\left(\mathbf{r}\right) + \sum_i^N \int \mathrm{d}^3r\,\tilde{v}|\phi_i\left(\mathbf{r}\right)|^2 \\ &= \sum_i^N \frac{-\hbar^2}{2m}\int \mathrm{d}^3r\,\phi_i^*\left(\mathbf{r}\right)\nabla^2\phi_i\left(\mathbf{r}\right) + \int \mathrm{d}^3r\,\tilde{v}\rho_0\left(\mathbf{r}\right). \end{aligned} \tag{1.28}$$

Even for the much simpler case of the noninteracting system, one still does not explicitly know the dependence of the kinetic energy on the density ρ_0 explicitly. Nevertheless, in a first approximation, one can start from this assumption and use the kinetic energy functional of the noninteracting system for the interacting many-body system:

$$\begin{aligned} E[\rho\left(\mathbf{r}\right)] = {} & \sum_i^N \frac{-\hbar^2}{2m}\int \mathrm{d}^3r\,\phi_i^*\left(\mathbf{r}\right)\nabla^2\phi_i\left(\mathbf{r}\right) + \int \mathrm{d}^3r\,v\rho\left(\mathbf{r}\right) \\ & + \frac{e^2}{2}\int \mathrm{d}^3r\int \mathrm{d}^3r'\,\frac{\rho\left(\mathbf{r}\right)\rho\left(\mathbf{r}'\right)}{|\mathbf{r}-\mathbf{r}'|} + E_{\text{xc}}[\rho\left(\mathbf{r}\right)]. \end{aligned} \tag{1.29}$$

With this approach, we assume that the ground state density of the many-body Hamiltonian can be constructed by single-particle orbitals, which is not guaranteed. Even if that would be the case, it would not necessarily mean that the kinetic energy of the interacting system has the same form as in the noninteracting system. One can nevertheless choose this ansatz, assuming that corrections to the kinetic energy functional, introduced by the interaction between the carriers, are included in the still unknown exchange-correlation functional E_{xc}. By varying/minimizing the energy functional under the constraint that the single-particle states are normalized one can find the ground state energy [10,12]

$$\frac{\delta}{\delta\phi_i^*}\left\{E[\rho\left(\mathbf{r}\right)] - \sum_j^N \epsilon_j\left(\int \mathrm{d}^3r|\phi_j|^2 - 1\right)\right\} = 0. \tag{1.30}$$

This results in

$$\left\{-\frac{\hbar^2}{2m}\nabla^2 + v(\mathbf{r}) + \int \mathrm{d}^3r'\,\frac{e^2\rho\left(\mathbf{r}'\right)}{|\mathbf{r}-\mathbf{r}'|} + \frac{\delta E_{\text{xc}}[\rho\left(\mathbf{r}\right)]}{\delta\rho}\right\}\phi_i(\mathbf{r}) = \left\{-\frac{\hbar^2}{2m}\nabla^2 + v_s(\mathbf{r})\right\}\phi_i(\mathbf{r}) = \epsilon_i\phi_i\left(\mathbf{r}\right), \tag{1.31}$$

which is obviously of the form of a single-particle Schrödinger equation with the effective potential

$$v_s(\mathbf{r}) = v(\mathbf{r}) + \int \mathrm{d}^3 r' \frac{e^2 \rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} + \frac{\delta E_{\mathrm{xc}}[\rho(\mathbf{r})]}{\delta \rho}. \tag{1.32}$$

Equations 1.31 and 1.32 together with

$$\rho(\mathbf{r}) = \sum_{i=1}^{N} |\phi_i(\mathbf{r})|^2 \tag{1.33}$$

from the so-called *Kohn–Sham equations* [12]. This set of equations has to be solved self-consistently [13].

To use the Kohn–Sham equations for practical calculations, the form of the exchange-correlation functional needs to be known. For this quantity, we have to rely on approximations. One of the first approximations for E_{xc} that has been and is still often used is the so-called local density approximation (LDA) [12,13]. In this case, one uses the homogenous electron gas as a starting point, since in this case one can calculate the dependence of exchange energy on the (**r**-independent) density ρ explicitly. Using this ansatz, the *LDA exchange-correlation functional* $E_{\mathrm{xc}}^{\mathrm{LDA}}[\rho(\mathbf{r})]$ is given by

$$E_{\mathrm{xc}}^{\mathrm{LDA}}[\rho(\mathbf{r})] = -\frac{3e^2}{4\pi}\left(3\pi^2\right)^{\frac{1}{3}} \int (\rho(\mathbf{r}))^{4/3}\,\mathrm{d}^3 r. \tag{1.34}$$

The corresponding exchange-correlation potential $v_{\mathrm{xc}}^{\mathrm{LDA}}$ required in Equation 1.32 is then given by

$$v_{\mathrm{xc}}^{\mathrm{LDA}} = \frac{\delta E_{\mathrm{xc}}^{\mathrm{LDA}}}{\delta \rho} = -e^2 \left(\frac{3}{\pi}\right)^{\frac{1}{3}} (\rho(\mathbf{r}))^{\frac{1}{3}}. \tag{1.35}$$

Consequently, in LDA we approximate inhomogeneous systems locally by a homogenous electron gas. Even though this is an extremely crude approximation, it turns out that this ansatz works remarkably well for many systems and properties, while it systematically fails for others [14,15]. The underestimation of the band gap of semiconductors and insulators is probably one of the best known problems of LDA [14]. To overcome problems related to LDA, different approaches have been suggested. We review and introduce some of them in the next section.

1.5 Beyond LDA: GGA, Hybrid Functional, GW, and Related Approaches

One way to go beyond LDA is the *gradient expansion approximation* (GEA) [16]. In this case, one includes also gradients of the density ρ in the exchange-correlation functional, presenting therefore a systematic nonlocal extension of the LDA. The exchange-correlation functional $E_{\mathrm{xc}}^{\mathrm{GEA}}[\rho(\mathbf{r})]$ in GEA can be written as [16]

$$E_{\mathrm{xc}} = E_{\mathrm{xc}}^{\mathrm{LDA}}[\rho] + \int f_1(\rho(\mathbf{r}))(\nabla\rho(\mathbf{r}))^2\,\mathrm{d}^3 r + \int f_2(\rho(\mathbf{r}))\left(\nabla^2\rho(\mathbf{r})\right)^2 \mathrm{d}^3 r \ldots. \tag{1.36}$$

The functions f_i can be determined from response theory. Unfortunately, the GEA gives no real improvement compared with LDA. Furthermore, it has been shown that the GEA even violates the sum rule [17,18]

$$\int \mathrm{d}^3 r' \left(g(\mathbf{r}, \mathbf{r}') - 1\right) \rho\left(\mathbf{r}'\right) = 1, \tag{1.37}$$

where $g(\mathbf{r}, \mathbf{r}')$ is the so-called *pair correlation function*. This quantity describes the probability of finding an electron at $\mathbf{r}'$ while there is at the same time an electron at position $\mathbf{r}$. The integrand in Equation 1.37 is usually referred to as the *exchange-correlation hole* [17,18].

To overcome these problems, exchange-correlation functionals of the following form have been suggested [17,19]:

$$E_{\mathrm{xc}}^{\mathrm{GGA}} = \int \mathrm{d}^3 r\, f(\rho(\mathbf{r}), \nabla\rho(\mathbf{r}))\,. \tag{1.38}$$

This approach is the so-called *generalized gradient approximation* (GGA). The function f is now chosen in a form that it satisfies the sum-rule of the exchange-correlation hole [17]. Additional constraints have also been included, leading to a variety of different GGA functionals. An overview is given in Ref. [20], for example.

However, even when going beyond LDA by using GGA, it does not solve the "band gap problem." In general, it is a misconception that the band gap cannot be calculated from DFT. For example, we can express the band gap of a system in the following way [21]:

$$E_{\mathrm{g}} = \underbrace{\left(E_{N-1} - E_N\right)}_{\text{Ionization Energy}} - \underbrace{\left(E_N - E_{N+1}\right)}_{\text{Electron Affinity}}\,. \tag{1.39}$$

Here, E_N denotes the ground state energy of the N-particle system, while E_{N+1} (E_{N-1}) denotes the energy with $N + 1$ ($N - 1$) particles. Since DFT gives the correct ground state energies, E_{g} should in principle be an accessible quantity from DFT calculations. For the N-particle system, the band gap E_{g} can also be obtained using the Kohn–Sham eigenvalues via [21],

$$E_{\mathrm{g}} = E_{\mathrm{g}}^{\mathrm{KS}} + \Delta_{\mathrm{xc}}, \tag{1.40}$$

where $E_{\mathrm{g}}^{\mathrm{KS}}$ is the gap from Kohn–Sham theory and Δ_{xc} is given by [21]

$$\Delta_{\mathrm{xc}} = \frac{\delta\epsilon}{\delta\rho}|_{N+\delta N} - \frac{\delta\epsilon}{\delta\rho}|_{N-\delta N}, \tag{1.41}$$

where the number of particles is changed by an integer number δN. However, in Kohn–Sham theory, even when going beyond LDA by, for example, using gradient corrections in the exchange-correlation functionals, $\Delta_{\mathrm{xc}} = 0$. This originates from the fact that calculating the band gap is intrinsically a many-body effect, for which the independent electron picture underlying the Kohn–Sham approach therefore breaks down [21].

On the other hand, when dealing with isolated atoms *Hartree–Fock* (HF) *theory* allows for an accurate description of these systems. This stems from the fact that HF theory accounts for the exchange part exactly, and includes self-interaction and a derivative discontinuity [22,23]. However, correlation effects are not included in HF, which are important for solids or larger molecules. These effects are included in the local exchange-correlation functionals. Thus, combining these techniques provides a promising route to establish *hybrid functionals*. Becke introduced this idea in 1993 and considered an adiabatic connection between a noninteracting and a fully interacting system [24]

$$E_{\mathrm{xc}}[\rho] = \int_0^1 \mathrm{d}\lambda\, E_{\mathrm{xc},\lambda}[\rho]. \tag{1.42}$$

Here, $E_{xc}[\rho]$ is the exchange-correlation functional, while $E_{xc,\lambda}[\rho]$ denotes the exchange-correlation functional for a given value of λ, which determines the strength of the electron–electron interaction. In the limit of the noninteracting system, HF provides an accurate description, since the wave function is given by a single Slater determinant. In the limit of the fully interacting system, as an approximation, local exchange-correlation functionals are used. Initially, Becke [24] proposed the following expression for $E_{xc,\lambda}[\rho]$:

$$E_{xc,\lambda}[\rho] = (1-\lambda)\, E_x^{HF} + \lambda E_{xc}^{LDA}. \tag{1.43}$$

The functional E_x^{HF} is the HF exchange functional and E_{xc}^{LDA} denotes the exchange-correlation functional in LDA. Evaluating Equation 1.42 with $E_{xc,\lambda}[\rho]$ from Equation 1.43, one is left with [24]

$$E_{xc,\lambda}[\rho] = \frac{1}{2} E_x^{HF} + \frac{1}{2} E_{xc}^{LDA}. \tag{1.44}$$

This results in a hybrid functional, made up of 50% HF exchange and 50% LDA exchange correlation.

Using a semiempirical approach, Becke modified this approach further by adding nonlocal GGA type contributions [25,26]. These semiempirical hybrid functionals have then been benchmarked against sets of molecules. Becke introduced different sets of these semiempirical functionals, which differ in their number of free parameters. An example is the so-called *B1X functional*, which is of the form [26]

$$E_{xc}^{B1X}[\rho] = E_{xc}^{X}[\rho] + a\left(E_x^{HF}[\rho] - E_x^{X}[\rho]\right), \tag{1.45}$$

where X denotes a particular GGA functional.

Another hybrid functional, denoted as *PBE0*, was introduced by Perdew, Burke, and Ernzerhof [27], by assuming that

$$E_{xc,\lambda}^{PBE0}[\rho] = (1-\lambda)^{n-1}\left(E_x^{HF} - E_x^{PBE}\right) + E_{xc}^{PBE}, \tag{1.46}$$

where the GGA functional is $X =$ PBE and $n \geq 1$ is an integer, which has to be determined. Equations 1.46 and 1.42 yield

$$E_{xc}^{PBE0}[\rho] = E_{xc}^{PBE}[\rho] + \frac{1}{n}\left(E_x^{HF}[\rho] - E_x^{PBE}[\rho]\right). \tag{1.47}$$

Choosing $\frac{1}{n} = a$ gives the B1X functional (cf. Equation 1.45). The advantage of the hybrid functional $E_{xc}^{PBE0}[\rho]$ is that it can be related to the *Møller–Plesset perturbation theory* [28]. Møller–Plesset perturbation theory yields, when taken to the fourth order, very good atomization energies for many materials. Such an expansion up to fourth order results in $n = 4$ and thus [27]

$$\begin{aligned} E_{xc}^{PBE0}[\rho] &= E_{xc}^{PBE}[\rho] + \frac{1}{4}\left(E_x^{HF}[\rho] - E_x^{PBE}[\rho]\right) \\ &= E_c^{PBE}[\rho] + \frac{1}{4}E_x^{HF}[\rho] + \frac{3}{4}E_x^{PBE}[\rho]. \end{aligned} \tag{1.48}$$

These types of hybrid functionals have been successfully applied to a variety of different materials to predict electronic and structural properties [29–32]. However, the computational effort is much higher than for conventional local functionals. This results mainly from the calculation of the long-range (LR) part of the Coulomb potential in the HF exchange [33]. Therefore, within this framework only systems with a few atoms can be treated.

However, the LR exchange part vanishes in a real system since interactions between electrons far apart are screened. In 2003, Heyd et al. introduced the idea of separating the exchange part in PBE0 functionals into LR and short-range (SR) contributions [5]:

$$E_x^{\mathrm{PBE0}} = aE_x^{\mathrm{HF,SR}} + aE_x^{\mathrm{HF,LR}} + (1-a)E_x^{\mathrm{PBE,SR}} + (1-a)E_x^{\mathrm{PBE,LR}}. \tag{1.49}$$

Heyd et al. [5] proposed then to split the Coulomb interaction $v(r)$ into LR and SR contributions by applying the error function $\mathrm{erf}(x)$ and its complement $\mathrm{erfc}(x) = 1 - \mathrm{erf}(x)$:

$$v(r) = S_\mu(r) + L_\mu(r) = \frac{\mathrm{erfc}(\mu r)}{r} + \frac{\mathrm{erf}(\mu r)}{r}, \tag{1.50}$$

where the screening parameter μ is adjustable and $r = |\mathbf{r} - \mathbf{r}'|$. Numerical tests revealed that for realistic screening parameter values, for example, $\mu = 0.15a_0^{-1}$, where a_0 is the Bohr radius, the contribution to the functional from the LR parts in Equation 1.49 is small and that additional HF and PBE contributions tend to cancel each other. Therefore, neglecting these terms and assuming $a = 1/4$, the resulting Heyd–Scuseria–Ernzerhof *screened hybrid functional*, referred to as HSE, is given by [5]

$$\begin{aligned} E_{\mathrm{xc}}^{\mathrm{HSE}} &= E_{\mathrm{c}}^{\mathrm{PBE}} + aE_{\mathrm{x}}^{\mathrm{HF,SR}}(\mu) + (1-a)E_{\mathrm{x}}^{\mathrm{PBE,SR}}(\mu) + E_{\mathrm{x}}^{\mathrm{PBE,LR}}(\mu) \\ &= E_{\mathrm{c}}^{\mathrm{PBE}} + \frac{1}{4}E_{\mathrm{x}}^{\mathrm{HF,SR}}(\mu) + \frac{3}{4}E_{\mathrm{x}}^{\mathrm{PBE,SR}}(\mu) + E_{\mathrm{x}}^{\mathrm{PBE,LR}}(\mu). \end{aligned} \tag{1.51}$$

As discussed by Paier et al. [34] in detail, the HSE functional can significantly reduce the computational cost. This results from the increased locality of the exchange interactions, which allows for the evaluation of the SR part on a coarser k mesh in comparison to a calculation using the full exchange operator. The chosen value for μ was refined over the years so that one finds in the literature calculations using so-called HSE03 or HSE06 functionals, where the HSE functionals differ in the value of μ [5].

Hybrid functionals are a step forward in comparison to local functionals. However, they do not present a universal improvement. This is related to the presence of the exact exchange mixing parameter a. For example, the mixing parameter a required to reproduce correct defect geometries might be very different from the value chosen to optimize the band gap. Thus, for certain systems there is not a unique value for a that describes both the bulk band edges and the ground state geometries for certain defects [35,36]. Also for band gaps of semiconductor alloys, the choice of a is problematic since the binary compounds may require different a values for an accurate band gap description [37]. Therefore, the choice for alloys is difficult.

As discussed in detail in Ref. [38], there are several ways to go beyond standard DFT that can offer better accuracy, but keeping it still systematic and general. Here, *many-body perturbation theory* (MBPT) using an efficiently achievable and trustworthy reference state has been widely used. In the solid-state community, Hedin's *GW approximation* [39], which represents such an MBPT treatment, has become very popular since it allows for calculations of quasi-particle band structures which can be compared to direct and inverse photoemission spectra [23,40]. A detailed description of MBPT approaches is beyond the scope of the present work and can be found, for example, in Ref. [38].

Overall, the basics concepts of DFT and some of the current state-of-the-art first-priciples approaches have been discussed in this section. These approaches allow in general for a description of electronic band structure over the entire first Brillouin zone. However, when looking back at the general discussion of the electronic band structure of GaAs in Section 1.3, it was highlighted that this material system is of particular interest for optoelectronic devices because of the band structure specifics around $\mathbf{k} = \mathbf{0}$. Thus, when interested in describing the optical properties of these systems this region is of central importance, while the behavior of the highest valence and lowest conduction bands at the boundaries of the first Brillouin zone is generally of secondary importance. Therefore, more empirical methods, which describe a particular part of the band structure accurately, have been developed over the years. These (semi)empirical models are computationally less demanding and, due to their semiempirical nature, they can overcome, for example, the band gap problem of standard DFT calculations. Additionally, because they are computationally relatively cheap, they become highly attractive for the calculation of electronic and optical properties of semiconductor nanostructures. In the next sections the basic concepts of these methods are reviewed, starting with the so-called *(multiband)* $\mathbf{k}\cdot\mathbf{p}$ *method.*

1.6 k•p and Envelope-Function Approaches

$\mathbf{k}\cdot\mathbf{p}$ theory is a perturbation method, whereby if the exact energy levels are known at a given point in the Brillouin zone, for example, say $\mathbf{k} = \mathbf{0}$, it can be used to find the band structure near the $\mathbf{k}$ in question. We outline in the following the general idea behind the $\mathbf{k}\cdot\mathbf{p}$ approach. More details can be found, for example, in Refs. [8,41].

We start again from the Hamiltonian, H_0, of a single electron moving in a periodic solid:

$$H_0 = -\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r}), \tag{1.52}$$

with $V(\mathbf{r} + \mathbf{R}) = V(\mathbf{r})$. From Bloch's theorem we know that the eigenstates, $\psi_{n\mathbf{k}}(\mathbf{r})$, can be written as the product of a plane wave, $e^{i\mathbf{k}\cdot\mathbf{r}}$, times a periodic function, $u_{n\mathbf{k}}(\mathbf{r})$, with associated energy levels, $E_n(\mathbf{k})$. For a chosen value of $\mathbf{k}$, say $\mathbf{k}_0$, Schrödinger's equation can be written as

$$H_0\psi_{n\mathbf{k}_0}(\mathbf{r}) = \left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right)\left(e^{i\mathbf{k}_0\cdot\mathbf{r}}u_{n\mathbf{k}_0}(\mathbf{r})\right) = E_n(\mathbf{k}_0)\left(e^{i\mathbf{k}_0\cdot\mathbf{r}}u_n\mathbf{k}_0(\mathbf{r})\right). \tag{1.53}$$

In the following, we assume that the allowed energy levels $E_n(\mathbf{k}_0)$ at $\mathbf{k}_0$ are known. The aim is now to find the energy levels, $E_n(\mathbf{k})$, at a wave vector $\mathbf{k}$ close to $\mathbf{k}_0$, for which the Schrödinger equation can be written as

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right)\left(e^{i\mathbf{k}\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r})\right) = E_n(\mathbf{k})\left(e^{i\mathbf{k}\cdot\mathbf{r}}u_n\mathbf{k}(\mathbf{r})\right). \tag{1.54}$$

Since we are interested in a $\mathbf{k}$ close to $\mathbf{k}_0$, we rewrite Equation 1.54 as follows:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right)e^{i(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}}\left(e^{i\mathbf{k}_0\cdot\mathbf{r}}u_{nk}(\mathbf{r})\right) = E_n(\mathbf{k})e^{i(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}}\left(e^{i\mathbf{k}_0\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r})\right). \tag{1.55}$$

Multiplying both sides of Equation 1.55 from the left by $e^{-i(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}}$, the following modified differential equation is obtained:

$$\begin{aligned}\left[e^{-i(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}}H_0e^{i(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}}\right]\left(e^{i\mathbf{k}_0\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r})\right) &= \left[e^{-i(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}}E_n(\mathbf{k})e^{i(\mathbf{k}-\mathbf{k}_0)\cdot\mathbf{r}}\right]e^{i\mathbf{k}_0\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r}) \\ &= E_n(\mathbf{k})\left(e^{i\mathbf{k}_0\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r})\right).\end{aligned} \tag{1.56}$$

When transforming Equation 1.55 into Equation 1.56, the Hamiltonian is now a $\mathbf{k}$-dependent Hamiltonian, which is denoted in the following by $H_\mathbf{q}$, where $\mathbf{q} = \mathbf{k} - \mathbf{k}_0$. Equation 1.56 can then be rewritten as

$$H_\mathbf{q}\phi_{n\mathbf{k}}(\mathbf{r}) = e^{-i\mathbf{q}\cdot\mathbf{r}}\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right)e^{i\mathbf{q}\cdot\mathbf{r}}\phi_{n\mathbf{k}}(\mathbf{r}), \tag{1.57}$$

where $\phi_{n\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}_0\cdot\mathbf{r}}u_{n\mathbf{k}}(\mathbf{r})$. With a small amount of algebra, Equation 1.57 becomes

$$\begin{aligned} H_{\mathbf{q}}\phi_{n\mathbf{k}}(\mathbf{r}) &= \left(-\frac{\hbar^2}{2m}\nabla^2 + \frac{\hbar^2}{m}\mathbf{q}\cdot\frac{1}{i}\nabla + \frac{\hbar^2 q^2}{2m} + V(r)\right)\phi_{n\mathbf{k}}(\mathbf{r}) \\ &= \left[H_0 + \frac{\hbar}{m}\mathbf{q}\cdot\mathbf{p} + \frac{\hbar^2 q^2}{2m}\right]\phi_{n\mathbf{k}}(\mathbf{r}), \end{aligned} \tag{1.58}$$

where we have used Equation 1.52 to introduce H_0, and replaced $\frac{\hbar}{i}\nabla$ by the momentum operator, $\mathbf{p}$.

Equation 1.58 forms the foundation of the $\mathbf{k}\cdot\mathbf{p}$ method. In the case of $\mathbf{q} = \mathbf{0}$, Equation 1.58 reduces to the standard form of Schrödinger's time-independent equation. When dealing with a direct gap semiconductor one chooses, in general, $\mathbf{k}_0 = \mathbf{0}$, the Γ-point. At the Γ-point one generally knows or can estimate the values of all the relevant zone center energies, $E_n(\mathbf{0})$. The contribution H',

$$H' = \frac{\hbar}{m}\mathbf{q}\cdot\mathbf{p} + \frac{\hbar^2 q^2}{2m}, \tag{1.59}$$

to the full Hamiltonian can be regarded as a perturbation to the zone center Hamiltonian, H_0. Using *second-order perturbation theory* one can evaluate the variation of the energy levels $E_n(\mathbf{k})$ with wave vector $\mathbf{k}$ (= $\mathbf{q}$) close to the Γ-point. For a nondegenerate band, the energy of the nth band in the neighborhood of $\mathbf{k} = \mathbf{0}$ is given by

$$E_n(\mathbf{k}) = E_n(0) + \frac{\hbar}{m}\mathbf{k}\cdot\mathbf{p}_{nn} + \frac{\hbar^2 k^2}{2m} + \frac{\hbar^2}{m^2}\sum_{n'\neq n}\frac{|\mathbf{k}\cdot\mathbf{p}_{nn'}|^2}{E_n(0) - E_{n'}(0)}. \tag{1.60}$$

Here $\mathbf{p}_{nn'}$ denotes the *momentum matrix element* between the nth and n'th zone center states, $u_{n\mathbf{0}}(\mathbf{r})$ and $u_{n'\mathbf{0}}(\mathbf{r})$,

$$\mathbf{p}_{nn'} = \int_V \mathrm{d}^3 r\, u^*_{n\mathbf{0}}(\mathbf{r})\mathbf{p}u_{n'\mathbf{0}}(\mathbf{r}) = \langle u_{n\mathbf{0}}|\mathbf{p}|u_{n'\mathbf{0}}\rangle, \tag{1.61}$$

with the integration taken over a unit cell of the crystal structure. Equation 1.60 shows that the energy $E_n(\mathbf{k})$ at a given wave vector $\mathbf{k}$ can be expressed in terms of the known energies at $\mathbf{k} = \mathbf{0}$, and the interactions between the zone center states through the momentum matrix elements $\mathbf{p}_{nn'}$. From Equation 1.60 one can also infer that corrections linear in $\mathbf{k}$ ($\sim \mathbf{k}\cdot\mathbf{p}_{nn}$) and quadratic in $\mathbf{k}$ are present. It has been shown by Kane [42] that in a diamond crystal structure the linear term is by symmetry identically equal to zero. For III–V semiconductors, the effects of the linear in $\mathbf{k}$ terms are generally negligibly small and are most often ignored.

To derive Equation 1.60, we have assumed a single nondegenerate band. As discussed in Section 1.3, for a realistic description of the bulk band structure near the Γ-point of semiconductors, effects such as spin–orbit interaction and also of band degeneracies have to be included. The practical application of the $\mathbf{k}\cdot\mathbf{p}$ model, therefore, looks rather complicated. However, two main reasons help ensure its usefulness [42,43]:

- Given that in Equation 1.60 the quadratic in $\mathbf{k}$ correction to $E_n(\mathbf{0})$ depends inversely on $E_n(\mathbf{0}) - E_{n'}(\mathbf{0})$, in most cases one has to deal with a very small number of bands, which are close to each other in energy. Higher and lower bands can be ignored, since the denominator in this case will be large, resulting in small corrections.
- When restricting the number of bands in Equation 1.60, and then starting to fit to experimental data, such as measured energy gaps $E_n(\mathbf{0}) - E_{n'}(\mathbf{0})$, it is found that the matrix elements $\mathbf{p}_{nn'}$ are remarkably constant between different semiconductors, with $\mathbf{p}_{nn'}$ identically zero for many pairs of bands n and n'.

When describing the band structures of realistic semiconductor materials such as GaAs or Si, grown in the zincblende or diamond crystal structure, respectively, the valence band structure, in particular, presents a complicated situation (cf. Section 1.3). This originates from the fact that on a microscopic level the conduction band is mainly described by an s-like state, while the valence band states are predominantly p-like in character near $\mathbf{k} = \mathbf{0}$. The "Bloch" wave functions have therefore the symmetry of p_x, p_y, and p_z-like atomic orbitals and will all contribute to the valence band states. This needs to be included in Equation 1.60. Focusing on the valence band structure, from Figure 1.2 one can infer that at the Γ-point one is left with two doubly degenerate valence bands and one which is shifted downward in energy. This third band is usually referred to as the *split-off* (SO) *band*. The two doubly degenerate bands at Γ are the so-called *heavy-hole* (HH) and *light-hole* (LH) *bands*. The degeneracy between the HH and LH bands is lifted for finite values of $\mathbf{k}$. They receive their names from the fact that the energy of the HH band varies at a slower rate than the LH band when moving away from $\mathbf{k} = \mathbf{0}$.

For bulk systems, these bands are eigenstates of the system. However, in the case of a nanostructure where the potential varies spatially, this is no longer the case [44]. For example, in a QW system, the valence sub-band states can be mixtures of HH, LH, and SO bands. Thus, Hamiltonians describing this kind of mixing between different states are given by matrices or a system of coupled time-independent Schrödinger equations [44]. The solution of this set of coupled time-independent Schrödinger equations will give the energy eigenvalues and the corresponding eigenstates as a set of *envelope functions*, which are themselves given in terms of the contribution of the underlying basis states, for example, the bulk HH, LH, or SO bands. This approach is referred to as the *multiband envelope function method* or the *multiband effective mass approximation*. When the interaction between the different bands is described in terms of $\mathbf{k}\cdot\mathbf{p}$ perturbation theory, it is also called a "multiband $\mathbf{k}\cdot\mathbf{p}$ method" [44].

These multiband envelope or $\mathbf{k}\cdot\mathbf{p}$ methods can be at different levels of sophistication, basically depending on the number of bands included in the description. The number of explicitly included bulk bands in the model also describes the name of the multiband Hamiltonian [44]. For instance, four-band (HH, LH) and six-band (HH, LH, SO) Hamiltonians for valence band envelope wave functions for zincblende materials were introduced by Luttinger and Kohn [45]. Pidgeon and Brown [46] extended these models later on to an eight-band model, which includes also the conduction band. Chuang and Chang introduced in 1996 an eight-band model for wurtzite semiconductors [47]. A detailed overview of different types of multiband $\mathbf{k}\cdot\mathbf{p}$ models is given in the textbook by Lew Yan Voon and Willatzen [41].

It is important to note that all these Hamiltonians implicitly include contributions from more remote bands, which are not explicitly included. This originates from the fact that the used material parameters incorporate the effects of the omitted bands. For example, the simple conduction band effective mass model is a special case of an envelope-function approximation. Here, the existence of the other bands in the system is accounted for by using the effective electron mass instead of the free electron mass. In this single-band approach, only the conduction band envelope function is calculated.

Introductions to multiband $\mathbf{k}\cdot\mathbf{p}$ models and their applications to nanostructures can be found in several textbooks [41,44,48]. These approaches have been use extensively in the literature to study the electronic and optical properties of semiconductor heterostructures. Due to the limited number of bands normally included in such a method, these models can be applied, for example, to embedded semiconductor quantum dots of realistic size and shape, which is beyond a DFT treatment, since supercell sizes can easily reach hundreds of thousands of atoms [49]. However, one of the main drawbacks of conventional six- or eight-band models is that they do not necessarily capture the symmetry of the whole system (crystal structure plus quantum dot geometry) correctly [50–52]. Consider the example of a pyramidal, zincblende InAs/-GaAs quantum dot with a square base. Within eight-band $\mathbf{k}\cdot\mathbf{p}$ theory, neglecting piezoelectric effects, this system exhibits a C_{4v} symmetry [50,51], with 90° rotations about the growth direction, leaving the structure unchanged. However, when accounting for the underlying zincblende structure on an atomistic level (two atoms per unit cell), the symmetry is actually reduced to a C_{2v} symmetry [51,52]. This symmetry reduction leads to a lifting of degeneracies in the electronic structure and affects the optical properties of these systems [51,52]. Therefore, semiempirical models, which can treat large numbers of atoms plus provide an atomistic description of the electronic structure, are required. *Empirical tight-binding models* and *empirical pseudopotential theory* can provide this and are the topic of the next section.

1.7 Need for Semiempirical Atomistic Approaches—Tight-Binding and Pseudopotential Methods

In this section, we introduce the basic concepts of (empirical) tight-binding (TB) and empirical pseudopotential theory. Detailed descriptions of these approaches can be found in several textbooks [3,6,8,44]. Here, the basic concepts are summarized. We start with the (empirical) TB approach in the following section. In a second step, in Section 1.7.2, insight into the empirical pseudopotential theory is given. In both cases, the inclusion of the spin–orbit interaction is discussed.

1.7.1 The Empirical Tight-Binding Method

In contrast to the free-electron picture, within which the crystal potential in perturbation theory is often treated, the TB model describes the electronic band structure starting from the limit of isolated atoms. The basis states correspond to the localized orbitals of the different atoms. In this way, one obtains a description of the electronic properties, which simultaneously takes into account the microscopic structure of the solid and offers a transparent approach. The matrix elements in the TB approach have a simple physical interpretation, as they represent interactions between electrons on adjacent atoms.

The localized basis states $|\mathbf{R}, \alpha, \nu, \sigma\rangle$ are classified according to their unit cell $\mathbf{R}$, the type of atom α at which they are centered, the orbital type ν, and the spin σ. The basic assumptions of the TB model are that (1) a small number of basis states per unit cell is already sufficient to describe the bulk band structure and (2) that the overlap of the strongly localized atomic orbitals decreases rapidly with increasing distance between the atomic sites.

Since the inner electronic shells are only slightly affected by the field of all the other atoms, for the description of the bulk band structure it is sufficient to take into account the states of the outer shells. These orbitals then form the highest valence band and lowest conduction band. We start from electron wave functions of an isolated atom. The time-independent, single-electron Schrödinger equation for an atom located at the position $\mathbf{R}_l$ is given by

$$H^{\mathrm{at}}|\mathbf{R}_l, \alpha, \nu, \sigma\rangle = E^{\mathrm{at}}_{\alpha,\nu}|\mathbf{R}_l, \alpha, \nu, \sigma\rangle,$$

with

$$H^{\mathrm{at}} = \frac{p^2}{2m_0} + V^0(\mathbf{R}_l, \alpha),$$

where $V^0(\mathbf{R}_l, \alpha)$ denotes the atomic potential of the atom at the position $\mathbf{R}_l$. Due to the presence of all of the other atoms in the crystal, the wave functions are modified. The single-particle Hamiltonian of the periodic system can be written in the following way:

$$H^{\mathrm{bulk}} = H^{\mathrm{at}}(\mathbf{R}_l, \alpha) + \underbrace{\sum_{\substack{n\neq l\\ \alpha'}} V(\mathbf{R}_n, \alpha')}_{\Delta V(\mathbf{R}_l)}.$$

Here, $H^{\mathrm{at}}(\mathbf{R}_l, \alpha)$ is the Hamiltonian of the isolated atom α at the position $\mathbf{R}_l$ and $\Delta V(\mathbf{R}_l)$ is the potential generated by all the other ions in the lattice. The Schrödinger equation for the periodic solid is then

$$H^{\mathrm{bulk}}|\mathbf{k}\rangle = E(\mathbf{k})|\mathbf{k}\rangle, \tag{1.62}$$

where $\mathbf{k}$ denotes the crystal wave vector. To solve the Schrödinger equation, Equation 1.62, the electronic wave functions $|\mathbf{k}\rangle$ are approximated by linear combinations of atomic orbitals. Because of the translational symmetry of the crystal, the wave functions can be expressed in terms of Bloch functions:

$$|\mathbf{k}\rangle = \sqrt{\frac{V_0}{V}} \sum_n e^{i\mathbf{k}\cdot\mathbf{R}_n} \sum_{\alpha,\nu,\sigma} e^{i\mathbf{k}\cdot\boldsymbol{\Delta}_\alpha} u_{\alpha,\nu,\sigma}(\mathbf{k}) |\mathbf{R}_n, \alpha, \nu, \sigma\rangle \quad \text{with} \quad \mathbf{k} \in 1.\ \text{BZ}. \tag{1.63}$$

The position of the atom α in the unit cell $\mathbf{R}_n$ is denoted by $\boldsymbol{\Delta}_\alpha$. The volume of the unit cell is given by V_0 and the volume of the system is V. Due to the periodicity of the crystal, it is sufficient to restrict $\mathbf{k}$ to the first Brillouin zone (1. BZ). This ansatz fulfills the Bloch theorem. By plugging the wave function of Equation 1.63 into Schrödinger's equation

$$H^{\text{bulk}}|\mathbf{k}\rangle = \left[H^{\text{at}}(\mathbf{R}_l, \alpha) + \Delta V(\mathbf{R}_l)\right] |\mathbf{k}\rangle = E(\mathbf{k})|\mathbf{k}\rangle, \tag{1.64}$$

and applying the bra-vector $\langle \mathbf{k}, \alpha', \nu', \sigma'| = \sqrt{\frac{V_0}{V}} \sum_m e^{-i(\mathbf{k}\cdot\mathbf{R}_m + \mathbf{k}\cdot\boldsymbol{\Delta}_{\alpha'})} \langle \mathbf{R}_m, \alpha', \nu', \sigma'|$, one is left with a matrix equation

$$\sum_{\alpha,\nu,\sigma} H^{\text{bulk}}_{\alpha',\nu'\,\sigma';\alpha,\nu,\sigma}(\mathbf{k}) u_{\alpha,\nu,\sigma}(\mathbf{k}) = E(\mathbf{k}) \sum_{\alpha,\nu,\sigma} S_{\alpha',\nu'\,\sigma';\alpha,\nu,\sigma}(\mathbf{k}) u_{\alpha,\nu,\sigma}(\mathbf{k}), \tag{1.65}$$

instead of a differential equation (Equation 1.64). For each $\mathbf{k}$, this constitutes a *generalized eigenvalue problem* for the matrix $H^{\text{bulk}}(\mathbf{k})$, with a matrix S instead of the identity that occurs in an ordinary eigenvalue problem. $E(\mathbf{k})$ represents the eigenvalue, corresponding to the eigenvector $\mathbf{u}(\mathbf{k})$. The elements of the Hamiltonian matrix read:

$$H^{\text{bulk}}_{\alpha',\nu'\,\sigma';\alpha,\nu,\sigma}(\mathbf{k}) = \frac{V_0}{V} \sum_{n,m} e^{i\mathbf{k}\cdot(\mathbf{R}_n + \boldsymbol{\Delta}_\alpha - \mathbf{R}_m - \boldsymbol{\Delta}_{\alpha'})} \langle \mathbf{R}_m, \alpha', \nu', \sigma'|H^{\text{bulk}}|\mathbf{R}_n, \alpha, \nu, \sigma\rangle, \tag{1.66}$$

and the overlap-matrix elements are given by $S_{\alpha',\nu'\,\sigma';\alpha,\nu,\sigma}(\mathbf{k})$:

$$S_{\alpha',\nu'\,\sigma';\alpha,\nu,\sigma}(\mathbf{k}) = \frac{V_0}{V} \sum_{n,m} e^{i\mathbf{k}\cdot(\mathbf{R}_n + \boldsymbol{\Delta}_\alpha - \mathbf{R}_m - \boldsymbol{\Delta}_{\alpha'})} \langle \mathbf{R}_m, \alpha', \nu', \sigma'|\mathbf{R}_n, \alpha, \nu, \sigma\rangle. \tag{1.67}$$

According to the basic assumptions of a TB model, the electrons stay close to the atomic sites and the electronic wave functions centered around neighboring sites have little overlap. Consequently, there is almost no overlap between wave functions for electrons that are separated by two or more atoms (second-nearest neighbors, third-nearest neighbors, etc.). Nevertheless, the basis orbitals, and thus the Bloch sums, are in general not fully orthogonal to one another. It turns out, however, that if the localized TB basis $\mathcal{B}$ is not orthogonal, one can use a so-called *Löwdin transformation* to transform it into an orthogonal one [53]. These Löwdin orbitals are also localized and preserve the symmetry of the orbital from which they are derived. Therefore, we assume in the following an orthogonal basis set, so that the overlap-matrix elements, Equation 1.67, are given simply by

$$S_{\alpha',\alpha,\nu',\nu,\sigma',\sigma} = \delta_{\mathbf{R}_m,\mathbf{R}_n} \delta_{\alpha',\alpha} \delta_{\nu',\nu} \delta_{\sigma',\sigma}\,, \tag{1.68}$$

and the Schrödinger equation, Equation 1.65, is reduced to

$$\sum_{\alpha,\nu,\sigma} H^{\text{bulk}}_{\alpha',\nu'\,\sigma';\alpha,\nu,\sigma}(\mathbf{k}) u_{\alpha,\nu,\sigma}(\mathbf{k}) = E(\mathbf{k}) u_{\alpha',\nu',\sigma'}(\mathbf{k})\,. \tag{1.69}$$

For the sake of a simplified illustration, in the next couple of paragraphs, a system with one atom per unit cell and one orbital per site is assumed. Furthermore, a spin-independent Hamiltonian H^{bulk} is assumed. Consequently, we are left with $\sigma' = \sigma$, $\alpha' = \alpha$, and $\nu' = \nu$. In this case, the matrix elements $H^{\text{bulk}}_{\alpha',\nu',\sigma';\alpha,\nu,\sigma}(\mathbf{k}) = H^{\text{bulk}}_{\alpha,\nu,\sigma;\alpha,\nu,\sigma}(\mathbf{k})$ are given by

$$\begin{aligned}\langle \mathbf{k},\alpha,\nu,\sigma|H^{\text{bulk}}|\mathbf{k},\alpha,\nu,\sigma\rangle &= \frac{V_0}{V}\sum_{m,n} e^{i\mathbf{k}\cdot(\mathbf{R}_n-\mathbf{R}_m)}\langle \mathbf{R}_m,\alpha,\nu,\sigma|H^{\text{bulk}}|\mathbf{R}_n,\alpha,\nu,\sigma\rangle \\ &= \frac{V_0}{V}\sum_{m,n} e^{i\mathbf{k}\cdot(\mathbf{R}_n-\mathbf{R}_m)} I_{mn}.\end{aligned} \tag{1.70}$$

Due to the localized structure of the atomic-like wave functions, the integrals I_{mn} become exponentially small for large $R = |(\mathbf{R}_m - \mathbf{R}_n)|$. It is therefore reasonable to ignore all integrals outside some R_{max}, as they would bring only negligible corrections to the band structure $E(\mathbf{k})$. The leading contribution is $n = m$, the so-called onsite contributions, then nearest neighbor contributions are denoted by $n = m \pm 1$, and so on. Keeping only the leading order of the expression of Equation 1.70, the matrix elements I_{mn} can be written as

$$\begin{aligned} I_{mn} &= \delta_{n,m}\Big[\langle \mathbf{R}_m,\alpha,\nu,\sigma|H^{at}|\mathbf{R}_n,\alpha,\nu,\sigma\rangle \\ &\quad + \sum_{l\neq n}\langle \mathbf{R}_m,\alpha,\nu,\sigma|V(\mathbf{R}_l,\alpha)|\mathbf{R}_n,\alpha,\nu,\sigma\rangle\Big] \\ &\quad + \delta_{n\pm1,m}\sum_{l}\langle \mathbf{R}_m,\alpha,\nu,\sigma|V(\mathbf{R}_l,\alpha)|\mathbf{R}_n,\alpha,\nu,\sigma\rangle + \ldots \\ &\equiv \delta_{n,m}\tilde{E} + \delta_{n\pm1,m}\,\lambda + \ldots .\end{aligned} \tag{1.71}$$

$\tilde{E}$ can be interpreted as the renormalized atomic energy level in the presence of all the other atoms in the lattice. The matrix elements containing orbitals from different atomic sites are denoted by λ. Finally, one obtains from Equation 1.70:

$$\langle \mathbf{k},\alpha,\nu,\sigma|H^{\text{bulk}}|\mathbf{k},\alpha,\nu,\sigma\rangle = \frac{V_0}{V}\sum_{m,n} e^{i\mathbf{k}\cdot(\mathbf{R}_n-\mathbf{R}_m)}(\delta_{n,m}\tilde{E} + \delta_{n\pm1,m}\,\lambda + \ldots). \tag{1.72}$$

The symmetry properties of the crystal together with the symmetries of the basis states determine which matrix elements vanish and which are equal. For the calculation of the TB matrix elements, two different approaches are possible. From the knowledge of the atomic potential V^0 and the atomic orbitals, one could calculate the matrix elements of H^{bulk} as well as the overlap matrix elements $S_{\alpha',\alpha,\nu',\nu,\sigma',\sigma}$. A different way to obtain these matrix elements is to treat them as parameters. This approach leads to the so-called *empirical* TB (ETB) model. These parameters are fitted to characteristic properties of the bulk band structure, such as band gaps and effective masses. With both methods, a microscopic description of the solid arises. The empirical approach has been widely used to model semiconductor nanostructures, which consist of different semiconductor materials [54–58]. In the following, we focus on the ETB model.

As already discussed, we assume that the basis states are Löwdin orbitals. Thus, the different TB matrix elements can be symbolized by

$$\langle \mathbf{R}',\alpha',\nu',\sigma'|H^{\text{bulk}}|\mathbf{R},\alpha,\nu,\sigma\rangle = \underbrace{E_{\nu',\nu}(\mathbf{R}-\mathbf{R}')_{\alpha',\alpha}\delta_{\sigma,\sigma'}}_{\text{TB-parameter}}. \tag{1.73}$$

Here, the translational invariance of the crystal is already used. In addition, we have assumed that H^{bulk} is spin independent. A spin-dependent component of H^{bulk}, the spin–orbit coupling, will be introduced in the following section. In this section, we drop the spin index σ and denote the Hamiltonian by H_0^{bulk}.

The onsite matrix elements, $E_{\nu,\nu}(\mathbf{0})_{\alpha,\alpha}$, which are the expectation values of the Hamiltonian H_0^{bulk} between two identical atomic orbitals at the same site, correspond to the *orbital energies* which are renormalized and shifted due to the other atoms in the crystal. The matrix element $E_{\nu,\nu}(\mathbf{0})_{\alpha,\alpha}$ equals $\tilde{E}$ in Equation 1.71. The values taken for the onsite matrix elements tend in many cases to follow trends in the atomic orbital energies of the isolated atoms [59]. The off-diagonal matrix elements, which describe the coupling between different orbitals at different sites, are called *hopping matrix elements*, because they are related to the probability amplitude of an electron moving from one site to another [3].

To further reduce the number of relevant ETB matrix elements, one can use the so-called *two-center approximation* of Slater and Koster [60]. Slater and Koster proposed to neglect the so-called *three-center integrals*, which are considerably smaller than the so-called *two-center integrals*. The three-center integrals involve two orbitals located at different atoms, and a potential part at a third atom. This corresponds to contributions in Equation 1.71 where $l \neq n \neq m$. Thus, in the Slater and Koster approach, only the potential due to the two atoms at which the orbitals are located is taken into account. Thus, the (effective) potential is symmetric around the axis $\mathbf{d} = \mathbf{R}' + \mathbf{\Delta}_{\alpha'} - (\mathbf{R} + \mathbf{\Delta}_{\alpha})$ between the two atoms. In this approximation, all hopping matrix elements vanish, if the two involved orbitals are eigenstates of the angular momentum $L_{\mathbf{d}} = \mathbf{L}\cdot\frac{\mathbf{d}}{d}$ in $\mathbf{d}$-direction with different eigenvalues $\hbar m_{\mathbf{d}} \neq \hbar m'_{\mathbf{d}}$. Since the (effective) Hamiltonian H, which contains the axially symmetric potential of the two atoms, commutes with $L_{\mathbf{d}}$, this statement follows from

$$0 = \langle i| \left[H, L_{\mathbf{d}}\right] |i'\rangle = \langle i|HL_{\mathbf{d}}|i'\rangle - \langle i|L_{\mathbf{d}}H|i'\rangle = \hbar \left(m'_{\mathbf{d}} - m_{\mathbf{d}}\right) \langle i|H|i'\rangle. \tag{1.74}$$

For an *sp*-bonding, there are only four nonzero hopping integrals as indicated in Figure 1.3, in which σ ($m_{\mathbf{d}} = 0$) and π ($m_{\mathbf{d}} = \pm 1$) bondings are defined such that the axes of the involved *p*-orbitals are parallel and normal to the interatomic vector **d**, respectively. So far, the electronic wave functions are expanded in terms of the *p*-orbitals along the Cartesian *x*-, *y*-, and *z*-axes, whereas the hopping integrals are parameterized for *p*-orbitals that are parallel or normal to the bonding directions. To construct the ETB matrix elements, in the general case it is fruitful to decompose the Cartesian *p*-orbitals into bond-parallel and bond-normal *p*-orbital components.

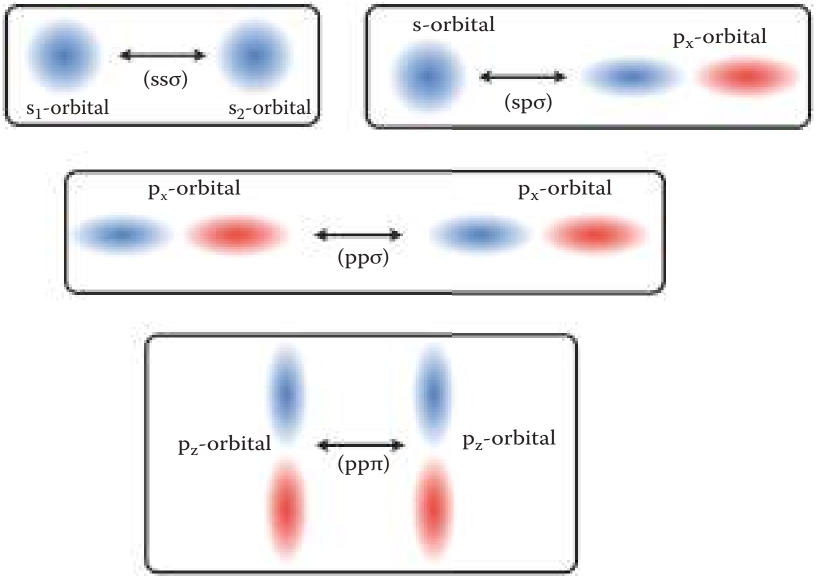


FIGURE 1.3 Schematic illustration of the nonzero hopping matrix elements $V_{ll'm}$ for *sp*-bonding. Orbital types are denoted by the indices l and l', while the z-component of the orbital angular momentum for rotation around the bonding direction is given by m.

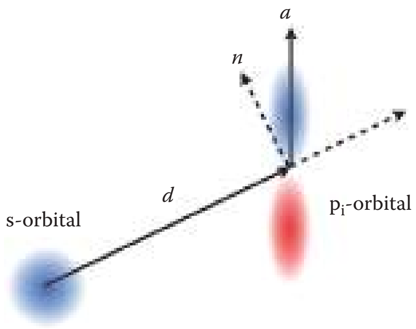


FIGURE 1.4 Schematic illustration of the overlap between s- and p_i-orbital along the vector **d** joining the two atoms.

As an example of this procedure, we consider the Hamiltonian matrix element $\langle s|H|p_i\rangle$, between the s-orbital, $|s\rangle$, and one of the p-orbitals, $|p_i\rangle (i = x, y, z)$, localized at different atoms. Let $\mathbf{e}_d$ be the unit vector along the bond **d** from the first atom to the second and **a** the unit vector along one of the Cartesian (x−, y−, or z−) axes, as shown in Figure 1.4. We first decompose the p-orbital along **a**, $|p_a\rangle$, into two p-orbitals that are parallel and normal to **d**, respectively,

$$|p_a\rangle = \mathbf{a} \cdot \mathbf{e_d}|p_\sigma\rangle + \mathbf{a} \cdot \mathbf{n}|p_\pi\rangle,$$

where **n** is the unit vector normal to **d** within the plane spanned by **d** and **a**. The Hamiltonian matrix element is then given by

$$\begin{aligned}\langle s_1|H|p_{2,a}\rangle &= \mathbf{a} \cdot \mathbf{e_d}\langle s|H|p_\sigma\rangle + \mathbf{a} \cdot \mathbf{n}\langle s|H|p_\pi\rangle \\ &= \mathbf{a} \cdot \mathbf{e_d}\langle s|H|p_\sigma\rangle \\ &= \mathbf{a} \cdot \mathbf{e_d} V_{sp\sigma}.\end{aligned}$$

The s-orbital centered around atom 1 is labeled by $|s_1\rangle$ and the p-orbital localized at atom 2 by $|p_{2,a}\rangle$. The matrix element $V_{sp\pi} = \langle s|H|p_\pi\rangle$ vanishes according to Equation 1.74. To obtain an explicit formula in terms of p_x, p_y, and p_z, let us introduce the directional cosines (d_x, d_y, d_z) along the x-, y-, and z- axes via $\mathbf{d} = |\mathbf{d}|(d_x, d_y, d_z)$. Then we obtain from the previous equation with $\mathbf{a} = \mathbf{e}_i$, where $\mathbf{e}_i$ is the unit vector along the Cartesian axes

$$\begin{pmatrix}\langle s_1|H|p_{2,x}\rangle \\ \langle s_1|H|p_{2,y}\rangle \\ \langle s_1|H|p_{2,z}\rangle\end{pmatrix} = \begin{pmatrix}d_x V_{sp\sigma}(d) \\ d_y V_{sp\sigma}(d) \\ d_z V_{sp\sigma}(d)\end{pmatrix}. \tag{1.75}$$

A similar analysis can be carried out for the matrix elements $\langle p_{1,i}|H|p_{2,j}\rangle$, which leads to [60]

$$\begin{aligned}\langle p_{1,x}|H|p_{2,x}\rangle &= d_x^2 V_{pp\sigma}(d) + (1 - d_x^2)V_{pp\pi}(d), \\ \langle p_{1,x}|H|p_{2,y}\rangle &= d_x d_y V_{pp\sigma}(d) - d_x d_y V_{pp\pi}(d).\end{aligned} \tag{1.76}$$

The other hopping matrix elements can be obtained by cyclical permutation of the coordinates and direction cosines.

Consequently, by constructing the basis orbitals $|\mathbf{R}, \alpha, \nu\rangle$ as linear combinations of the basis states $|\mathbf{R}, \alpha, \{l, m\}\rangle$, the hopping parameters $E_{\nu',\nu}(\mathbf{R} - \mathbf{R}')_{\alpha',\alpha}$ can be described in terms of the parameters $V_{ll'm}$

and the directional cosines of **d** [60]:

$$
\begin{aligned}
E_{ss}(\mathbf{d})_{\alpha',\alpha} &= V_{ss\sigma}(d,\alpha',\alpha),\\
E_{sx}(\mathbf{d})_{\alpha',\alpha} &= d_x V_{sp\sigma}(d,\alpha',\alpha),\\
E_{xx}(\mathbf{d})_{\alpha',\alpha} &= d_x^2 V_{pp\sigma}(d,\alpha',\alpha) + (1-d_x^2)V_{pp\pi}(d,\alpha',\alpha),\\
E_{xy}(\mathbf{d})_{\alpha',\alpha} &= d_x d_y V_{pp\sigma}(d,\alpha',\alpha) - d_x d_y V_{pp\pi}(d,\alpha',\alpha).
\end{aligned}
$$

Note that interchanging the order of the indices l and l' of $V_{ll'm}$ has no effect if the sum of the parities of the two orbitals is even, but changes sign if the sum of the parities is odd [60]. Similar formulas can be derived for each combination of the localized orbitals and are listed, for example, in Ref. [60]. It is also clear from the above considerations that the hopping matrix elements in the two-center approximation depend only on the orbital type and the distance d between the two sites.

Because of the translational invariance of the bulk system, H_0^{bulk} can be divided into sub-blocks, which are diagonal in **k**:

$$H_0^{\text{bulk}}(\mathbf{k}) = \big(\langle \mathbf{k},\alpha',\nu',\sigma'|H_0^{\text{bulk}}|\mathbf{k},\alpha,\nu,\sigma\rangle\big). \tag{1.77}$$

The dimension of this matrix depends on the number of basis states and atoms within one unit cell. From the diagonalization of this matrix, one obtains the energy dispersion $E(\mathbf{k})$ as a function of the ETB parameters. The energy eigenstates $|n\mathbf{k}\rangle$ are related to the eigenvector $(\mathbf{u}_{n\mathbf{k}})_{\alpha,\nu,\sigma} = u_{n,\alpha,\nu,\sigma}(\mathbf{k})$ via

$$|n\mathbf{k}\rangle = \sum_{\alpha,\nu,\sigma} u_{n,\alpha,\nu,\sigma}(\mathbf{k})|\mathbf{k},\alpha,\nu,\sigma\rangle, \tag{1.78}$$

where n denotes the different bands.

Examples for the fitting of TB band structures to DFT band structures are given in Figure 1.5. In Figure 1.5a, a (nearest neighbor) ETB model with one s- and three p-orbitals per atom (sp^3-model) is used to describe the bulk band structure of wurtzite InN. Even with this limited basis set one obtains a very good description of the band structure around $\mathbf{k} = \mathbf{0}$ and also a reasonable description near the Brillouin zone boundaries. The description of the valence band structure and the first few conduction bands over the full first Brillouin zone can be further improved by including more basis states. An example is given in

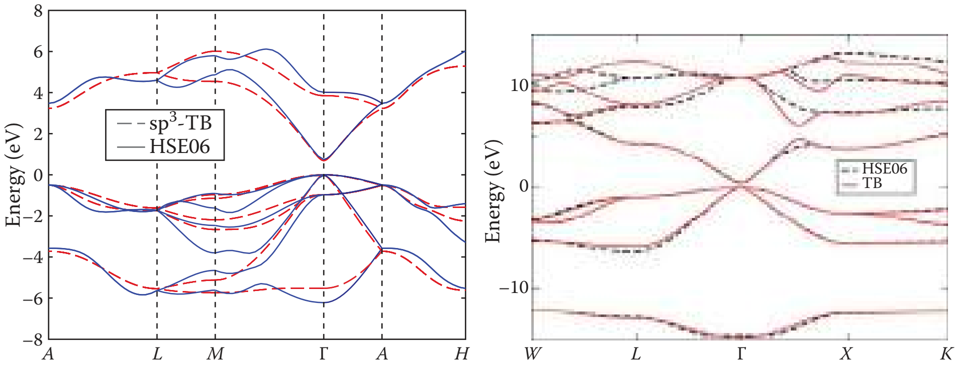


FIGURE 1.5 (a) Electronic band structure of *wurtzite* InN calculated within a nearest neighbor sp^3 TB model and HSE06-DFT. For the HSE06-DFT the cut-off energy was set to 600 eV and Γ-centered $6 \times 6 \times 4$ k-point mesh was used (figure adapted from Ref. [61]). (b) Electronic band structure of *zincblende* InN within a nearest neighbor $sp^3d^5s^*$ TB model and HSE06-DFT. The cut-off energy for the HSE06-DFT was chosen to be 600 eV and a Γ-centered $6 \times 6 \times 6$ k-point mesh was used [62].

Figure 1.5b where an $sp^3d^5s^*$ model is used to describe the band structure of *zincblende* InN. A variety of different ETB models can be found in the literature [59,63–65].

So far, we have neglected the spin of the electron in our discussion. In order to obtain an accurate description of the bulk band structure of semiconductor materials containing heavier atoms such as CdSe and InSb, relativistic effects on the electronic states in crystals have to be considered. In semiconductors containing lighter atoms such as AlP and InN, these effects are negligible for many purposes [66]. This is because the potential is very strong near the nuclei and the kinetic energy is consequently very large, so that the electron velocity is comparable to the velocity of light. Thus, relativistic corrections become more important for heavy elements. One can include these contributions for the motion of an electron in a potential $V(\mathbf{r})$ by considering the *Dirac equation* [67,68]. From a *Foldy–Wouthuysen transformation* of the Dirac equation one obtains relativistic corrections to the time-independent Schrödinger equation. These additional terms are related to (1) relativistic corrections to the kinetic energy, (2) relativistic contributions to the potential $V(\mathbf{r})$, and (3) the spin–orbit interaction [68]. The spin–orbit coupling (SOC) originates from the interaction of the electron spin magnetic moment with the magnetic field "seen" by the electron. The first two components (1) and (2) do not depend on the spin of the electron. Therefore these terms do not change the symmetry properties of the nonrelativistic Hamiltonian. However, the term corresponding to spin–orbit interaction, as we see, couples operators in spin space and ordinary spatial space, thus reducing the symmetry. The effect of the SOC in removing degeneracies of the nonrelativistic results can be deduced using group theory [9].

Therefore, for certain semiconductor materials, it is essential for an accurate bulk band structure description in the framework of an ETB model to include the contributions of the SOC. This issue will be discussed in the following subsection.

1.7.1.1 Spin–Orbit Coupling

In this section, we discuss the inclusion of the SOC in the ETB band structure formalism. Here, the approach of Chadi [69] is employed for the TB scheme. The bulk Hamiltonian of a perfect crystal, including SOC effects can be written as

$$H^{\text{bulk}} = H_0^{\text{bulk}} + H_{\text{so}}, \tag{1.79}$$

where H_0^{bulk} is the spin-independent part and H_{so} denotes the operator for the SOC. Assuming here only s- and p-orbitals, one has to investigate the action of this operator on the basis states of the previous section. The part H_0^{bulk} yields nonzero interactions only between states of the same spin. The matrix elements arising from H_{so} of the Hamiltonian H^{bulk} have the potential to connect states of different spins. To calculate these terms, we assume here that the operator H_{so} acts on the TB basis states in a manner analogous to the atomic spin–orbit operator on atomic orbitals [68]

$$H_{\text{so}}^{\text{atom}} = \frac{1}{2m^2c^2}\frac{1}{r}\frac{\partial V^0}{\partial r}\mathbf{L}\cdot\mathbf{s}. \tag{1.80}$$

Here, V^0 is the atomic potential, $\mathbf{s}$ the spin operator, and $\mathbf{L}$ denotes the angular momentum operator. All matrix elements $\langle p_i\sigma'|H_{\text{so}}|p_j\sigma\rangle$ can be determined in the following way:[†] The operators s_x and s_y can be described by the flip/flop operators $s_\pm$:

$$s_x = \frac{s_+ + s_-}{2},$$
$$s_y = -i\frac{s_+ - s_-}{2}.$$

[†] An s state with $l = 0$ is not split, since the angular part is constant.

The operators s_x, s_y, and s_z then act on the spin states $|\uparrow\rangle$ and $|\downarrow\rangle$ in the following way:

$$\begin{aligned}
s_x|\uparrow\rangle &= \frac{\hbar}{2}|\downarrow\rangle \quad ; \quad s_x|\downarrow\rangle = \frac{\hbar}{2}|\uparrow\rangle; \\
s_y|\uparrow\rangle &= i\frac{\hbar}{2}|\downarrow\rangle \quad ; \quad s_y|\downarrow\rangle = -i\frac{\hbar}{2}|\uparrow\rangle; \\
s_z|\uparrow\rangle &= \frac{\hbar}{2}|\uparrow\rangle \quad ; \quad s_z|\downarrow\rangle = -\frac{\hbar}{2}|\downarrow\rangle.
\end{aligned} \tag{1.81}$$

Furthermore, we have to define the action of the angular momentum operator $\mathbf{L}$ on the states $|p_x\rangle$, $|p_y\rangle$, and $|p_z\rangle$. The states can be written in terms of the *spherical harmonics* Y_{lm} [7]

$$\begin{aligned}
|p_z\rangle &= Y_{10} = \sqrt{\frac{3}{4\pi}}\cos\theta = \sqrt{\frac{3}{4\pi}}\frac{z}{r}, \\
|p_y\rangle &= \frac{i}{\sqrt{2}}(Y_{11} + Y_{1-1}) = \sqrt{\frac{3}{4\pi}}\frac{y}{r}, \\
|p_x\rangle &= \frac{1}{\sqrt{2}}(Y_{1-1} - Y_{11}) = \sqrt{\frac{3}{4\pi}}\frac{x}{r},
\end{aligned} \tag{1.82}$$

which fulfill

$$\begin{aligned}
\mathbf{L}^2 Y_{lm} &= \hbar^2 l(l+1) Y_{lm}, \\
L_z Y_{lm} &= \hbar m Y_{lm}.
\end{aligned} \tag{1.83}$$

The calculation of the matrix elements will exemplarily be demonstrated for the matrix element $\langle p_x \uparrow |H_{\text{so}}|p_y \uparrow\rangle$, for which we obtain

$$\begin{aligned}
\langle p_x \uparrow |H_{\text{so}}|p_y \uparrow\rangle &= \langle p_x \uparrow |CL_x s_x|p_y \uparrow\rangle + \langle p_x \uparrow |CL_y s_y|p_y \uparrow\rangle + \langle p_x \uparrow |CL_z s_z|p_y \uparrow\rangle \\
&= \frac{\hbar}{2}\langle p_x|CL_z|p_y\rangle = \frac{1}{i}\langle p_x|\frac{\hbar^2}{2}C|p_x\rangle,
\end{aligned}$$

with C given by

$$C = \frac{1}{2m^2c^2}\frac{1}{r}\frac{\partial V^0}{\partial r}.$$

Here, we have used Equation 1.81 and that the states $|\uparrow\rangle$ and $|\downarrow\rangle$ are orthogonal to each other to derive the second line from the first. The last line is obtained from Equations 1.82 and 1.83. One can show that all matrix elements can be expressed in terms of

$$\langle p_x|\frac{\hbar^2}{4m^2c^2}\frac{1}{r}\frac{\partial V_{\text{atom}}}{\partial r}|p_x\rangle = \lambda,$$

and can be deduced along the same line. To obtain a compact notation, we denote the states $|\uparrow\rangle$ and $|\downarrow\rangle$ by $|+\rangle$ and $|-\rangle$, respectively. Since the different states $|p_i\pm\rangle$ are orthogonal, many of the matrix elements will be zero. Evaluation of all possible terms gives nonzero results only for

$$\begin{aligned}
\langle p_x \pm |H_{\text{so}}|p_z\mp\rangle &= \pm\lambda, \\
\langle p_x \pm |H_{\text{so}}|p_y\pm\rangle &= \mp i\lambda, \\
\langle p_y \pm |H_{\text{so}}|p_z\mp\rangle &= -i\lambda,
\end{aligned} \tag{1.84}$$

and their complex conjugates. The magnitude of the parameter λ can then be chosen to reproduce the splitting of the valence bands in the vicinity of the Brillouin zone center.

To summarize this section, we have presented a technique that allows us to include the SOC in the ETB formalism. In semiconductor materials with a large spin–orbit splitting at the Brillouin zone center, the inclusion of the SOC is important for a more accurate calculation of the bulk band structure and therefore a more realistic description of the single-particle states in semiconductor nanostructures. However, it should be noted that when including the SOC in the calculations, the size of Hamiltonian matrix is doubled, since one has to explicitly account for the spin. This obviously increases the computational effort.

In addition to ETB models, empirical pseudopotential theory is widely used to study the electronic structure of semiconductor materials, alloys, and nanostructures [49,70]. As mentioned earlier, this approach also accounts for the underlying crystal structure on a microscopic level. The next section describes the fundamental ideas behind this method.

1.7.2 Empirical Pseudopotential Method

Another empirical approach to obtain the electronic structure of semiconductors and related nanostructures on an atomistic level is the *empirical pseudopotential method* (EPM). In a similar fashion as in the ETB approach, the EPM solves the time-independent Schrödinger equation for a bulk system without knowing the exact potential in which the electrons are moving but rather by assuming an effective potential [44,70,71].

The general idea of the EPM is to assume that core electrons are tightly bound to the nuclei and have only a small influence on conduction and valence band states and therefore the electronic properties. Thus, one is left with the problem of removing the contributions of the core electrons from the total wave function. The wave function inside the core region varies rapidly compared to the behavior of the wave function outside this region. Therefore, the idea is to introduce a pseudopotential that leads to a slowly varying wave function in the core region but reproduces the “real” wave function outside the core region. This leads to the task of finding such a pseudopotential. A schematic illustration of the situation is shown in Figure 1.6.

To achieve this, the starting point is again the time-independent Schrödinger equation:

$$\left(-\frac{\hbar^2}{2m}\nabla^2 + V(\mathbf{r})\right)\psi_{\mathbf{k}}(\mathbf{r}) = E(\mathbf{k})\psi_{\mathbf{k}}(\mathbf{r}). \tag{1.85}$$

Due to the periodicity of the bulk crystal structure both the wave function $\psi_{\mathbf{k}}(\mathbf{r})$ and the potential $V(\mathbf{r})$ can be expanded in terms of plane waves:

$$\psi_{\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}\sum_m c_{\mathbf{G}_m} e^{i\mathbf{G}_m\cdot\mathbf{r}}, \tag{1.86}$$

$$V(\mathbf{r}) = \sum_l V_{\mathbf{G}_l} e^{i\mathbf{G}_l\cdot\mathbf{r}}. \tag{1.87}$$

The Fourier coefficients for a given reciprocal lattice vector $\mathbf{G}_n$ are denoted by $c_{\mathbf{G}_n}$ and $V_{\mathbf{G}_n}$, respectively. Using Equations 1.86 and 1.87 the time-independent Schrödinger equation reads

$$\frac{\hbar^2}{2m}\sum_m |\mathbf{k}+\mathbf{G}_m|^2 c_{\mathbf{G}_m} e^{i(\mathbf{k}+\mathbf{G}_m)\cdot\mathbf{r}} + \sum_m\sum_l V_{\mathbf{G}_l} c_{\mathbf{G}_m} e^{i(\mathbf{k}+\mathbf{G}_m+\mathbf{G}_l)\cdot\mathbf{r}} = E(\mathbf{k})\sum_m c_{\mathbf{G}_m} e^{i(\mathbf{k}+\mathbf{G}_m)\cdot\mathbf{r}}. \tag{1.88}$$

Multiplying by $e^{-i(\mathbf{k}+\mathbf{G}_n)\cdot\mathbf{r}}$ and integrating over the volume of the crystal yields

$$\frac{\hbar^2}{2m}\sum_m |\mathbf{k}+\mathbf{G}_m|^2 c_{\mathbf{G}_m}\delta_{n,m} + \sum_m\sum_l V_{\mathbf{G}_l} c_{\mathbf{G}_m}\delta_{l,m-n} = E(\mathbf{k})\sum_m c_{\mathbf{G}_m}\delta_{n,m}, \tag{1.89}$$

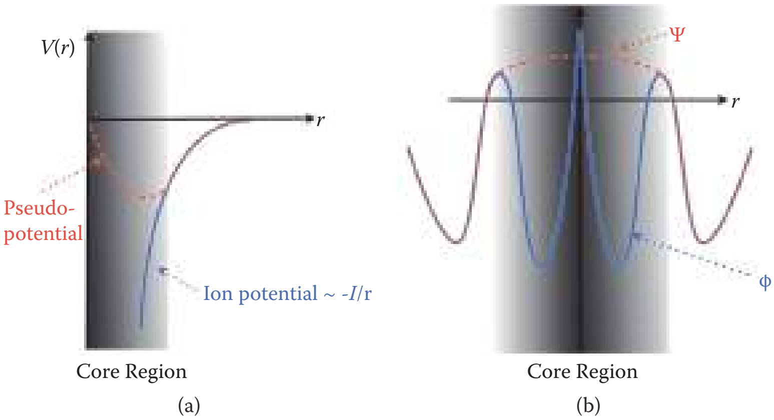


FIGURE 1.6 (a) Schematic illustration of the real ionic potential (solid line) and a pseudopotential (dotted line). (b) Schematic representation of a wave function ϕ in the real ionic potential and the corresponding pseudo-wave function ψ, when considering the pseudopotential in (a).

where $\delta_{\alpha,\beta}$ is the Kronecker delta. Thus, Equation 1.89 becomes

$$\frac{\hbar^2}{2m}|k+\mathbf{G}_n|^2 c_{\mathbf{G}_n} + \sum_m V_{\mathbf{G}_{m-n}} c_{\mathbf{G}_m} = E(\mathbf{k}) c_{\mathbf{G}_n}. \tag{1.90}$$

When truncating the Fourier series $\sum_m V_{\mathbf{G}_{m-n}} c_{\mathbf{G}_m}$ one is left with a finite-size eigenvalue problem. Solving this eigenvalue problem for each **k** point gives the energy dispersion $E(\mathbf{k})$.

However, to be able to solve the eigenvalue problem by diagonalizing the Hamiltonian matrix one still has to determine the potential V. We focus our attention in the following on the zincblende or diamond structure with two atoms per unit cell. Here, following Cohen and Bergstresser [72], we assume that the origin of the coordinate system is halfway between the atoms in the cell. Additionally, the assumption is made that the potential can be split into symmetric and antisymmetric parts [72]:

$$V_{\mathbf{G}_\alpha} = V^S_{\mathbf{G}_\alpha} \cos\left(\mathbf{G}_\alpha \cdot \boldsymbol{\tau}\right) + iV^A_{\mathbf{G}_\alpha} \sin\left(\mathbf{G}_\alpha \cdot \boldsymbol{\tau}\right). \tag{1.91}$$

Here, $\boldsymbol{\tau}$ denotes the position of the atom under consideration with respect to the origin. Since in a zincblende structure we have two atoms per unit cell, we have one atom at $+\boldsymbol{\tau}$ and the second one at $-\boldsymbol{\tau}$, where $\boldsymbol{\tau}$ is given by

$$\boldsymbol{\tau} = \left(\frac{1}{8}, \frac{1}{8}, \frac{1}{8}\right) a. \tag{1.92}$$

Here a denotes the lattice constant of the material. The prefactors $V^S_{\mathbf{G}_\alpha}$ and $V^A_{\mathbf{G}_\alpha}$ in Equation 1.91 are the so-called *form factors*. Qualitatively the form factor $V^S_{\mathbf{G}_\alpha}$ can be understood as the covalent part to the bonding of the atoms, while $V^A_{\mathbf{G}_\alpha}$ describes the ionic contribution [70]. The form factors are usually determined empirically. It can be shown that the pseudopotential $V_{\mathbf{G}_\alpha}$ can be written as the product of the form factors and the *structure factor* [70]. Making use of the fact that the structure factor for a given material is only nonzero at a few of the allowed reciprocal lattice vectors, the number of unknown parameters that has to be determined empirically can be significantly reduced [70].

1.7.2.1 Spin–Orbit Coupling

As described earlier, the spin–orbit interaction plays an important role for an accurate description of the electronic structure of systems containing heavier elements (e.g., CdSe, InSb). In the following, we discuss briefly how to include SOC effects into EPM. The approach described here was introduced by Weisz [73] and later modified by Bloom and Bergstresser [74]. In reciprocal space, one can write the Hamiltonian H_{so} describing the spin–orbit interaction in the following way:

$$H_{\text{so}}^{\mathbf{G},\mathbf{G}'}(\mathbf{k}) = \left(\mathbf{K} \times \mathbf{K}'\right) \cdot \boldsymbol{\sigma} \left[-i\lambda^s \cos\left(\left(\mathbf{G} - \mathbf{G}'\right) \cdot \boldsymbol{\tau}\right) + \lambda^a \sin\left(\left(\mathbf{G} - \mathbf{G}'\right) \cdot \boldsymbol{\tau}\right)\right], \tag{1.93}$$

where $\mathbf{K} = \mathbf{k} + \mathbf{G}$, $\mathbf{K}' = \mathbf{k} + \mathbf{G}'$, and $\boldsymbol{\sigma}$ are the vectors of the Pauli spin matrices. Here λ^s and λ^a are given by

$$\lambda^s = \frac{1}{2}\left(\lambda_1 + \lambda_2\right), \tag{1.94}$$

$$\lambda^a = \frac{1}{2}\left(\lambda_1 - \lambda_2\right) \tag{1.95}$$

with

$$\lambda_1 = \mu B_{1nl}(K) B_{1nl}\left(K'\right), \tag{1.96}$$

$$\lambda_2 = \alpha\mu B_{2nl}(K) B_{2nl}\left(K'\right). \tag{1.97}$$

Here 1 and 2 denote the two atoms in the unit cell, α is the ratio of the free atom spin–orbit splitting energies of the two atoms, $K = |\mathbf{K}|$ and $K' = |\mathbf{K}'|$ and μ is a fitting parameter. The terms B_{nl} are in general given by (the additional index $i = 1, 2$ just denotes the different atoms):

$$B_{nl} = \beta \int_0^\infty j_{nl}(Kr) R_{nl}(r) r^2 \mathrm{d}r. \tag{1.98}$$

Here, $j_{nl}(Kr)$ are the spherical Bessel functions and $R_{nl}(r)$ denotes the radial part of the core function. The indices n and l denote the principal quantum number and the angular moment, respectively. The prefactor β is a normalization factor and can be derived from

$$\lim_{k \to 0} k^{-l} B_{nl}(k) = 1. \tag{1.99}$$

Similar to the ETB approach presented in the previous section, when including the SOC effects into EPM, one is left with one additional fitting parameter, namely μ (cf. Equations 1.96 and 1.97). Even though we add only one additional fitting parameter, the computational expense is increased since the size of the matrix that has to be diagonalized is doubled due to the spin degree that has to be accounted for.

1.8 Summary

In this chapter, we have presented a detailed overview of different, widely used methods to describe the electronic band structure of semiconductor (bulk and nanostructure) materials. The focus here is on crystalline solid-state materials with a periodic lattice. We have seen that this leads to the situation that individual electron wave functions can be chosen to reflect this periodicity, satisfying a condition referred to as Bloch's theorem. Its application considerably simplifies the calculation and description of the electronic structure of crystalline solids.

However, even when using Bloch's theorem, the calculation of the electronic structure of a semiconductor material is in general an extremely complicated task. There are about 10^{23} valence electrons, which contribute to the bonding in each cubic centimeter of a typical solid. This implies that the calculation of the electronic structure is in general a complex many-body problem, as the exact wave function and energy of each electron depends on those of all the others. However, when using the ideas of Hohenberg and Kohn [10], it follows that the ground state energy E_{GS} of the many-particle system is an unambiguous functional of the ground state electron density ρ_{GS}. Thus, the Hohenberg–Kohn theorem allowed us to find the ground state density $\rho_{\mathrm{GS}}(\mathbf{r})$ and energy E_{GS} of an interacting N-particle system by minimizing the energy functional $E[\rho]$. Even though this provides a general framework, it does not give the form of the different functionals, for example, how the kinetic energy depends on ρ. We discussed in a second step the ideas of Kohn and Sham [12], which resulted in an effective single-particle equation in which a particular part of the effective potential, namely the exchange-correlation part, requires approximations. The derived effective single-particle Schrödinger equation is known as the Kohn–Sham equation, forming the foundation of density functional theory (DFT). We reviewed different approximations used in the literature for the exchange-correlation functional. This ranges from the simple so-called local density approximation (LDA) up to hybrid functionals. Ways to go beyond standard DFT that can offer better accuracy, but keeping it still systematic and general, have also been briefly discussed.

These highly accurate first-principles calculations come at the cost of high computational efforts. When interested in the electronic structure of semiconductor nanostructures with dimensions of at least several nanometers, such as quantum wells (QW) and dots, one has to deal with hundreds of thousands of atoms. These cell sizes are not manageable with standard DFT approaches. We have, therefore, discussed semiempirical models, which make use of available experimental or DFT-based data but allow for the description of realistically sized semiconductor nanostructures. These methods, namely $\mathbf{k}\cdot\mathbf{p}$ and empirical tight-binding (TB) and empirical pseudopotential theory, are commonly used to calculate and develop an understanding of the electronic structure of solids. An overview of the general concepts of these three approaches has been provided here. We also considered the main differences between standard $\mathbf{k}\cdot\mathbf{p}$ approaches and TB or empirical pseudopotential theory. While the latter two provide an atomistic treatment of the electronic structure of a semiconductor, $\mathbf{k}\cdot\mathbf{p}$ theory operates on a continuum-based level. This difference can be of critical importance when it comes to the description of the electronic and optical properties of nanostructures. The discussion of the application of the different techniques to nanostructures is beyond the scope of this chapter. Overviews can be found, for example, in Refs. [48,49,75] or in other chapters of this handbook.

References

1. J. Singh. *Electronic and Optoelectronic Properties of Semiconductor Structures*. Cambridge: Cambridge University Press, 2003.
2. D. Bimberg, M. Grundmann and N. N. Ledentsov. *Quantum Dot Heterostructures*. Chichester: Wiley, 2001.
3. G. Czycholl. *Theoretische Festkörperphysik*. Braunschweig/Wiesbaden: Vieweg, 2000.
4. N. W. Ashcroft and D. N. Mermin. *Solid State Physics*. New York: Harcourt College Publishers, 1976.
5. J. Heyd, G. E. Scuseria and M. Ernzerhof. Hybrid functionals based on a screened coulomb potential. *J. Chem. Phys.*, 118:8207, 2003.
6. O. Madelung. *Introduction to Solid-State Theory*. Solid State Sciences Series. Berlin: Springer, 1996.
7. A. I. M. Rae. *Quantum Mechanics*. Bristol: Institute of Physics, 2002.
8. E. P. O'Reilly. *Quantum Theory of Solids*. Masters Series in Physics and Astronomy. London: Taylor & Francis, 2002.
9. H. W. Streitwolf. *Group Theory in Solid-State Physics*. London: Macdonald, 1971.
10. P. Hohenberg and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.*, 136:B864, 1964.
11. H. Englisch and R. Englisch. Hohenberg-Kohn theorem and non-V-representable densities. *Phy. A: Stat. Mech. Its App.*, 121:253, 1983.
12. W. Kohn and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.*, 140:A1133, 1965.
13. R. M. Martin. *Electronic Structure: Basic Theory and Practical Methods*. Cambridge: Cambridge University Press, 2011.

14. J. P. Perdew. Density functional theory and the band gap problem. *Int. J. Quantum Chem.*, 28:497, 1985.
15. A. J. Cohen, P. Mori-Sanchez and W. Yang. Challenges for density functional theory. *Chem. Rev.*, 112:289, 2012.
16. D. J. W. Geldart. Volume 180 of *Topics in Current Chemistry*. Berlin: Springer, 1996.
17. J. P. Perdew, K. Burke and Y. Wang. Generalized gradient approximation for the exchange-correlation hole of a many-electron system. *Phys. Rev. B*, 54:16533, 1996.
18. K. Capelle. A bird's-eye view of density-functional theory. *Braz. J. Phys.*, 36:1318, 2006.
19. J. P. Perdew, K. Burke and M. Ernzerhof. Generalized gradient approximation made simple. *Phys. Rev. Lett.*, 77:3865, 1996.
20. J. P. Perdew and S. Kurth. Density Functionals for Non-Relativistic Coulomb Systems, in *Density Functionals: Theory and Applications*, volume 500 of Lecture Notes in Physics. Berlin: Springer, 1998.
21. L. J. Sham and M. Schlüter. Density-functional theory of the energy gap. *Phys. Rev. Lett.*, 51:1888, 1983.
22. A. Seidl, A. Görling, P. Vogl, J. A. Majewski and M. Levy. Generalized Kohn–Sham schemes and the band-gap problem. *Phys. Rev. B.*, 53:3764, 1996.
23. P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt and M. Scheffler. Exciting prospects for solids: Exact-exchange based functionals meet quasiparticle energy calculation. *Phys. Stat. Sol (b)*, 245:929, 2008.
24. A. D. Becke. A new mixing of Hartree–Fock and local density-functional theories. *J. Chem. Phys.*, 98:1372, 1993.
25. A. D. Becke. Density-functional thermochemistry. III. The role of exact exchange. *J. Chem. Phys.*, 98:5648, 1993.
26. A. D. Becke. Density-functional thermochemistry. IV. A new dynamical correlation functional and implications for exact-exchange mixing. *J. Chem. Phys.*, 104:1040, 1996.
27. J. P. Perdew, M. Ernzerhof and K. Burke. Rationale for mixing exact exchange with density functional approximations. *J. Chem. Phys.*, 105:9982, 1996.
28. J. A. Pople, M. Head-Gordon, D. J. Fox, K. Raghavachari and L. A. Curtiss. Gaussian-1 theory: A general procedure for prediction of molecular energies. *J. Chem. Phys.*, 90:5622, 1989.
29. C. Adamo and V. Barone. Toward reliable density functional methods without adjustable parameters: The PBE0 model. *J. Chem. Phys.*, 110:6158, 1999.
30. T. Bredow and A. R. Gerson. Effect of exchange and correlation on bulk properties of MgO, NiO, and CoO. *Phys. Rev. B*, 61:5194, 2000.
31. J. Muscat, A. Wander and N. Harrison. On the prediction of band gaps from hybrid functional theory. *Chem. Phys. Lett.*, 342:397, 2001.
32. W. Perger. Calculation of band gaps in molecular crystals using hybrid functional theory. *Chem. Phys. Lett.*, 368:319, 2003.
33. M. Schlipf. *Heyd–Scuseria–Ernzerhof Screened-Exchange Hybrid Functional for Complex Materials: All-Electron Implementation and Application*. Schriften des Forschungszentrums Jülich Reihe Schlüssel technologien/Key Technologies. Jülich, Germany: Forschungszentrum Jülich GmbH, RWTH Aachen University, 2012.
34. J. Paier, M. Marsman, K. Hummer, I. C. Gerber G. Kresse and J. G. Angyan. Screened hybrid density functionals applied to solids. *J. Chem. Phys.*, 124:154709, 2006.
35. A. Carvalho, A. Alkauskas, Alfredo Pasquarello, A. K. Tagantsev and N. Setter. A hybrid density functional study of lithium in ZnO: Stability, ionization levels, and diffusion. *Phys. Rev. B*, 80:195205, 2009.
36. A. Carvalho, A. Alkauskas, Alfredo Pasquarello, A. K. Tagantsev and N. Setter. Li-related defects in ZnO: Hybrid functional calculations. *Physica B*, 404:4797, 2009.
37. P. G. Moses, M. Miao, Q. Yan and C. G. Van de Walle. Hybrid functional investigations of band gaps and band alignments for AlN, GaN, InN, and InGaN. *J. Chem. Phys.*, 134:084703, 2011.

38. X. Ren, P. Rinke, V. Blum, J. Wieferink, A. Tkatchenko, A. Sanfilippo, K. Reuter and M. Scheffler. Resolution-of-identity approach to Hartree-Fock, hybrid density functionals, RPA, MP2 and GW with numeric atom-centered orbital basis functions. *New J. Phys.*, 14:053020, 2012.
39. L. Hedin. New method for calculating the one-particle Green's function with application to the electron–gas problem. *Phys. Rev.*, 139:A796, 1965.
40. P. Rinke, A. Qteish, J. Neugebauer, C. Freysoldt and M. Scheffler. Combining GW calculations with exact-exchange density-functional theory: An analysis of valence-band photoemission for compound semiconductors. *New J. Phys.*, 7:126, 2005.
41. L. C. Lew Yan Voon and M. Willatzen. *The* **k·p** *Method: Electronic Properties of Semiconductors*. Heidelberg: Springer, 2009.
42. E. O. Kane. Volume 1 of *Physics of III–V Compounds*. New York: Academic Press, 1966.
43. G. Bastard. *Wave Mechanics Applied to Semiconductor Heterostructures*. Les Editions de Physique. Courtaboeuf: EDP Sciences, 1988.
44. P. Harrison. *Quantum Wells, Wires and Dots: Theoretical and Computational Physics of Semiconductor Nanostructures*. New York: John Wiley & Sons, 2005.
45. J. M. Luttinger and W. Kohn. Motion of electrons and holes in perturbed periodic fields. *Phys. Rev.*, 97:869, 1955.
46. C. R. Pidgeon and R. N. Brown. Interband magneto-absorption and Faraday rotation in InSb. *Phys. Rev.*, 146:575, 1966.
47. S. L. Chuang and C. S. Chang. **k·p** method for strained wurtzite semiconductors. *Phys. Rev. B*, 54:2491, 1996.
48. M. Ehrhardt and T. Koprucki, editors. *Multi-Band Effective Mass Approximations: Advanced Mathematical Models and Numerical Techniques*, volume 94 of Lecture Notes in Computational Science and Engineering. London: Springer, 2014.
49. G. Bester. Electronic excitations in nanostructures: An empirical pseudopotential based approach. *J. Phys. Condens. Matter*, 21:023202, 2009.
50. A. Zunger. On the farsightedness (hyperopia) of the standard **k·p** model. *Phys. Stat. Sol. (a)*, 190:467, 2002.
51. G. Bester and A. Zunger. Cylindrically shaped zinc blende semiconductor quantum dots do not have cylindrical symmetry: Atomistic symmetry, atomic relaxation, and piezoelectric effects. *Phys. Rev. B*, 71:045318, 2005.
52. N. Baer, S. Schulz, P. Gartner, S. Schumacher, G. Czycholl and F. Jahnke. Influence of symmetry and Coulomb-correlation effects on the optical properties of nitride quantum dots. *Phys. Rev. B*, 76:075310, 2007.
53. P. O. Löwdin. On the non-orthogonality problem connected with the use of atomic wave functions in the theory of molecules and crystals. *J. Chem. Phys.*, 18:365, 1950.
54. S. Lee, O. L. Lazarenkova, P. von Allmen, F. Oyafuso and G. Klimeck. Effect of wetting layers on the strain and electronic structure of In As self-assembled quantum dots. *Phys. Rev. B*, 70:125307–1, 2004.
55. W. Sheng, S. J. Cheng and P. Hawrylak. Multiband theory of multi-exciton complexes in self-assembled quantum dots. *Phys. Rev. B*, 71:035316, 2005.
56. S. Schulz, S. Schumacher and G. Czycholl. Tight-binding model for semiconductor quantum dots with a wurtzite crystal structure: From one-particle properties to Coulomb correlations and optical spectra. *Phys. Rev. B*, 73:245327, 2006.
57. M. Zielinski, M. Korkusinski and P. Hawrylak. Atomistic tight-binding theory of multi-exciton complexes in a self-assembled InAs quantum dot. *Phys. Rev. B*, 81:085301, 2010.
58. K. Schuh, S. Barthel, O. Marquardt, T. Hickel, J. Neugebauer, G. Czycholl and F. Jahnke. Strong dipole coupling in nonpolar nitride quantum dots due to Coulomb effects. *Appl. Phys. Lett.*, 100:092103, 2012.

59. P. Vogl, H. P. Hjalmarson and J. D. Dow. A semi-empirical tight-binding theory of the electronic structure of semiconductors. *J. Phys. Chem. Solids*, 44(5):365–378, 1983.
60. J. C. Slater and G. F. Koster. Simplified LCAO method for periodic potential problem. *Phys. Rev.*, 94:1498, 1954.
61. M. A. Caro, S. Schulz and E. P. O'Reilly. Theory of local electric polarization and its relation to internal strain: Impact on the polarization potential and electronic properties of group-III nitrides. *Phys. Rev. B*, 88:214103, 2013.
62. R. Benchamekh, S. Schulz and E. P. O'Reilly. Tight-binding and $\mathbf{k}\cdot\mathbf{p}$ for cubic nitrides. *In Preparation*, 2016.
63. A. Kobayashi, O. F. Sankey, S. M. Volz and J. D. Dow. Semiempirical tight-binding band structures of wurtzite semiconductors: AlN, CdS, CdSe, ZnS, and ZnO. *Phys. Rev. B*, 28:935, 1983.
64. J. M. Jancu, R. Scholz, F. Beltram and F. Bassani. Empirical *spds** tight-binding calculation for cubic semiconductors: General method and material parameters. *Phys. Rev. B*, 57:6493, 1998.
65. S. Schulz and G. Czycholl. Tight-binding model for semiconductor nanostructures. *Phys. Rev. B*, 72:165317, 2005.
66. P. Y. Yu and M. Cardona. *Fundamentals of Semiconductors*. Physics and Material Properties. Berlin: Springer, 1999.
67. P. A. M. Dirac. *Principles of Quantum Mechanics*. International Series of Monographs on Physics. New York: Oxford University Press, 1958.
68. F. Schwabl. *Quantenmechanik für Fortgeschrittene*. Berlin: Springer, 2000.
69. D. J. Chadi. Spin–orbit splitting in crystalline and compositionally disordered semiconductors. *Phys. Rev. B*, 16:790–796, 1977.
70. M. L. Cohen and J. R. Chelikowski. *Electronic Structure and Optical Properties of Semiconductors*, 2nd Edition. Series in Solid-State Sciences 75. Berlin: Springer, 1989.
71. J. C. Phillips and L. Kleinman. New method for calculating wave functions in crystals and molecules. *Phys. Rev.*, 116:287, 1959.
72. M. L. Cohen and T. K. Bergstresser. Band structures and pseudopotential form factors for fourteen semiconductors of the diamond and zinc-blende structures. *Phys. Rev.*, 141:789, 1966.
73. G. Weisz. Band structure and Fermi surface of white tin. *Phys. Rev.*, 149:504, 1966.
74. S. Bloom and T. K. Bergstresser. Band structure of α-Sn, InSb and CdTe including spin–orbit effects. *Solid State Commun.,* 6:465, 1968.
75. A. Di Carlo. Microscopic theory of nanostructured semiconductor devices: Beyond the envelope-function approximation. *Semicond. Sci. Technol.*, 18:R1, 2003.

2

Electron Transport

Francesco Bertazzi
Michele Goano
Giovanni Ghione
Alberto Tibaldi
Pierluigi Debernardi
and
Enrico Bellotti

2.1 Introduction 35
2.2 From the Boltzmann Transport Equation to the Drift-Diffusion Model 36
Direct Solution of the BTE: Monte Carlo Transport Simulation • Hierarchical Approximations of the BTE: The Method of Moments • The Hydrodynamic Model • The Energy Balance Model • The Drift-Diffusion Model
2.3 Quantum-Corrected Drift-Diffusion 50
Capture and Escape Processes • Coupling Carrier Transport Models with Optical Solvers
2.4 Towards a Genuine Quantum Approach 58
Contour-Ordered NEGF • Boundary Self-Energies • Scattering Self-Energies • Conclusions

2.1 Introduction

A theory of carrier transport in highly nanostructured optoelectronic devices cannot be entirely formulated without addressing the coupling of the current-carrying extended states to the localized states of the system, and the interaction of such localized carriers with coherent fields, as obtained from the classical solution of the Maxwell's equations, or with incoherent field fluctuations, if spontaneous emission is of interest. We will illustrate this complex interplay between carrier transport and optical transitions for a specific class of light-emitting devices, but similar considerations also apply to the inverse regime of light detection and photovoltaics, as a consequence of the principle of detailed balance, which links emission of light by radiative recombination to light absorption by generation of electron–hole pairs, leading e.g., to reciprocity relations between the photovoltaic and the electroluminescent properties of solar cells and light-emitting diodes (LEDs) [1]. For illustrative purposes, we will focus on spatially resolved approaches to describe the far-above threshold, possibly multimode, lasing regime of vertical cavity surface emitting lasers (VCSELs). We will not discuss lumped models based on rate equations for spatially integrated carrier and photon densities, being these approaches restricted to the near-threshold dynamic behavior of single-mode lasers.

One viable solution to the simulation of the temporal and three-dimensional (3D) spatial dynamics of state-of-the-art VCSELs is the coupling of a full-wave optical solver to a semiclassical carrier transport model derived from the Boltzmann transport equation (BTE). In this framework, the interaction of the optical wave with the semiconductor active region is considered from two closely related perspectives, namely the electromagnetic wave standpoint, leading to wave absorption and gain, and the semiconductor standpoint, leading to the electron–hole pair generation and recombination rates [2]. In addition to the interaction with the optical field, the localized states in the active region are also coupled to the extended states of the system, which requires complementing the carrier transport model with appropriate quantum corrections. The issue of including nonclassical corrections within classical frameworks has generated quite some confusion, at least in the context of visible LEDs. The ongoing debate over the loss mechanisms in these devices emphasizes the need for a unified treatment of both optical and electronic aspects by means of genuine quantum kinetic approaches, such as the density matrix (DM) and the nonequilibrium Green's function (NEGF) theory.

Given this context, an appraisal of the theoretical underpinnings of carrier transport in nanostructured material for optoelectronic applications seemed to us far more compelling than a comprehensive review of the models, which would be unavoidably incomplete and possibly redundant, considering the vast literature dedicated to the topic. The aim of this chapter is to review the assumptions behind the BTE, to present how a hierarchical set of equations for the moments of the distribution can be derived from it (Section 2.2), and to discuss advantages and limitations of different approaches to go beyond the semiclassical limit, from quantum-corrected semiclassical models (Section 2.3) to quantum kinetic models in which dissipative carrier transport and quantum optics are described on equal footing (Section 2.4). Our hope is to be informative, while at the same time to call attention to some of the tacit assumptions of electronic transport in optoelectronic devices which are still largely untested.

2.2 From the Boltzmann Transport Equation to the Drift-Diffusion Model

The starting point of the semiclassical picture recalled in Section 2.1 is the BTE—the first equation ever written to govern the time evolution of a probability—proposed in the late nineteenth century by Ludwig Boltzmann in his kinetic theory of gases and generalized in different fields of physics [3] including, of course, carrier transport in semiconductors, where it takes the form

$$\frac{\partial f}{\partial t} + \underbrace{\underline{v}\cdot\nabla_{\underline{r}} f}_{\text{position}} + \underbrace{\frac{1}{\hbar}\underline{F}\cdot\nabla_{\underline{k}} f}_{\text{momentum}} = \left.\frac{\mathrm{d}f}{\mathrm{d}t}\right|_{\text{collision}}. \tag{2.1}$$

The relevant probability function here is the distribution $f(\underline{r}, \underline{k}, t)$ that describes the evolution in the phase space $(\underline{r}, \underline{k})$ of carriers treated as point-like particles with well-defined geometrical position $\underline{r}$ and wave vector $\underline{k}$ (corresponding to a momentum $\underline{p} = \hbar\underline{k}$). The symbols $\nabla_{\underline{r}}$ and $\nabla_{\underline{k}}$ denote differentiation in the position and momentum spaces, respectively. The description of the electronic band structure $E(\underline{k})$ is included in the BTE through the velocity term $\underline{v} = \nabla_{\underline{k}} E(\underline{k})/\hbar$, where the band index has been omitted to simplify the notation. The "position" term in Equation 2.1 accounts for the motion of carriers in real space, due to their group velocity $\underline{v}$; the "momentum" term describes the variation of momentum impressed by the classical force $\underline{F}$. The momentum of carriers (but not their position) is also affected by scattering events with photons, phonons, impurities, and other carriers, all included in the collision term on the right-hand side (RHS) of Equation 2.1,

$$\left.\frac{\mathrm{d}f}{\mathrm{d}t}\right|_{\text{collision}} = \sum_{\underline{k}'} \left\{ f(\underline{r}, \underline{k}', t)[1 - f(\underline{r}, \underline{k}, t)]\, W(\underline{k}', \underline{k}) - f(\underline{r}, \underline{k}, t)[1 - f(\underline{r}, \underline{k}', t)]\, W(\underline{k}, \underline{k}') \right\} \mathrm{d}\underline{k}', \tag{2.2}$$

where $W(\underline{k}, \underline{k}')$ is the rate at which a particle scatters from a state $\underline{k}$ to $\underline{k}'$ (the first term represents the flow of particles into state $\underline{k}$, the second is the loss by scattering into another state). Each carrier population (electrons and holes, or possibly their subsets, e.g., heavy/light/split-off band holes) is described by its own distribution function and the corresponding BTEs are coupled together (and to the crystal lattice) through the collision terms.

Heuristic derivations[†] of the BTE by means of conservation principles may be found in [4,5]. In introducing the theory of semiclassical carrier transport, it is important to recall some of the basic hypotheses that we are forced to assume to reach the BTE.

1. The carriers have a well-defined position $\underline{r}$ and momentum $\underline{p}$ as if they were classical particles. For this approximation to be acceptable, the size of the device should be much larger than the size of the electron wavepacket given by the mean free path between collisions and the energy scale of interest should be much larger than the uncertainty implied by the spread of the electron momentum.
2. At any given point in $\underline{k}$-space, the energy bands are well separated in energy and the electric fields are sufficiently small so that we can describe the free flight of the carriers between collisions with Newton's laws.[‡]
3. The energy bands do not change too quickly in space, so that the concept of a local band structure may be accepted and level quantization can be neglected.
4. Boltzmann's Stosszahlansatz holds, i.e., all scattering processes are independent (no memory is conserved of where and when the previous collisions happened), local (they involve no change of $\underline{r}$), instantaneous (their duration is negligible with respect to the free-flight time between successive collisions), and are perturbations weak enough to justify the use of Fermi's golden rule [5, Appendix E] to estimate their probability per unit time.

These assumptions break down in a few important cases: very fast transients that have to be studied on a time scale comparable to the scattering rate [6]; high scattering rates, so that the initial state may decay appreciably by the time the scattering is completed (collision broadening) [6]; and presence of high electric fields, which may transfer a significant amount of energy to the carriers during the finite time required for a collision, or, in other words, renormalize the states involved in the scattering process (intracollisional field effect [7]).

An additional assumption, unrelated to the BTE derivation, is that time-domain transients are slow enough to justify the use of electrostatics rather than Maxwell's equations, meaning that the electric field may be determined only by the instantaneous distributions of the carriers via Poisson's equation, therefore neglecting propagation effects. A carrier transport model coupled to a full-wave electromagnetic solver is presented in [10] in the context of semiconductor waveguides for microwave applications; the coupling of the transport equations with an optical modal solver will be discussed in Section 2.3.2 in the context of VCSELs.

2.2.1 Direct Solution of the BTE: Monte Carlo Transport Simulation

The BTE is an integro-differential equation, nonlinear because the collision term depends on the distribution function $f(\underline{r}, \underline{k}, t)$, itself due to the Fermi statistical factors appearing in Equation 2.2 (that also describe phase-space filling) or, in dealing with Coulomb interactions, via the Lindhart screening formula [6]. Closed-form approximate solutions of the BTE may be obtained under homogeneous conditions and small electric fields, which allow linearizing the equation with respect to the field (relaxation-time approximation). Partial differential equations (PDEs) may be obtained from the BTE by expanding the unknown $f(\underline{r}, \underline{k}, t)$ into suitable basis functions (e.g., spherical harmonics [11]) and then solving for the expansion coefficients (expansion methods).

A direct statistical solution of the BTE for the distribution function is provided by the Monte Carlo method [5,12–14], which consists of the simulation of the motion of an ensemble of carriers, subject to the action of electric and magnetic fields and of given scattering mechanisms. The duration of the carrier free flights, i.e., the time between two successive collisions, and the states after scattering are determined stochastically according to the scattering probabilities.

[†] Note that thus far the BTE has resisted a rigorous derivation. What remains elusive is the problem of deriving macroscopic irreversibility (that defines the "arrow of time") from microscopic reversibility [8].

[‡] In wide band-gap materials, due to the small size of the Brillouin zone, multiple bands exist in relatively narrow ranges of energy, and band-to-band tunneling at band "kissing points" may be of importance [9].

As with any time-domain technique, the Monte Carlo approach is much less efficient, at least in its self-consistent version (when Poisson's equation is solved self-consistently with the charge distribution) if different time constants are involved. This is the case in the simulation of bipolar devices, where intraband and interband (recombination) processes contribute with very different scattering times to determine the profile of the minority carriers. In addition to this limitation, the treatment of minority carriers (whose diffusion determines the current across a *p-n* junction) is problematic within a Monte Carlo framework, requiring statistical enhancement techniques that are quite cumbersome to implement; in fact, most Monte Carlo studies of optoelectronic devices are restricted to non-self-consistent approximations or to unipolar devices [15–18]. This does not imply that Monte Carlo device simulation is not viable in optoelectronics. On the contrary, it is probably the most general approach to the BTE, for its ability to incorporate physical models beyond what other techniques can handle (full-Brillouin-zone descriptions of the electronic structure [19,20], first-principles phonon dispersions [9], wave vector–dependent scattering rates), limited only by the time one is willing to wait for a converged solution. As an example of how much physics can be included in a Monte Carlo code, we mention the pioneering works [21–23] on carrier–phonon and carrier–carrier scattering, with the aim of explaining high-field carrier transport starting from the sole input of the ion pseudopotentials employed for the description of the electronic structure.[†] Another example along these lines is the inclusion of interband tunneling within full-band Monte Carlo codes (by means of the time-dependent solution of the Schrödinger equation) to describe the coherent (between scattering events) drift of the carriers at band crossing points [24].

2.2.2 Hierarchical Approximations of the BTE: The Method of Moments

From an engineering perspective, in several technologically relevant cases a good description of the device behavior can be obtained at a much reduced computational cost by solving a small set of coupled nonlinear PDEs derived from the BTE by applying the *method of moments* (MOM) [25,26]. Contrary to Monte Carlo simulation, MOM does not try to obtain the full distribution function $f(\underline{r}, \underline{k}, t)$, but settles instead for knowledge of some of its ensemble averages, closely related to relevant quantities such as carrier density, velocity, and energy. The derivation of the MOM PDEs from the BTE is to some extent comparable to the definition of the moments of a probability density in statistics (i.e., average, variance, etc.), hence its name. The BTE Equation 2.1 is multiplied by a scalar function $\psi(\underline{k})$ and then integrated over the $\underline{k}$ space, which gives

$$\int_{\underline{k}} \psi \frac{\partial f}{\partial t}\, \mathrm{d}\underline{k} + \int_{\underline{k}} \psi \underline{v} \cdot \nabla_{\underline{r}} f\, \mathrm{d}\underline{k} + \int_{\underline{k}} \psi \frac{1}{\hbar} \underline{F} \cdot \nabla_{\underline{k}} f\, \mathrm{d}\underline{k} = \int_{\underline{k}} \psi \left.\frac{\mathrm{d}f}{\mathrm{d}t}\right|_{\mathrm{collision}} \mathrm{d}\underline{k}.$$

Since ψ and $\underline{v}$ do not depend on time and position, we have

$$\int_{\underline{k}} \psi \frac{\partial f}{\partial t}\, \mathrm{d}\underline{k} = \frac{\partial}{\partial t} \int_{\underline{k}} \psi f\, \mathrm{d}\underline{k}$$

$$\int_{\underline{k}} \psi \underline{v} \cdot \nabla_{\underline{r}} f\, \mathrm{d}\underline{k} = \nabla_{\underline{r}} \cdot \int_{\underline{k}} \underline{v} \psi f\, \mathrm{d}\underline{k}.$$

[†] Wave vector–dependent deformation potentials that rigorously satisfy selection rules along symmetry lines were derived from nonlocal empirical pseudopotential calculations of the electronic structure and density functional calculations of the lattice dynamics [9,23]. Their use shifted the empiricism of earlier Monte Carlo investigations (based on phenomenological deformation potential parameters) to a more fundamental level, with no fitting parameters involved, save for the initial choice for the pseudopotentials. Electron–phonon interactions and impact ionization were never treated in such detail within any other carrier transport code.

Moreover, $\nabla_{\underline{k}}(\psi f) = f\nabla_{\underline{k}}\psi + \psi\nabla_{\underline{k}} f$, so

$$\int_{\underline{k}} \psi \frac{1}{\hbar}\underline{F} \cdot \nabla_{\underline{k}} f \, \mathrm{d}\underline{k} = \frac{1}{\hbar}\underline{F} \cdot \int_{\underline{k}} \psi \nabla_{\underline{k}} f \, \mathrm{d}\underline{k} = \frac{1}{\hbar}\underline{F} \cdot \int_{\underline{k}} \nabla_{\underline{k}}(\psi f) \, \mathrm{d}\underline{k} - \frac{1}{\hbar}\underline{F} \cdot \int_{\underline{k}} f \nabla_{\underline{k}} \psi \, \mathrm{d}\underline{k}.$$

Applying Green's theorem we also have

$$\int_{\underline{k}} \nabla_{\underline{k}}(\psi f) \, \mathrm{d}\underline{k} = \oint_{\Sigma} \hat{\mathrm{n}} \psi f \, \mathrm{d}\sigma,$$

where the surface Σ, with normal vector $\hat{\mathrm{n}}$, has an infinite radius. Assuming that ψ is a polynomial function of $\underline{k}$ (see discussion below) and f is well behaved (i.e., decaying exponentially, Maxwellian-like, for $\underline{k} \to \infty$), the surface integral is zero, and we are left with

$$\frac{\partial}{\partial t} \int_{\underline{k}} \psi f \, \mathrm{d}\underline{k} + \nabla_{\underline{r}} \cdot \int_{\underline{k}} \underline{v} \psi f \, \mathrm{d}\underline{k} - \frac{1}{\hbar}\underline{F} \cdot \int_{\underline{k}} f \nabla_{\underline{k}} \psi \, \mathrm{d}\underline{k} = \int_{\underline{k}} \left.\frac{\mathrm{d}(\psi f)}{\mathrm{d}t}\right|_{\text{collision}} \mathrm{d}\underline{k}. \tag{2.3}$$

By choosing the moment functions ψ_m as integer powers of k, e.g., $\psi_0 = a_0$, $\psi_1 = a_1 k$, $\psi_2 = a_2 k^2$, etc. (where a_0, a_1, a_2 ... are proportionality coefficients), one may formally replace the BTE with a hierarchy of balance PDEs whose unknowns are the mth-order moments

$$M_m = \int_{\underline{k}} \psi_m f \, \mathrm{d}\underline{k}.$$

In principle, the prescription above may be applied to derive an arbitrary number of moments, from which all the relevant information about the distribution function $f(\underline{r}, \underline{k}, t)$ could be recovered. However, most device simulators account for only a few moments connected to well-defined physical quantities, such as the carrier density ($m = 0$), the average momentum ($m = 1$), and the average energy ($m = 2$). Higher-order moments are usually neglected as they lack a simple physical interpretation (save for the third-order moment, which is related to the heat flow).

In truncating the infinite hierarchy of PDEs, one has to face the so-called "closure problem," since the balance equation for any given moment also includes the next higher-order moment [27]. With the help of approximate closure relations, the PDE system can be conveniently reduced to the three lower-order equations, expressing the conservation of particle density, average momentum, and average energy. Owing to the similarities between carrier transport and fluid dynamics, the resulting system is commonly referred to as the HD model [28,29]. Further simplifications lead to the so-called energy balance (EB) model [30,31], which, in turn, can be reduced to the familiar drift-diffusion (DD) picture of carrier transport, assuming local instantaneous equilibrium between the average energy and the electric field. HD models can describe nonstationary effects such as velocity overshoot, carrier heating, and negative differential mobility, but they require additional information concerning the energy and momentum relaxation times, which can be provided by Monte Carlo simulations [32,33]. On the other hand, the DD approximation, complemented by appropriate recombination and mobility models, has proved remarkably successful in explaining the electrical behavior of devices that violate most of the assumptions under which it should be applied (small electric fields, large device size with respect to the mean free path).

In the rest of this section, unless otherwise stated, only the electron population will be considered, but the same discussion can be replicated for holes with nearly identical results.

2.2.2.1 Zeroth-Order Moment: Carrier Conservation

Setting $\psi(\underline{k}) = 1$, one has

$$M_0 = \int_{\underline{k}} f(\underline{r}, \underline{k}, t) \, \mathrm{d}\underline{k} = n(\underline{r}, t),$$

and Equation 2.3 becomes

$$\frac{\partial}{\partial t}\int_{\underline{k}} f(\underline{r},\underline{k},t)\,\mathrm{d}\underline{k} + \nabla_{\underline{r}}\cdot\int_{\underline{k}} \underline{v}(\underline{k})f(\underline{r},\underline{k},t)\,\mathrm{d}\underline{k} = \int_{\underline{k}} \left.\frac{\mathrm{d}f(\underline{r},\underline{k},t)}{\mathrm{d}t}\right|_{\text{collision}},$$

which is a balance equation for the electron density n

$$\frac{\partial n}{\partial t} + \nabla_{\underline{r}}\cdot(n\underline{v}_n) = \left.\frac{\mathrm{d}n}{\mathrm{d}t}\right|_{\text{collision}},$$

where the collision term includes GR mechanisms coupling the electron and hole populations, and $\underline{v}_n$ is the average electron velocity

$$\underline{v}_n = \frac{\int_{\underline{k}} \underline{v}(\underline{k})f(\underline{r},\underline{k},t)\,\mathrm{d}\underline{k}}{\int_{\underline{k}} f(\underline{r},\underline{k},t)\,\mathrm{d}\underline{k}},$$

closely related to the first-order moment of the electron distribution.

2.2.2.2 First-Order Moment: Momentum Conservation

Setting $\psi(\underline{k}) = \hbar\underline{k}$, which is the carrier momentum, and taking into account that the average momentum of the electron distribution is

$$\underline{p}_n = \frac{\int_{\underline{k}} \hbar\underline{k}f(\underline{r},\underline{k},t)\mathrm{d}\underline{k}}{\int_{\underline{k}} f(\underline{r},\underline{k},t)\mathrm{d}\underline{k}},$$

one has in Equation 2.3

$$\frac{\partial}{\partial t}\int_{\underline{k}} \psi f\,\mathrm{d}\underline{k} = \frac{\partial}{\partial t}\int_{\underline{k}} \hbar\underline{k}f\,\mathrm{d}\underline{k} = \frac{\partial(n\underline{p}_n)}{\partial t}$$

and

$$\nabla_{\underline{r}}\cdot\int_{\underline{k}} \psi\underline{v}f\mathrm{d}\underline{k} = \nabla_{\underline{r}}\cdot\int_{\underline{k}} \hbar\underline{k}\,\underline{v}f\mathrm{d}\underline{k} = \frac{\hbar^2}{m_n^*}\nabla_{\underline{r}}\cdot\int_{\underline{k}} \underline{k}\,\underline{k}f\mathrm{d}\underline{k},$$

where it has been assumed (*parabolic approximation*) that the electron effective mass m_n^* is independent of the electron energy. By defining the average wave vector $\underline{k}_n = \underline{p}_n/\hbar$, one has

$$\underline{k}\underline{k} = (\underline{k}-\underline{k}_n)(\underline{k}-\underline{k}_n) + 2\underline{k}\,\underline{k}_n - \underline{k}_n\underline{k}_n,$$

hence

$$\begin{aligned}\int_{\underline{k}} \underline{k}\,\underline{k}f\mathrm{d}\underline{k} &= \int_{\underline{k}} (\underline{k}-\underline{k}_n)(\underline{k}-\underline{k}_n)f\mathrm{d}\underline{k} + 2\underline{k}_n\int_{\underline{k}} \underline{k}f\mathrm{d}\underline{k} - \underline{k}_n\underline{k}_n\int_{\underline{k}} f\mathrm{d}\underline{k} = \\ &= \int_{\underline{k}} (\underline{k}-\underline{k}_n)(\underline{k}-\underline{k}_n)f\mathrm{d}\underline{k} + n\underline{k}_n\underline{k}_n.\end{aligned}$$

The dyad

$$\int_{\underline{k}} (\underline{k} - \underline{k}_n)(\underline{k} - \underline{k}_n) f \, \mathrm{d}\underline{k}$$

is proportional to the *covariance matrix* $\underline{C}[\underline{k}]$ of the $\underline{k}$ components taken in pairs,

$$\underline{C}[\underline{k}] = \frac{\int_{\underline{k}} (\underline{k} - \underline{k}_n)(\underline{k} - \underline{k}_n) f \, \mathrm{d}\underline{k}}{\int_{\underline{k}} f \, \mathrm{d}\underline{k}}.$$

Considering, e.g., a drifted Gaussian distribution (see Figure 2.1), symmetric with respect to its central value (the average velocity), the off-diagonal terms of the covariance matrix are identically zero (integrals of odd functions over symmetric intervals), while the diagonal terms include the variances of the distribution with respect to each spatial coordinate. If the variance of the distribution is the same along all directions, $\underline{C}$ is proportional to the identity matrix $\underline{I}$ through a coefficient that can be defined in terms of the *temperature* of the distribution, usually called the *electron temperature* T_n

$$m_n^* k_{\mathrm{B}} T_n \underline{I} = \hbar^2 \underline{C}[\underline{k}] = \underline{C}[\underline{p}], \tag{2.4}$$

where $k_{\mathrm{B}} T_n$ is proportional to the average energy of the electron gas due to the *disordered* electron motion. In the general case, it is possible to define an anisotropic *temperature tensor*

$$\mathbf{T}_n = \frac{\hbar^2}{m_n^* k_{\mathrm{B}}} \underline{C}[\underline{k}],$$

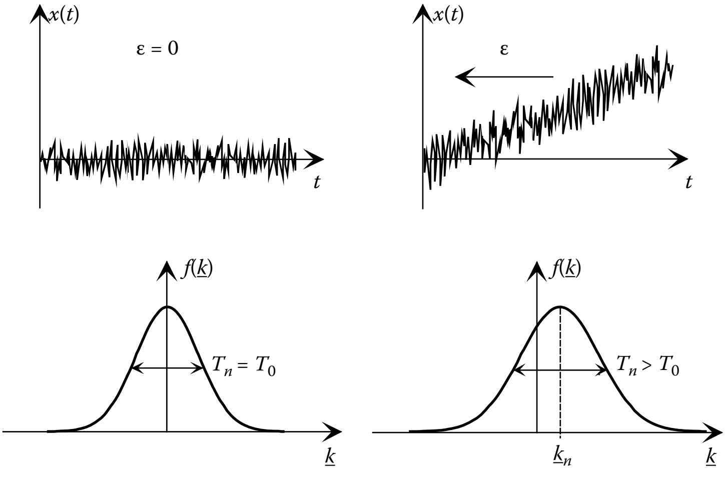


FIGURE 2.1 Drift motion and heating of an electron gas determined by an electric field $\mathcal{E}$. The carrier momentum distribution is Gaussian (under the parabolic approximation) and symmetric with respect to its central value (average momentum $\underline{k}_n$), which is zero at thermal equilibrium and shifts under the application of $\mathcal{E}$. The temperature of the electron gas T_n is proportional to the *variance* of the distribution, distribution (see Equation 2.4), which increases with $\mathcal{E}$.

but it is common to use a diagonal approximation for $\mathbf{T}_n$, so that

$$\frac{\hbar^2}{m_n^*}\int_{\underline{k}} \underline{k}\,\underline{k} f\,\mathrm{d}\underline{k} = \frac{\hbar^2}{m_n^*}\int_{\underline{k}} (\underline{k}-\underline{k}_n)(\underline{k}-\underline{k}_n) f\,\mathrm{d}\underline{k} + \frac{\hbar^2 n}{m_n^*}\underline{k}_n\underline{k}_n = n k_{\mathrm{B}} T_n \underline{I} + \frac{1}{m_n^*} n \underline{p}_n \underline{p}_n. \tag{2.5}$$

Therefore, one has

$$\nabla_{\underline{r}} \cdot \int_{\underline{k}} \hbar \underline{k}\,\underline{v} f\,\mathrm{d}\underline{k} = \nabla_{\underline{r}} \cdot \left(n k_{\mathrm{B}} T_n \underline{I} + \frac{1}{m_n^*} n \underline{p}_n \underline{p}_n \right) = \nabla_{\underline{r}}(n k_{\mathrm{B}} T_n) + \frac{1}{m_n^*} \nabla_{\underline{r}} \cdot (n \underline{p}_n \underline{p}_n).$$

Also,

$$\int_{\underline{k}} (\nabla_{\underline{k}} \psi) f\,\mathrm{d}\underline{k} = \hbar \int_{\underline{k}} (\nabla_{\underline{k}} \underline{k}) f\,\mathrm{d}\underline{k} = \hbar \underline{I} \int_{\underline{k}} f\,\mathrm{d}\underline{k} = \hbar \underline{I} n.$$

Assuming that the only relevant forces are due to the electric field $\underline{\mathcal{E}}$, the previous equation becomes

$$\frac{\underline{F}}{\hbar} \cdot \int_{\underline{k}} \nabla_{\underline{k}} \underline{k} f\,\mathrm{d}\underline{k} = \underline{F} n = -\mathrm{q} n \underline{\mathcal{E}},$$

where q is the elementary charge. Finally, assembling all the previous partial results and introducing a phenomenological collision term, one has

$$\frac{\partial (n \underline{p}_n)}{\partial t} + \nabla_{\underline{r}}(n k_{\mathrm{B}} T_n) + \frac{1}{m_n^*} \nabla_{\underline{r}} \cdot (n \underline{p}_n \underline{p}_n) + \mathrm{q} n \underline{\mathcal{E}} = \left. \frac{\mathrm{d}(n \underline{p}_n)}{\mathrm{d}t} \right|_{\text{collision}}, \tag{2.6}$$

which is called the *momentum transport equation*, and includes a term (the electron temperature) closely related to the second-order moment (the variance) of the electron distribution.

2.2.2.3 Second-Order Moment: Energy Conservation

Setting $\psi(\underline{k}) = \dfrac{(\hbar k)^2}{2m_n^*} = E_n$, which is the electron kinetic energy under the parabolic approximation, and taking into account that the average energy of the electron distribution is

$$\overline{E}_n = \frac{\displaystyle\int_{\underline{k}} E_n f(\underline{r}, \underline{k}, t)\,\mathrm{d}\underline{k}}{\displaystyle\int_{\underline{k}} f(\underline{r}, \underline{k}, t)\,\mathrm{d}\underline{k}},$$

the second-order moment of the electron distribution is

$$\int_{\underline{k}} \frac{(\hbar k)^2}{2m_n^*} f(\underline{r}, \underline{k}, t)\,\mathrm{d}\underline{k} = n \overline{E}_n.$$

Therefore, the first term in Equation 2.3 becomes

$$\frac{\partial}{\partial t}\int_{\underline{k}} \psi f\,\mathrm{d}\underline{k} = \frac{\partial}{\partial t}\int_{\underline{k}} \frac{(\hbar k)^2}{2m_n^*} f\,\mathrm{d}\underline{k} = \frac{\partial (n \overline{E}_n)}{\partial t}.$$

The second term in Equation 2.3,

$$\nabla_{\underline{r}} \cdot \int_{\underline{k}} \psi \underline{v} f \mathrm{d}\underline{k} = \nabla_{\underline{r}} \cdot \int_{\underline{k}} \frac{(\hbar k)^2}{2m_n^*} \underline{v} f \mathrm{d}\underline{k} = \nabla_{\underline{r}} \cdot \underline{S}_n,$$

now represents the divergence of the energy flow of the electron gas. In terms of $\underline{k}$, $\underline{S}_n$ can be written as

$$\underline{S}_n = \frac{\hbar^3}{2(m_n^*)^2} \int_{\underline{k}} k^2 \underline{k} f \mathrm{d}\underline{k},$$

and it can be demonstrated that

$$\underline{S}_n = \frac{1}{m_n^*} n \underline{p}_n \overline{E}_n + \frac{1}{m_n^*} n \underline{p}_n k_{\mathrm{B}} T_n + \underline{Q}, \tag{2.7}$$

where the vector $\underline{Q}$,

$$\underline{Q} = \frac{\hbar^3}{2(m_n^*)^2} \int_{\underline{k}} (\underline{k} - \underline{k}_n) \cdot (\underline{k} - \underline{k}_n)(\underline{k} - \underline{k}_n) f \mathrm{d}\underline{k},$$

is the *heat flow* of the electron gas and is related to the third-order moment of the distribution. (The heat flow is identically zero when the distribution is symmetric with respect to its central value, as for example in the drifted Gaussian case.)

A closure relation for the system of three PDEs corresponding to the $m = (0, 1, 2)$ moments of the BTE is usually provided by a phenomenological approximation of the heat flow, which can be assumed proportional to the gradient of the electron temperature

$$\underline{Q} \approx -\kappa_n \nabla_{\underline{r}} T_n, \tag{2.8}$$

where the proportionality constant κ_n is the *thermal conductivity of the electron gas*. Near equilibrium, κ_n can be estimated with the *Wiedemann–Franz law*,

$$\kappa_n \approx r \left(\frac{k_{\mathrm{B}}}{\mathrm{q}}\right)^2 T_n \sigma_n,$$

where r is a unitless coefficient (close to 1), $\sigma_n = \mathrm{q} n \mu_n$ is the electrical conductivity, and μ_n is the electron mobility (see Section 2.2.4).

2.2.2.4 The Collision Terms

Within the relaxation-time approximation, the collision term of the BTE is approximated as

$$\left.\frac{\mathrm{d}f}{\mathrm{d}t}\right|_{\text{collision}} \approx \frac{f - f_0}{\tau},$$

where f_0 is the distribution at thermal equilibrium and τ is the *relaxation time* of the distribution, which is assumed to be either a constant or a function of some parameters defining the distribution (average, variance, etc.). A similar approach is usually adopted in all the PDEs derived from the BTE with the MOM. In the balance equation for the electron density (electron continuity equation), the collision term is the net recombination rate (see Section 2.2.5 for a more detailed discussion):

$$\left.\frac{\mathrm{d}n}{\mathrm{d}t}\right|_{\text{collision}} = -U_n.$$

The collision term in the momentum transport equation can be defined through a *momentum relaxation time*, possibly dependent on the average carrier energy (or temperature):

$$\left.\frac{\mathrm{d}\underline{p}_n}{\mathrm{d}t}\right|_{\mathrm{coll}} \approx -\frac{\underline{p}_n}{\tau_{pn}}.$$

Finally, in the energy transport equation, the collision term is approximated by defining an *energy relaxation time*:

$$\left.\frac{\mathrm{d}\overline{E}_n}{\mathrm{d}t}\right|_{\mathrm{collision}} \approx -\frac{\overline{E}_n - \overline{E}_0}{\tau_{En}}.$$

The application of the relaxation time approximation is straightforward, also in the formulation of the momentum and energy transport equations where the collision terms involve the products $n\underline{p}_n$ and $n\overline{E}_n$:

$$\left.\frac{\mathrm{d}(n\overline{E}_n)}{\mathrm{d}t}\right|_{\mathrm{collision}} = n\left.\frac{\mathrm{d}\overline{E}_n}{\mathrm{d}t}\right|_{\mathrm{collision}} + \overline{E}_n\left.\frac{\mathrm{d}n}{\mathrm{d}t}\right|_{\mathrm{collision}} = -n\frac{\overline{E}_n - \overline{E}_0}{\tau_{En}} - \overline{E}_n U_n.$$

2.2.3 The Hydrodynamic Model

The HD model includes the transport PDEs of energy, momentum, and density for electrons and holes—or possibly for one carrier species only, the other being approximated with a simpler description. Under the relaxation time approximation, the HD transport model for electrons is

$$\frac{\partial n}{\partial t} + \nabla\cdot(n\underline{v}_n) = -U_n \tag{2.9}$$

$$\frac{\partial(n\underline{p}_n)}{\partial t} + \nabla(nk_{\mathrm{B}}T_n) + \frac{1}{m_n^*}\nabla\cdot(n\underline{p}_n\underline{p}_n) + \mathrm{q}n\underline{\mathcal{E}} = -n\frac{\underline{p}_n}{\tau_{pn}} - \underline{p}_n U_n \tag{2.10}$$

$$\frac{\partial(n\overline{E}_n)}{\partial t} + \nabla\cdot\underline{S}_n = -\frac{1}{m_n^*}\mathrm{q}n\underline{p}_n\cdot\underline{\mathcal{E}} - n\frac{\overline{E}_n - \overline{E}_0}{\tau_{En}} - \overline{E}_n U_n, \tag{2.11}$$

where, from now on, $\nabla \equiv \nabla_{\underline{r}}$, the energy flow $\underline{S}_n$ is reported in Equation 2.7, and the relationship between average energy and average temperature of the electron gas can be written as

$$\overline{E}_n = \frac{3}{2}k_{\mathrm{B}}T_n + \frac{1}{2m_n^*}p_n^2. \tag{2.12}$$

An alternative, equivalent form of Equations 2.10 and 2.11 can be derived by using the continuity equation and the vector identity $\nabla\cdot(\underline{a}\,\underline{b}) = (\nabla\cdot\underline{a})\underline{b} + (\underline{a}\cdot\nabla)\underline{b}$:

$$\frac{\partial\underline{p}_n}{\partial t} + \frac{1}{n}\nabla(nk_{\mathrm{B}}T_n) + (\underline{v}_n\cdot\nabla)\underline{p}_n + \mathrm{q}\underline{\mathcal{E}} = -\frac{\underline{p}_n}{\tau_{pn}} \tag{2.13}$$

$$\frac{\partial\overline{E}_n}{\partial t} + \underline{v}_n\cdot\nabla\overline{E}_n + \frac{1}{n}\nabla\cdot\left(n\underline{v}_n k_{\mathrm{B}}T_n\right) + \frac{1}{n}\nabla\cdot\underline{Q}_n = -\mathrm{q}\underline{v}_n\cdot\underline{\mathcal{E}} - \frac{\overline{E}_n - \overline{E}_0}{\tau_{En}}. \tag{2.14}$$

In the HD picture, Equation 2.9 enforces carrier density conservation, whereas Equations 2.10 and 2.11 or Equations 2.13 and 2.14 impose conservation of carrier momentum and energy, respectively. The heat flow $\underline{Q}_n$, involved in the closure relation, can either be neglected or described with Equation 2.8.

2.2.4 The Energy Balance Model

The relaxation times and characteristic lengths associated with the momentum and energy transport PDEs Equations 2.10 and 2.11 or Equations 2.13 and 2.14 are often small enough to suggest the possibility of reducing the complexity of the HD model. As a first step, the momentum transport equation can be approximated to make carrier momentum a dependent variable, leading to the energy balance (or energy transport) model.

The derivation of the EB model may start by assuming that the transport phenomena under study are slow enough to allow for a quasistatic analysis of Equation 2.13, so that the time derivative can be neglected, leading to

$$\left(1 + \tau_{pn}\underline{v}_n \cdot \nabla\right)\underline{v}_n \approx -\frac{q\tau_{pn}}{m_n^*}\underline{\mathcal{E}} - \frac{\tau_{pn}}{m_n^* n}\nabla(nk_B T_n), \tag{2.15}$$

where the *convection term* $\tau_{pn}\underline{v}_n \cdot \nabla\underline{v}_n$ is potentially much smaller than $\underline{v}_n$. Focusing on a spatially one-dimensional (1D) case for the sake of clarity, this condition holds if

$$\frac{dv_n}{dx} \ll \frac{1}{\tau_{pn}}.$$

Here, τ_{pn} is shorter than 500 fs [32,34,35] and the maximum Δv_n is approximately equal to the saturation velocity $v_{sat} \approx 10^7$ cm/s,[†] so that the condition becomes

$$\frac{dv_n}{dx} \approx \frac{\Delta v_n}{\Delta x} \ll \frac{1}{\tau_{pn}}.$$

Therefore, the convection term can be non-negligible only for devices with spatial extension Δx smaller than $\tau_{pn}\Delta v_n \approx 50$ nm.[‡] By neglecting this term in Equation 2.15 one obtains

$$\underline{v}_n \approx -\frac{q\tau_{pn}}{m_n^*}\underline{\mathcal{E}} - \frac{\tau_{pn}}{m_n^* n}\nabla(nk_B T_n) = -\mu_n(\overline{E}_n)\underline{\mathcal{E}} - \frac{\mu_n(\overline{E}_n)}{qn}\nabla(nk_B T_n), \tag{2.16}$$

where the electron mobility μ_n has been introduced:

$$\mu_n(\overline{E}_n) = \frac{q\tau_{pn}(\overline{E}_n)}{m_n^*}.$$

Since both the effective mass (if one takes into account nonparabolicity effects) and especially the momentum relaxation time depend on the average energy of the distribution, the electron mobility also is a function of $\overline{E}_n$. According to Equation 2.16, the average velocity is the superposition of two effects, like in the DD model (see Section 2.2.5): the drift-like term involves an energy-dependent (rather than field-dependent) mobility, while the diffusion-like term depends on the gradients of carrier density and temperature (rather than of carrier density only).

[†] In nonstationary conditions, the average drift velocity can be larger than v_{sat} (velocity overshoot); however, under realistic conditions, the peak velocity is of the same order of magnitude of v_{sat}.

[‡] This upper bound is material dependent, and in general the related convection term may be negligible even for smaller extensions.

When substituted in Equations 2.9 and 2.14, the two surviving equations of the HD model, Equation 2.16 can be used to remove the average velocity from the independent variables, leading to the EB model,

$$\frac{\partial n}{\partial t} + \nabla \cdot \left(n\underline{v}_n\right) = -U_n \tag{2.17}$$

$$\frac{\partial \overline{E}_n}{\partial t} + \underline{v}_n \cdot \nabla \overline{E}_n + \frac{1}{n}\nabla \cdot \left(n\underline{v}_n k_\mathrm{B} T_n\right) + \frac{1}{n}\nabla \cdot \underline{Q}_n = -\mathrm{q}\underline{v}_n \cdot \underline{\mathcal{E}} - \frac{\overline{E}_n - \overline{E}_0}{\tau_{En}}, \tag{2.18}$$

whose closure relation is a heat flux constitutive equation (such as the Fourier law). The distinctive feature of EB is that the quantities defining the transport properties (in particular, the mobility) depend on the local average energy rather than on the local field; therefore, EB takes into account the effects of carrier heating caused by the dynamic variations of the electric field.

Another simplification can be introduced by noting that, in the relationship between average energy and temperature of the electron gas reported in Equation 2.12

$$\overline{E}_n = \frac{3}{2}k_\mathrm{B}T_n + \frac{1}{2m_n^*}p_n^2 = \frac{3}{2}k_\mathrm{B}T_n + \frac{1}{2}m_n^* v_n^2,$$

a carrier population having drift velocity close (in the worst case scenario) to v_sat exhibits a kinetic component ($\approx$ 10 meV) lower than its thermal one ($\approx$ 40 meV at 300 K). For this reason, the kinetic contribution can be neglected,

$$\overline{E}_n \approx \frac{3}{2}k_\mathrm{B}T_n,$$

and mobility is written as a function of the electron temperature.

2.2.4.1 From the Energy Balance to the Drift-Diffusion Model

Since the energy relaxation time is of the order of a picosecond, the variations of the average energy in space and time are often negligible, allowing to reduce Equation 2.18 (the energy transport equation) to its stationary form

$$\frac{\overline{E}_n - \overline{E}_0}{\tau_{En}} = -\mathrm{q}\underline{v}_n \cdot \underline{\mathcal{E}},$$

which can be used to write the average energy as

$$\overline{E}_n = \overline{E}_0 - \mathrm{q}\tau_{En}\underline{v}_n \cdot \underline{\mathcal{E}}. \tag{2.19}$$

From Equation 2.16, $\underline{v}_n$ includes a diffusion component depending on the carrier density gradient. Assuming that such contribution is negligible,[†] one has

$$\underline{v}_n \approx -\mu_n \underline{\mathcal{E}},$$

and Equation 2.19 becomes

$$\overline{E}_n \approx \overline{E}_0 + \mathrm{q}\tau_{En}(\overline{E}_n)\mu_n(\overline{E}_n)\mathcal{E}^2. \tag{2.20}$$

[†] This approximation may not be appropriate whenever significant drift and diffusion components coexist, e.g., in the depletion regions of *p*-*n* junctions.

This local (implicit) relationship between $\overline{E}_n$ and $\mathcal{E}$ allow us to write the electron mobility as a function of the electric field; correspondingly, the current density is approximated as

$$\begin{aligned}\underline{J}_n &= -\mathrm{q}n\underline{v}_n = \\ &= \mathrm{q}n\mu_n(\mathcal{E})\underline{\mathcal{E}} + \mu_n(\mathcal{E})\nabla(nk_\mathrm{B}T_n(\mathcal{E})) \\ &= \mathrm{q}n\mu_n(\mathcal{E})\underline{\mathcal{E}} + \mathrm{q}\mu_n(\mathcal{E})\frac{k_\mathrm{B}T_n}{\mathrm{q}}\nabla n + \mu_n(\mathcal{E})n\nabla(k_\mathrm{B}T_n(\mathcal{E})).\end{aligned}$$

By defining the electron diffusivity through a generalized Einstein relation involving T_n,

$$D_n(\mathcal{E}) = \mu_n(\mathcal{E})\frac{k_\mathrm{B}T_n}{\mathrm{q}},$$

the current can be written as

$$\underline{J}_n = \mathrm{q}n\mu_n(\mathcal{E})\left(\underline{\mathcal{E}} + \nabla\frac{k_\mathrm{B}T_n(\mathcal{E})}{\mathrm{q}}\right) + \mathrm{q}D_n\nabla n.$$

In this expression, the temperature gradient contribution can be assimilated to an equivalent electric field. However, this term is to some extent inconsistent with the approximations applied so far to the energy transport equation, where temperature gradients have been neglected. Therefore, the spatial variation of T_n is neglected as well,[†] and the classical expression of the drift-diffusion (DD) current density is finally obtained:

$$\underline{J}_n \approx \mathrm{q}n\mu_n(\mathcal{E})\underline{\mathcal{E}} + \mathrm{q}D_n\nabla n.$$

2.2.5 The Drift-Diffusion Model

With the drastic simplifications described in Section 2.2.4 and above, we arrive at the bipolar drift-diffusion model, which includes Poisson's equation and the electron and hole continuity equations,

$$\begin{cases} -\nabla^2\phi = \dfrac{\mathrm{q}}{\varepsilon}\left(p - n + N_\mathrm{D}^+ - N_\mathrm{A}^-\right) \\ \dfrac{\partial n}{\partial t} = \dfrac{1}{\mathrm{q}}\nabla\cdot\underline{J}_n - U_n \\ \dfrac{\partial p}{\partial t} = -\dfrac{1}{\mathrm{q}}\nabla\cdot\underline{J}_p - U_p \end{cases}, \tag{2.21}$$

with the following expressions for the drift and diffusion current densities:

$$\begin{aligned}\underline{J}_n &= +\mathrm{q}D_n\nabla n - \mathrm{q}\mu_n n\nabla[\phi + \tilde{\phi}_n] \\ \underline{J}_p &= -\underbrace{\mathrm{q}D_p\nabla p}_{\text{diffusion}} - \underbrace{\mathrm{q}\mu_p p\nabla[\phi + \tilde{\phi}_p]}_{\text{drift}}.\end{aligned} \tag{2.22}$$

[†] This last approximation implies that carriers are in equilibrium with the lattice, i.e., the carrier temperature T_n is approximately equal to the lattice temperature T_0. Whenever T_0 varies in space due to heating effects, the DD model is not strictly isothermal and the temperature gradient contribution should be included, leading to the so-called *thermoelectric model* [36].

Here, ϕ is the electrostatic potential, $\varepsilon = \varepsilon_0\varepsilon_r$ is the dielectric permittivity, n and p are the electron and hole densities, $U_{n,p}$ is the net recombination rates per unit time and volume, $D_{n,p}$, $\mu_{n,p}$ are the field-dependent diffusivities and mobilities, and N_D^+, N_A^- are the ionized donor and acceptor densities, which differ from the total densities N_D, N_A if the ionization of the impurities is incomplete [37]. The spatially varying terms $\tilde{\phi}_{n,p}$ (in Equation 2.22)

$$\begin{aligned}\tilde{\phi}_n &= \chi/q + V_T \ln N_C + V_T \ln \gamma_n \\ \tilde{\phi}_p &= E_g/q + \chi/q - V_T \ln N_V - V_T \ln \gamma_p,\end{aligned} \tag{2.23}$$

provide a natural treatment of heterointerfaces, i.e., of the effects of the position-dependent energy gap $E_g = E_C - E_V$, electron affinity χ, and effective density of states N_C, N_V of the conduction and valence bands.[†] Having defined the electron and hole quasi-Fermi levels $E_{F,n}$, $E_{F,p}$ in terms of Fermi–Dirac integrals of order 1/2 [38,39],

$$n = N_C \mathcal{F}_{\frac{1}{2}}\left(\frac{E_{F,n} - E_C}{k_B T}\right), \quad p = N_V \mathcal{F}_{\frac{1}{2}}\left(\frac{E_V - E_{F,p}}{k_B T}\right), \tag{2.24}$$

corrections to Boltzmann statistics are included by means of the fractions $\gamma_n = n/n_B$, $\gamma_p = p/p_B$ (in Equation 2.23), where n_B and p_B are the carrier densities in the nondegenerate limit ($n < N_C, p < N_V$):

$$n_B = N_C \exp\left(\frac{E_{F,n} - E_C}{k_B T}\right), \quad p_B = N_V \exp\left(\frac{E_V - E_{F,p}}{k_B T}\right). \tag{2.25}$$

2.2.5.1 Generation–Recombination Models

The approximations to derive the DD model discussed in Section 2.2.4 must be completed by introducing the collision term relevant in this simplified context, i.e., the expressions of the net recombination rates $U_{n,p}$ in the RHS of Equation 2.21. In a recombination process, an electron in the conduction band fills an empty state of the valence band, annihilating an electron–hole (e–h) pair and releasing energy, e.g., in the form of photons or phonons.[‡] In the reverse process (generation), an electron in the valence band is promoted to the conduction band, creating an e–h pair. As generation–recombination (GR) transitions must conserve energy and momentum, they can be phonon assisted (thermal), photon assisted (radiative), or assisted by other electrons or holes (Auger). Moreover, GR transitions can be either interband (direct mechanisms) or assisted by intermediate trap levels in the forbidden band (indirect or Shockley–Read–Hall mechanisms). In direct-gap semiconductors, direct optical GR is typically the dominant mechanism, whereas trap-assisted GR is usually prevalent in indirect-gap materials [2]. In stationary conditions, and whenever the GR process is band-to-band (i.e., involves direct and instantaneous transitions between the valence and conduction bands), $U_n = U_p = U$. In time-varying conditions, $U_n \neq U_p$ if GR takes place through intermediate traps or recombination centers acting as electron or hole reservoirs; in this case, the DD model should be complemented by additional trap equations [41].

Following the seminal work of Blakemore and Landsberg [42,43], recombination mechanisms in semiconductors have been the object of a large body of theoretical and experimental studies. Restricting our attention to steady-state carrier transport in light emitters, the net recombination rate U in the bulk regions of the device includes contributions from the spontaneous emission (U^{sp}), Auger (U^{A}), and Shockley–Read–Hall (U^{SRH}) rates,

---

[†] An alternative way of treating abrupt heterointerfaces involves the use of a thermionic boundary condition (BC) that constrains the current across the interface. This approach is limited to planar structures [40].

[‡] Transitions within the same band are not normally considered recombination events.

$$U^{\mathrm{sp}} = B(np - \gamma_n \gamma_p n_{\mathrm{i}}^2)$$

$$U^{\mathrm{SRH}} = \frac{np - \gamma_n \gamma_p n_{\mathrm{i}}^2}{\tau_p(n + \gamma_n n_{\mathrm{t}}) + \tau_n(p + \gamma_p p_{\mathrm{t}})}$$

$$U^{\mathrm{A}} = (C_n n + C_p p)(np - \gamma_n \gamma_p n_{\mathrm{i}}^2),$$

where $\gamma_{n,p}$ has been introduced to account for the effect of Fermi statistics as in Equation 2.23, B is the radiative recombination coefficient (in units of cm^3/s), n_{i} is the intrinsic carrier density of the material, $n_{\mathrm{t}} = n_{\mathrm{i}} \exp(E_{\mathrm{trap}}/k_{\mathrm{B}}T)$, $p_{\mathrm{t}} = n_{\mathrm{i}} \exp(-E_{\mathrm{trap}}/k_{\mathrm{B}}T)$, τ_n and τ_p are the average carrier lifetimes, E_{trap} is the difference between the defect level and the intrinsic level, C_n and C_p are the Auger coefficients (in units of cm^6/s). The electron- or hole-assisted recombination rate (Auger recombination) is proportional to p^2n or pn^2, implying proportionality not only with respect to the colliding populations (electrons and holes) but also to the population of the energy suppliers. Due to this dependence, Auger recombination is important in high-injection devices, where it can be an unwelcome competitor of radiative recombination.[†]

2.2.5.2 Boundary Conditions and Numerical Solution Schemes

All the physical models described earlier must be complemented with a suitable set of BCs. A first kind of BC concerns ideal insulating boundaries, for which the normal component of the current density is zero. This condition corresponds to zero normal derivative (also called homogeneous Neumann BC) for the potential and the carrier densities. Metallic contacts fall into two main classes: ohmic and rectifying (Schottky) contacts. In both cases, metallic contacts are assumed to be made of ideal conductors, i.e., they are equipotential. Ohmic contacts ideally behave as short circuits, since they do not support any potential drop, while rectifying Schottky contacts exhibit a diode-like behavior. For ohmic contacts, the local electron and hole densities assume the equilibrium values corresponding to the local doping, yielding a Dirichlet BC (i.e., the unknown is assigned). If the contact is controlled by an ideal voltage source, a Dirichlet BC also arises for the potential. The treatment of Schottky contacts is more involved, since several approximations can be exploited. A fairly accurate set of BCs can be derived from the Sze–Bethe thermionic-drift diffusion model [44].

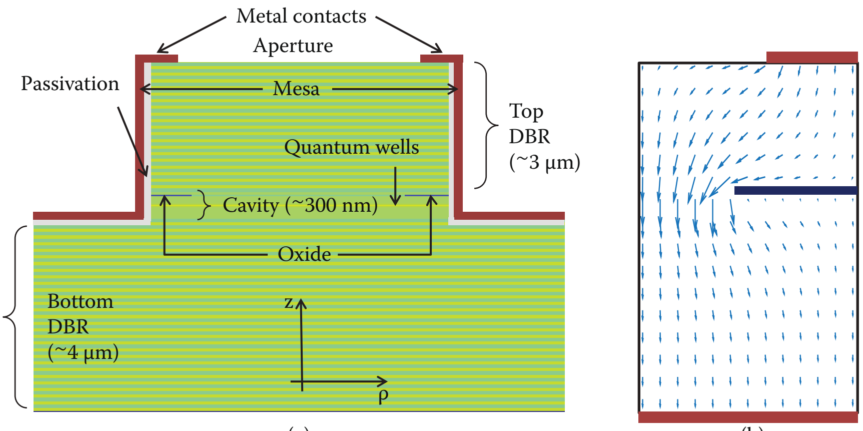
The numerical solution of Equation 2.21 requires discretization in space and time. Spatial discretization reduces the physical model to a set of ordinary time-domain differential equations; the need to perform a time-domain solution depends on the regime where the device operates. In static DC conditions time derivatives can be directly set to zero, and the resulting system of nonlinear equations can be solved through the Newton method and its variants (e.g., the Newton–Richardson approach [45]). Small-signal AC excitation can be dealt with by direct linearization and transformation into the frequency domain [41]. Large-signal analysis may also be carried out in the frequency domain by means of the harmonic balance method [46]. Conventional finite-difference (FD) or finite-element (FE) techniques are ill-suited to the spatial discretization of conservation equations, including both advection (drift) and diffusion terms. Advances in the development of stable numerical schemes for semiconductor device simulation were made in 1969, when Scharfetter and Gummel (SG) proposed a discretization scheme for the 1D current continuity equation [47], which was later extended to two-dimensional (2D) and 3D geometries by means of the generalized finite-box (FB) approach, applicable to arbitrary Delaunay grids and to axisymmetric geometries [45,48–52]. The SG–FB scheme is exploited today by most device simulators. Inspired by computational fluid dynamics [53], upwind FE schemes based on a Petrov–Galerkin method with asymmetric weight functions (equivalent to a Galerkin approach with the explicit addition of numerical diffusivity) were also proposed for the spatial discretization of the HD equations [28].

[†] The inverse process of Auger recombination is generation by impact ionization. In high-field conditions, electrons and holes gather enough energy from the electric field between two successive scattering events (e.g., collisions with phonons or impurities) to be able to interact with another electron (hole) and promote it to the conduction (valence) band. Each electron or hole is therefore able to generate, over a certain length, a number of electron–hole pairs, that undergo in turn the same process. The resulting chain can lead to a divergent current, i.e., to avalanche breakdown in the semiconductor. Avalanche breakdown occurs for electric fields of the order of the breakdown field, and is usually an undesirable effect, with the notable exception of avalanche photodetectors [17].

2.3 Quantum-Corrected Drift-Diffusion

For illustrative purposes, we find it convenient to focus the foregoing discussion on VCSELs. The main feature of these devices is the radiation normal to their layers, which eases fiber coupling owing to the circular emitting area. Figure 2.2a shows a typical 850-nm VCSEL with a one-λ cavity containing the quantum wells (QWs). An oxide aperture just above the QWs confines the current (see Figure 2.2b), and improves optical guiding at low bias voltages. The short gain path of VCSELs demands highly reflective (>99%) distributed Bragg reflectors (DBRs) to achieve lasing. In fact, most of the VCSEL volume is taken up by the DBRs, which include several pairs of high and low refractive index layers with $\lambda/4$ thicknesses [54]. From one perspective, VCSELs represent a paradigm of sorts in the modeling of optoelectronic devices, as their simulation requires a comprehensive multiphysics and multiscale approach to account for electrical, optical, and thermal aspects [80]. In this complex scenario, the only viable approach to carrier transport modeling seems to be a semiclassical one, complemented with appropriate models connecting the microscopic scale governed by coherent phenomena with the macroscopic scale governed by incoherent processes. These models, often referred to as “quantum corrections,” account for a variety of nonclassical effects, such as quantum confinement, (quasi-)ballistic transport, tunneling, and excitonic effects in optical spectra. The most accurate approach to introducing quantum corrections in a semiclassical device simulator is to complement the carrier transport model with a Poisson–Schrödinger solver (other approaches based, e.g., on the density gradient formalism account for nonlocality effects but they are limited to dispersions with one subband only [55]).

Multiband $k\cdot p$ envelope-function models represent the traditional approach to compute subband dispersions in optoelectronics, as they represent a fair compromise between accuracy and computational cost. The unknown nanostructure wavefunction is expanded in terms of zone-center Bloch bands $u_m(\underline{\rho}, z)$



(a) (b)

FIGURE 2.2 (a) Vertical cavity surface emitting laser (VCSEL) structure with semiconductor distributed Bragg reflectors. (b) The simulated current distribution profile.

(usually defined at the center of Brillouin zone) of the underlying bulk solids. An extremely concise representation of the nanostructure excitation spectrum is obtained by selecting a reduced set of M bulk states of interest, typically the three top valence bands and the first conduction band (inclusion of spin–orbit coupling doubles M to eight), and by removing the remaining remote bands by means of Luttinger–Kohn perturbation theory. Arresting the perturbative expansion to second-order terms, and writing the bulk Hamiltonian in a form that explicitly shows its wave vector dependence, we have

$$[H_{k\cdot p}] = \sum_{\alpha,\beta=x,y,z} [H^{(2)}_{\alpha\beta}]k_\alpha k_\beta + \sum_{\alpha=x,y,z} [H^{(1)}_{\alpha}]k_\alpha + [H^{(0)}], \tag{2.26}$$

where the expressions of the matrices above depend on crystal symmetry and on the zone-center functions employed in the $k\cdot p$ representation [56–58]. Second-order terms $H^{(2)}_{ij}$ include free carrier dispersion and the perturbation treatment of remote states in perturbation theory, first-order terms $H^{(1)}_{i}$ stem from the direct treatment of the $k\cdot p$ interaction and spin–orbit coupling, and the zeroth–order term $H^{(0)}$ contains zone-center energies.

We assume the following *ansatz* for the wavefunction of a planar quantum well nanostructure,

$$\psi(\underline{\rho}, z) = \frac{1}{\sqrt{\mathcal{A}}} e^{\mathrm{i}\underline{k}\cdot\underline{\rho}} \sum_{m=1}^{M} u_{m0}(\underline{\rho}, z)\zeta_{m\underline{k}}(z), \tag{2.27}$$

where the coordinates z and $\underline{\rho}$ are the symmetry broken and the translational invariant directions, respectively, $\mathcal{L}$, $\mathcal{A}$ their corresponding normalization volumes ($\mathcal{V} = \mathcal{A}\mathcal{L}$), $\underline{k}$ is the transversal crystal momentum, and $u_{m0}(\underline{\rho}, z)$ are $k\cdot p$ orthonormal lattice-periodic functions rapidly changing over the crystal primitive cell. The slowly varying envelopes $\zeta_{m\underline{k}}(z)$ describe, at every position in the symmetry broken direction z, how the lattice-periodic functions are mixed together. Upon substitution of the wave vector component k_z along the quantized direction with the corresponding operator $-\mathrm{i}\partial_z$, one obtains a coupled equation system for the nanostructure envelopes $\{F(z)\} = \{\zeta_1(z), \zeta_2(z), \ldots, \zeta_M(z)\}^T$,

$$\left(-\partial_z[H^{(2)}(\underline{k}, z)]\partial_z - [H^{(1)}_L(\underline{k}, z)]\mathrm{i}\partial_z - \mathrm{i}\partial_z[H^{(1)}_R(\underline{k}, z)] + [H^{(0)}(\underline{k}, z)]\right)\{F(z)\} = E\{F(z)\}, \tag{2.28}$$

parametrized for the wave vector $\underline{k}$ in the translational invariant space. Notice that $H^{(1)}$ has been split into two terms depending on the position of the operator ∂_z (terms of the type $\partial_z^2 H^{(2)}$ and $H^{(2)}\partial_z^2$ are usually neglected in $k\cdot p$ theories). Having discretized the device region into a number of elements, where the envelopes are represented with Lagrange polynomials $\zeta_m(z) = \sum_j \zeta_{mj}N_j(z)$, application of Galerkin's procedure leads to the weak form of Equation 2.28

$$\sum_j \int_e \mathrm{d}z \left(N^z_i[H^{(2)}]N^z_j - \mathrm{i}N_i[H^{(1)}_L]N^z_j + \mathrm{i}N^z_i[H^{(1)}_R]N_j + N_i[H^{(0)}]N_j\right)\{F_j\}$$

$$= E\sum_j \int_e \mathrm{d}z\, N_i N_j\,\{F_j\}, \tag{2.29}$$

with $N^z_i = \partial_z N_i$. Boundary terms, arising when $[H^{(2)}]$ and $[H^{(1)}_R]$ are integrated by parts, vanish by the assumed homogeneous Dirichlet or Neumann BCs, since we are interested in the bound states of closed systems.† Notice that integration by parts is also an elegant way to *naturally* treat material discontinuities [59]. Assembling all elements, i.e., summing over e, one obtains the eigenmatrix equation

---

† The integration-by-parts boundary terms neglected in Equation 2.29 play an important role in the calculation of current-carrying scattering states of open systems, as they describe how the device interacts with the semi-infinite leads; we return to this point in Section 2.4 with the derivation of the boundary self-energy of the contacts.

$$[H(\underline{k})]\{F\} = E[M]\{F\}, \tag{2.30}$$

where $[H]$ and $[M]$ are $MN \times MN$ sparse FE matrices, and the MN column vector $\{F\}$ is the FE representation of the nanostructure envelopes. Multiband $k \cdot p$ models are extremely concise and require only a few physically sensitive band parameters, which are usually obtained from a combination of experiments and first-principles electronic structure calculations.[†]

The Schrödinger equation should be reconciled with the solution of the Poisson equation to account for space charge effects. A simple iterative approach that computes the quantum charge from the Schrödinger equation and inserts it in the RHS of the Poisson equation, the solution of which updates in turn the Hartree potential, may be unstable due to the high sensitivity of the energy levels on the confining potential. On the other hand, an analytic expression of the Jacobian for a Newton updating scheme cannot be derived. An elegant and stable solution is provided by the predictor-corrector algorithm proposed in [65]. The idea is to define an approximate expression for the quantized charge as a function of the confining potential, so that most of the nonlinearities of the Schrödinger equation can be moved into the Poisson equation, which can be efficiently solved in an inner Newton loop. Stability is achieved by correcting the predicted results in an outer iteration loop by the exact solution of the Schrödinger's equation [65].

2.3.1 Capture and Escape Processes

Given our inability to quantize the whole VCSEL cross section, we must face the problem of letting bulk-like, 3D carriers coexist with 2D carriers populating the VCSEL active region, and connect them by means of appropriate scattering mechanisms. The choice of the energetic boundary E_{u}^0 between quantum and bulk regions is somewhat arbitrary [66]. A possible choice is to define the boundary by replacing the band edge of the spatial quantum region with the band edge of the barrier region in the equations for the unbound density, which leads to

$$n^{\mathrm{3D}}(\underline{\rho}, z) = \int_{E_{\mathrm{u}}^0}^{\infty} g_{\mathrm{3D}}(E) f_{\mathrm{FD}}(E, E_{\mathrm{F},n}^{\mathrm{3D}})\,\mathrm{d}E \tag{2.31}$$

$$n^{\mathrm{2D}}(\underline{\rho}, z) = \sum_i \left|\zeta_i(z)\right|^2 \int_{E_{\mathrm{i}}}^{E_{\mathrm{u}}^0} g_{\mathrm{2D}}(E) f_{\mathrm{FD}}(E, E_{\mathrm{F},n}^{\mathrm{2D}})\,\mathrm{d}E, \tag{2.32}$$

where $g_{\mathrm{3D}}(E)$, $g_{\mathrm{2D}}(E)$ are the density of states of the bulk population and the ith quantized level with the envelope function $\zeta_i(z)$, respectively, and $f_{\mathrm{FD}}(E, E_{\mathrm{F}}) = 1/\left[1 + \exp\left(\frac{E-E_{\mathrm{F}}}{k_{\mathrm{B}}T}\right)\right]$ is the Fermi–Dirac distribution. Having separated the two populations, the electron continuity equation in Equation 2.21 in the active region of the VCSEL is replaced with (a similar equation holds for holes) [67]

[†] The apparent simplicity of $k \cdot p$ models should not mislead the reader into thinking that first-principles/experimental information of bulk properties can be seamlessly combined to derive nanostructure properties. As it happens, band parameters are often derived from the fitting of bulk bands (sometimes with targets selected from different sources), disregarding the non-commutability of potential and position operators, and, most importantly, the ordering of the wave number operators, which ultimately means parameter splitting [58]. As terms of the type Hk_ik_j with $i \neq j$ in Equation 2.26 do not have a well-defined operator ordering, they are usually split symmetrically to $-\frac{1}{2}(\partial_i H \partial_j + \partial_j H \partial_i)$ [59]. What makes parameter splitting subtle is that it is inconsequential in the calculation of bulk material dispersions, but it becomes relevant when the sub-band dispersion of a nanostructure is considered. Symmetrized operator ordering and/or an ill-advised choice of the band parameters may lead to significant deviations of subband dispersions, incorrect matrix elements, or even to the appearance of spurious solutions [60,61]. A multiband envelope-function model for wurtzite nanostructures based on a rigorous numerical procedure to determine operator ordering and band parameters from full-Brillouin-zone nonlocal empirical pseudopotential calculations is presented in [58]. For a discussion about the validity of the envelope-function approximation, see [62,63]. Atomistic models beyond the envelope-function approximation are reviewed in [64].

$$\nabla \cdot \underline{J}^{\mathrm{n}} = \mathrm{q}\left(U^{\mathrm{SRH}} + U^{\mathrm{sp}} + U^{\mathrm{cap}} + \frac{\partial n_{3\mathrm{D}}}{\partial t}\right) \tag{2.33}$$

$$\nabla \cdot \underline{J}^{\mathrm{n}}_{2\mathrm{D}} = \mathrm{q}\left(R^{\mathrm{nr}} + R^{\mathrm{sp}} + R^{\mathrm{st}} - U^{\mathrm{cap}} + \frac{\partial n_{2\mathrm{D}}}{\partial t}\right), \tag{2.34}$$

where the stimulated and spontaneous radiative rates R^{st}, R^{sp} are defined by the optical response of the QWs (see Section 2.3.2), and R^{nr} is the nonradiative (SRH and Auger) recombination rate in the active region. The two populations are coupled by a capture rate U^{cap} that acts as a drain for the 3D charges and a source for the 2D carriers. Equation 2.34 describes the lateral currents (in the unconfined directions) of the 2D gas, whose density is determined by the localized wavefunctions given by the self-consistent solution of the Poisson–Schrödinger problem. From a physical standpoint, a capture process may be understood as a carrier propagating above a QW emitting phonons and finally relaxing into a bound state of the well.[†] Once in the bound state, the carrier may either absorb phonons and be excited back into the continuum or recombine by any of the radiative or dark recombination mechanisms. If interband scattering processes are slow enough compared to intraband processes in both the continuum and the bound states, the spectral distribution of the carriers may be approximated by Fermi distributions defined by quasi-Fermi levels.[‡] From the DD modeling standpoint, the capture of a carrier from the bulk to the quantum region may be seen as a recombination mechanism for the unbound population and as a generation mechanism for the bound population, and vice versa for an escape process. A net capture rate for electrons may be derived in a fashion similar to SRH recombination theory [68] (a similar expression holds for holes)

$$U^{\mathrm{cap}} = \left(1 - \exp\left(\frac{E^{2\mathrm{D}}_{\mathrm{F},n} - E^{3\mathrm{D}}_{\mathrm{F},n}}{k_{\mathrm{B}}T}\right)\right)\left(1 - \frac{n^{2\mathrm{D}}}{N_2}\right)\frac{n^{3\mathrm{D}}}{\tau_{\mathrm{scat},n}}, \tag{2.35}$$

where the parameter $N_2 = \sum_i \int_{E_{\mathrm{i}}}^{E^0_{\mathrm{u}}} g_{2\mathrm{D}}(E)\,\mathrm{d}E$ is the sum of all QW states per unit volume, i.e., the maximum bound electron charge. The net capture rate is, therefore, proportional to the unbound carrier density, decreases with the filling of the bound states, increases with the quasi-Fermi level separation, and vanishes in the case of a completely filled bound population (phase space filling) or equal quasi-Fermi levels (thermodynamic equilibrium). In principle, the scattering rate should also depend on the distribution of the initial states of the injected carriers, but in a DD framework void of any information concerning the spectral distribution of the carriers, the complex physics of scattering can only be lumped in one average scattering time $\tau_{\mathrm{scat},n}$, which is usually treated as a fitting/phenomenological parameter. Relating these scattering times to structural parameters of the nanostructure is essential to understand some critical aspects related to vertical carrier transport in high-injection conditions, e.g., the possible leakage of hot carriers above the barriers, or to predict the modulation bandwidth of lasers [71,72].

2.3.2 Coupling Carrier Transport Models with Optical Solvers

From the electromagnetic standpoint, a VCSEL is an open resonator whose resonance frequency is determined by the cavity length and the mirrors design. Lasing action is achieved when the gain of the active material compensates the losses of the open structure. This condition defines the threshold gain, which, together with the operation wavelength and the light pattern emitted by the VCSEL, is the output of the optical solver. Among full-wave optical solvers (beyond scalar-based approaches, e.g., on the effective index method [73]), one may distinguish fully numerical schemes from semianalytical techniques that exploit the physics/geometry of the problem. A remarkable example of the former class is the frequency-domain FE solver LUMI (that uses a 2D body-of-revolution expansion for axisymmetric devices and perfectly matched layers BCs), also included in the Synopsys Sentaurus Device simulator [36,67]. An example belonging to the latter class is the method of lines (MoL), which is based on a spatial discretization in the radial direction and on the transmission line formalism in the longitudinal direction [74]. Other noteworthy examples of physics-based techniques rely on plane-wave expansions, where, as in MoL, the field is propagated along the longitudinal direction by admittance [75,76] or reflection matrix [77] formulations.

[†] NEGF simulations suggest that the process is more likely to consist of tunneling to a bound state above the barriers (i.e., with a high transverse kinetic energy) followed by its subsequent relaxation assisted by phonons [69].

[‡] This assumption breaks down at high injections, if carrier–carrier collisions are not fast enough to replenish the spectral hole burned by the state-selective stimulated emission, leading to deviations of the carriers' spectral densities from their quasiequilibrium Fermi–Dirac distributions [70].

The authors of this chapter have proposed a 3D vectorial model of VCSELs [78–80] based on the expansion of the electromagnetic field in terms of the complete basis of the TE and TM modes of the cavity medium in cylindrical coordinates. All deviations with respect to the reference unperturbed structure, both in the transverse and longitudinal directions, are accounted by coupled-mode theory. By cascading the transmission matrices of each section of the device (evaluated by an exponential matrix when the section is longitudinally invariant), one can compute the overall transmission matrix between the two radiating sections, beyond which the device can be considered homogeneous. Resolution of the problem is obtained by introducing appropriate BCs, relating backward to forward waves at these two sections. Since the condition for an electromagnetic field to be a mode of the structure is to replicate itself after a full cavity round-trip, one finally obtains an eigenvalue problem, whose complex eigenvalues represent the modal wavelengths and the corresponding threshold gains (real and imaginary parts, respectively), and whose eigenvectors are the expansion coefficients of the field in the section where the round-trip condition is enforced.

In order to extend the semiconductor equations to model laser diodes, we still need to describe the time evolution of the electromagnetic energy in each lasing mode. The rate equation for the output stimulated power $P_{\mathrm{st},m}$ in the mth mode can be derived from quantum theory of semiconductor lasers assuming quasineutrality ($n \approx p$) and adiabatic elimination of the microscopic polarization [81]

$$\Gamma_z v_{\mathrm{ph}} \left(G_m - G_{\mathrm{th},m}\right) P_{\mathrm{st},m} + \Gamma_z \frac{\hbar\omega_m}{\tau_{\mathrm{out}}} R_{\mathrm{sp},m} = \frac{\partial P_{\mathrm{st},m}}{\partial t}. \tag{2.36}$$

The transverse modal gain G_m and the spontaneous emission rate $R_{\mathrm{sp},m}$ are

$$G_m = \int g(\underline{\rho}, \omega_m) \mathcal{E}_m^2(\underline{\rho})\, \mathrm{d}\underline{\rho} \tag{2.37}$$

$$R_{\mathrm{sp},m} = \int r_{\mathrm{sp}}(\underline{\rho}, \omega_m) \mathcal{E}_m^2(\underline{\rho})\, \mathrm{d}\underline{\rho}, \tag{2.38}$$

where $g(\underline{\rho}, \omega_m)$, $r_{\mathrm{sp}}(\underline{\rho}, \omega_m)$ are the local material gain and spontaneous emission at the lasing frequency ω_m; the longitudinal confinement factor Γ_z, the threshold gain $G_{\mathrm{th},m}$, and the photon escape time τ_{out} (s) are provided by the optical solver; $\mathcal{E}_m(\underline{\rho})$ is the mth mode field amplitude, at the QW section, whose intensity is normalized to unity. The conversion of the active region 2D carriers to coherent and incoherent photons is included in the DD transport model by the recombination terms

$$R^{\mathrm{st}}(\underline{\rho}) = \sum_m \frac{D_m}{\hbar\omega_m} g(\underline{\rho}, \omega_m) \mathcal{E}_m^2(\underline{\rho}) P_{\mathrm{st},m} \tag{2.39}$$

$$R^{\mathrm{sp}}(\underline{\rho}) = \int r_{\mathrm{sp}}(\underline{\rho}, \omega) \rho^{\mathrm{opt}}(\omega) \mathrm{d}\omega, \tag{2.40}$$

where the sum is over all lasing modes, $D_m^{-1} = \int \mathcal{E}_m^2(\underline{\rho})\,\mathrm{d}\underline{\rho}\big|_{\text{output}}$ links the field intensity in the active region to the output one radiated by the VCSEL [80], and $\rho^{\text{opt}}(\omega) = (n_b^2\omega^2)/(\pi^2\hbar c^3)$ is the photon density of states. The requirement that the solutions of the electromagnetic problem match the optical response of the active region provides a self-consistent mechanism for the selection of the lasing modes.

2.3.2.1 The Semiconductor Bloch Equations

The starting point for the calculation of the optical response of the active region is the single-particle Hamiltonian of the nanostructure describing the free carriers interacting with a classical field $\mathcal{E}(t)$

$$H = H_0 + H_{cf} + H_{cc} \tag{2.41}$$

with the following expressions for the free-carrier H_0, the carrier-field H_{cf}, and the carrier–carrier H_{cc} contributions

$$H_0 = \sum_A E_A \mathrm{a}_A^\dagger \mathrm{a}_A \tag{2.42}$$

$$H_{cf} = -\sum_{A,B} \mu_{A,B}\mathcal{E}\,(t)\,\mathrm{a}_A^\dagger \mathrm{a}_B \tag{2.43}$$

$$H_{cc} = \frac{1}{2}\sum_{A,B,C,D} \mathrm{V}^{A\,B}_{C\,D}\,\mathrm{a}_A^\dagger \mathrm{a}_B^\dagger \mathrm{a}_D \mathrm{a}_C \tag{2.44}$$

where E_A is the single-particle energy, and the a and $a^\dagger$ are the annihilation and creation operators, compound indices A, B, C, D include the transverse electron momentum $\underline{k}$ and the band index c, v. In the approximation of slowly varying envelopes, the optical dipole matrix element $\mu_{A,B}$ is related to the matrix element of the momentum operator $\hat{p}$ [59]

$$\mu_{AB} \approx -\frac{\mathrm{q}\mathcal{A}}{\mathrm{i}m_0\omega}\sum_{ij}\langle u_{i0}|\hat{p}|u_{j0}\rangle \int_{\mathcal{L}} \zeta_{Ai}^*(z)\zeta_{Bj}\,\mathrm{d}z. \tag{2.45}$$

The unscreened Coulomb matrix element in Equation 2.44 is

$$\mathrm{V}^{A\,B}_{C\,D} = \frac{\mathrm{q}^2}{4\pi\varepsilon_0}\iint_{\mathcal{V}} \psi_A^*(\underline{\rho}, z)\psi_B^*(\underline{\rho}', z')\frac{1}{|(\underline{\rho}, z) - (\underline{\rho}', z')|}\psi_D(\underline{\rho}', z)\psi_C(\underline{\rho}, z)\,\mathrm{d}\underline{\rho}\,\mathrm{d}\underline{\rho}'\mathrm{d}z\,\mathrm{d}z'. \tag{2.46}$$

Substituting the ansatz Equation 2.27 and changing the integration over $\underline{\rho}'$ into the integration over the distance $\underline{s} = \underline{\rho} - \underline{\rho}'$, we have [59]

$$\begin{aligned}\mathrm{V}^{A\,B}_{C\,D} &= \frac{\mathrm{q}^2}{4\pi\varepsilon_0}\iint_{\mathcal{L}} \mathrm{d}z\,\mathrm{d}z'\zeta_A^*(z)\zeta_B^*(z')\zeta_D(z')\zeta_C(z)\\ &\times\iint_{\mathcal{A}} \frac{1}{\sqrt{s^2 + (z - z')^2}}\mathrm{e}^{\mathrm{i}(\underline{k}_C + \underline{k}_D - \underline{k}_A - \underline{k}_B)\cdot\underline{\rho}}\ \ \mathrm{e}^{\mathrm{i}(\underline{k}_B - \underline{k}_D)\cdot\underline{s}}\,\mathrm{d}\underline{\rho}\,\mathrm{d}\underline{s}\end{aligned} \tag{2.47}$$

The integration over $\underline{\rho}$ yields the momentum conserving function $\delta_{\underline{k}_A+\underline{k}_B,\underline{k}_C+\underline{k}_D}$, reflecting the in-plane translational invariance of the system. Replacing $1/\sqrt{s^2 + (z - z')^2}$ with its Fourier expansion $\sum_{\underline{q}} \frac{2\pi}{\mathcal{A}q}\mathrm{e}^{-q|z-z'|}\mathrm{e}^{\mathrm{i}\underline{q}\cdot\underline{s}}$, we have

$$\mathrm{V}^{A\,B}_{C\,D} = \frac{\mathrm{q}^2}{4\pi\varepsilon_0}\iint_{\mathcal{L}} \mathrm{d}z\,\mathrm{d}z'\zeta_A^*(z)\zeta_B^*(z')\zeta_D(z')\zeta_C(z)\int_{\mathcal{A}}\sum_q \frac{2\pi}{\mathcal{A}q}\mathrm{e}^{-q|z-z'|}\mathrm{e}^{\mathrm{i}\underline{q}\cdot\underline{s}}\ \mathrm{e}^{\mathrm{i}(\underline{k}_2 - \underline{k}_3)\cdot\underline{s}}\,\mathrm{d}\underline{s}. \tag{2.48}$$

The integral over $\underline{s}$ yields $\mathcal{A}\delta_{\underline{q},\underline{k}_B-\underline{k}_D}$ and we are left with

$$\mathrm{V}^{A\ B}_{C\ D} = \frac{\mathrm{q}^2}{2\varepsilon_s}\frac{1}{q}\iint_{\mathcal{L}} \zeta_A^*(z)\zeta_B^*(z')\mathrm{e}^{-q|z-z'|}\zeta_D(z')\zeta_C(z)\,\mathrm{d}z\,\mathrm{d}z'. \tag{2.49}$$

Having defined the two compound indices $1 = (\underline{k}, v)$ and $2 = (\underline{k}, c)$, the dynamics of the microscopic interband polarization $p_{\underline{k}} = \langle \mathrm{a}_1^\dagger \mathrm{a}_2\rangle$ may be obtained from the Heisenberg equation of motion

$$\mathrm{i}\hbar\frac{\mathrm{d}p_{\underline{k}}}{\mathrm{d}t} = [p_{\underline{k}}, H], \tag{2.50}$$

which gives for the different Hamiltonian components [82]

$$\frac{\mathrm{d}}{\mathrm{d}t}p_{\underline{k}}|_{H_0} = \frac{\mathrm{i}}{\hbar}\left(E_1 - E_2\right)\langle \mathrm{a}_1^\dagger \mathrm{a}_2\rangle \tag{2.51}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}p_{\underline{k}}|_{H_{cf}} = -\frac{\mathrm{i}}{\hbar}\mu_{\underline{k}}[\langle \mathrm{a}_2^\dagger \mathrm{a}_2\rangle - \langle \mathrm{a}_1^\dagger \mathrm{a}_1\rangle]\mathcal{E}(t) \tag{2.52}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}p_{\underline{k}}|_{H_{cc}} = &+\frac{\mathrm{i}}{2\hbar}\sum_{A,B,D}\mathrm{V}^{A\ B}_{1\ D}\ \langle \mathrm{a}_A^\dagger \mathrm{a}_B^\dagger \mathrm{a}_D \mathrm{a}_2\rangle - \frac{\mathrm{i}}{2\hbar}\sum_{A,B,C}\mathrm{V}^{A\ B}_{C\ 1}\ \langle \mathrm{a}_A^\dagger \mathrm{a}_B^\dagger \mathrm{a}_C \mathrm{a}_2\rangle \\ &+\frac{\mathrm{i}}{2\hbar}\sum_{A,C,D}\mathrm{V}^{A\ 2}_{C\ D}\ \langle \mathrm{a}_1^\dagger \mathrm{a}_A^\dagger \mathrm{a}_D \mathrm{a}_C\rangle - \frac{\mathrm{i}}{2\hbar}\sum_{B,C,D}\mathrm{V}^{2\ B}_{C\ D}\ \langle \mathrm{a}_1^\dagger \mathrm{a}_B^\dagger \mathrm{a}_D \mathrm{a}_C\rangle\end{aligned} \tag{2.53}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}p_{\underline{k}}|_{H_{cp}} = -\frac{\mathrm{i}}{2\hbar}\sum_{3u}\left[g_{23}\langle \mathrm{a}_1^\dagger \mathrm{a}_3 \mathrm{b}_u\rangle - g_{31}\langle \mathrm{a}_3^\dagger \mathrm{a}_2 \mathrm{b}_u\rangle + g_{32}^*\langle \mathrm{a}_1^\dagger \mathrm{a}_3 \mathrm{b}_u^\dagger\rangle - g_{13}^*\langle \mathrm{a}_3^\dagger \mathrm{a}_2 \mathrm{b}_u^\dagger\rangle\right]. \tag{2.54}$$

Renaming summation indices and exploiting the Coulomb matrix element symmetries ($\mathrm{V}^{A\ B}_{C\ D} = \mathrm{V}^{B\ A}_{D\ C}$ and $\mathrm{V}^{A\ B}_{C\ D} = \mathrm{V}^{C\ D\,*}_{A\ B}$), we obtain

$$\frac{\mathrm{d}}{\mathrm{d}t}p_{\underline{k}}|_{H_{cc}} = \frac{\mathrm{i}}{\hbar}\sum_{A,B,D}\mathrm{V}^{A\ B}_{1\ D}\ \langle \mathrm{a}_A^\dagger \mathrm{a}_B^\dagger \mathrm{a}_D \mathrm{a}_2\rangle - \frac{\mathrm{i}}{\hbar}\sum_{A,B,D}\mathrm{V}^{2\ A}_{B\ D}\ \langle \mathrm{a}_1^\dagger \mathrm{a}_A^\dagger \mathrm{a}_D \mathrm{a}_B\rangle. \tag{2.55}$$

Because of the many-particle interaction, the equation above is not closed. The dynamics of the single-particle quantity $\sigma_{AB} = \langle \mathrm{a}_A^\dagger \mathrm{a}_B\rangle$ couples to two-particle contributions $C^{AB}_{CD} = \langle \mathrm{a}_A^\dagger \mathrm{a}_B^\dagger \mathrm{a}_C \mathrm{a}_D\rangle$. Applying the correlation expansion

$$C^{AB}_{CD} = \sigma_{AD}\sigma_{BC} - \sigma_{AC}\sigma_{BD} + \tilde{C}^{AB}_{CD}, \tag{2.56}$$

where $\tilde{C}^{AB}_{CD} = \langle \mathrm{a}_A^\dagger \mathrm{a}_B^\dagger \mathrm{a}_C \mathrm{a}_D\rangle^c$, the first-order Hartree–Fock (HF) contribution and the second-order correlation terms are separated:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}p_{\underline{k}}|_{H_{cc}^{HF}} = &\frac{\mathrm{i}}{\hbar}\sum_{A,B,D}\mathrm{V}^{A\ B}_{1\ D}\ [\langle \mathrm{a}_A^\dagger \mathrm{a}_2\rangle\langle \mathrm{a}_B^\dagger \mathrm{a}_D\rangle - \langle \mathrm{a}_A^\dagger \mathrm{a}_D\rangle\langle \mathrm{a}_B^\dagger \mathrm{a}_2\rangle] \\ &-\frac{\mathrm{i}}{\hbar}\sum_{A,B,D}\mathrm{V}^{2\ A}_{B\ D}\ [\langle \mathrm{a}_1^\dagger \mathrm{a}_B\rangle\langle \mathrm{a}_A^\dagger \mathrm{a}_D\rangle - \langle \mathrm{a}_1^\dagger \mathrm{a}_D\rangle\langle \mathrm{a}_A^\dagger \mathrm{a}_B\rangle]\end{aligned} \tag{2.57}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}p_{\underline{k}}|_{H_{cc}^{corr}} = \frac{\mathrm{i}}{\hbar}\sum_{A,B,D}\mathrm{V}^{A\ B}_{1\ D}\ \tilde{C}^{A\ B}_{D\ 2} - \frac{\mathrm{i}}{\hbar}\sum_{A,B,D}\mathrm{V}^{2\ A}_{B\ D}\ \tilde{C}^{1\ A}_{D\ B}. \tag{2.58}$$

In a spatially homogeneous system only DM elements that are diagonal with respect to the electronic wave vector, have a nonvanishing expectation value. Therefore, within the HF approximation, the equation of motion for the microscopic polarization $p_{\underline{k}}$ is

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}p_{\underline{k}}|_{H_{cc}^{HF}} = -\frac{\mathrm{i}}{\hbar}\sum_{k'}[\,&(\mathrm{V}^{vk\ vk'}_{vk\ vk'} - \mathrm{V}^{ck\ vk'}_{ck\ vk'})p_{\underline{k}}n^{v}_{k'} + (\mathrm{V}^{vk\ ck'}_{vk\ ck'} - \mathrm{V}^{ck\ ck'}_{ck\ ck'})p_{\underline{k}}n^{c}_{k'}\\ &-\left(\mathrm{V}^{ck\ \ vk'}_{vk'\ ck} - \mathrm{V}^{vk'\ vk}_{vk\ \ vk'}\right)p_{\underline{k}}n^{v}_{k'} - \left(\mathrm{V}^{ck\ \ ck'}_{ck'\ ck} - \mathrm{V}^{ck'\ vk}_{vk\ \ ck'}\right)p_{\underline{k}}n^{v}_{k'}\\ &-\left(\mathrm{V}^{ck\ ck'}_{vk\ vk'} - \mathrm{V}^{ck\ \ ck'}_{vk'\ vk}\right)p^{*}_{k'}(n^{c}_{k} - \rho^{v}_{k})\\ &-\left(\mathrm{V}^{ck\ vk'}_{vk\ ck'} - \mathrm{V}^{ck\ \ vk'}_{ck'\ vk}\right)p_{k'}(n^{c}_{k} - n^{v}_{k})\,]\,.\end{aligned} \tag{2.59}$$

The first line cancels for symmetry arguments. Terms that are nonlinear in the microscopic polarization have been omitted. Neglecting also terms proportional to $p^{*}_{k'} = \langle \mathrm{a}^{\dagger}_{2}\mathrm{a}_{1}\rangle$ involving transitions between different bands, we are left with

$$\frac{\mathrm{d}}{\mathrm{d}t}p_{\underline{k}}|_{H_{cc}^{HF}} = -\frac{\mathrm{i}}{\hbar}\sum_{k'}\left[\left(\mathrm{V}^{vk'\ vk}_{vk\ \ vk'}n^{v}_{k'} - \mathrm{V}^{ck\ \ ck'}_{ck'\ ck}n^{c}_{k'}\right)p_{\underline{k}} + \mathrm{V}^{ck\ \ vk'}_{ck'\ vk}p_{k'}(n^{c}_{k} - n^{v}_{k})\right]. \tag{2.60}$$

Moving to the electron–hole picture, we find the familiar semiconductor Bloch equations in the HF approximation

$$\frac{\mathrm{d}}{\mathrm{d}t}p_{\underline{k}}|_{HF} = -\mathrm{i}\tilde{\omega}_{k}p_{k} - \gamma p_{\underline{k}} - \frac{\mathrm{i}}{\hbar}\left\{\mathcal{E}(t)\mu_{\underline{k}} + \sum_{k'}\mathrm{V}^{ck\ \ vk'}_{ck'\ vk}p_{k'}\right\}(n^{c}_{k} + n^{h}_{k} - 1), \tag{2.61}$$

where we have defined

$$\hbar\tilde{\omega}_{k} = (\varepsilon_{ck} + \varepsilon_{hk}) - \sum_{k'}\{\mathrm{V}^{ck\ \ ck'}_{ck'\ ck}n^{c}_{k'} + \mathrm{V}^{vk'\ vk}_{vk\ \ vk'}n^{h}_{k'}\} \tag{2.62}$$

$$\varepsilon_{hk} = -\left(\varepsilon_{vk} - \sum_{k'}\mathrm{V}^{vk'\ vk}_{vk\ \ vk'}\right). \tag{2.63}$$

The second term in the RHS of Equation 2.61 is added to account for polarization decay due to carrier–carrier or carrier–phonon scattering. Formally integrating Equation 2.61 and screening the Coulomb matrix elements with the Lindhard formula, one obtains the material gain and the spontaneous emission spectra [83]

$$g(\omega) = -\frac{2\omega}{\epsilon_{0}n_{b}c\mathcal{V}\mathcal{E}}\,\mathrm{Im}\left(\sum_{\underline{k}}(\mu_{\underline{k}})^{*}p_{\underline{k}}e^{\mathrm{i}\omega t}\right) \tag{2.64}$$

$$r_{\mathrm{sp}}(\omega) = \frac{1}{\hbar}\left(\frac{n_{b}\omega}{\pi c}\right)^{2}g(\omega)\left[\exp\left(\frac{\hbar\omega - (E_{\mathrm{F},n} - E_{\mathrm{F},p})}{k_{B}T}\right)\right], \tag{2.65}$$

where $\mathcal{E}$ and ω are the amplitude and frequency of the optical wave, ϵ_0 and c are the permittivity and the speed of light in vacuum, and n_b is the background refractive index.

2.4 Towards a Genuine Quantum Approach

High-field nonstationary effects, phase coherence, tunneling, and other quantum features of electronic transport may have barely found mention in a handbook of optoelectronic devices, were it not for the recent developments in novel material systems, highly nanostructured to offer unprecedented performance. As characteristic length scales become smaller than the scattering mean free path, the physics of transport changes drastically from an incoherent, semiclassical picture governed by the BTE to a coherent picture described in a purely quantum mechanical framework in terms of currents associated with probability fluxes, usually injected from idealized reservoirs of carriers. Electronic transport in nanostructures is somewhere between these two limits, in a regime where quantum mechanical phase-coherent phenomena and phase-breaking and dissipative scattering processes coexist.

On the one hand, the immediate point of quantum-corrected semiclassical approaches is obvious, since, although technologically relevant optoelectronic devices do have a nanostructured active region, they really are 3D macroscopic structures that typically do not lend themselves to a fully quantum kinetic treatment. The staggering computational cost of quantum kinetic approaches becomes apparent if one considers, e.g., carrier transport across the DBR stacks of a VCSEL. More generally, a nanostructure is always a 3D system, its dimensionality referring to the free-motion reciprocal subspace only. On the other hand, no quantum correction complementing the DD equations has ever been truly validated, the underlying semiclassical framework being intrinsically local in space and void of any information about the energy distributions of the carriers.

A corollary to the above considerations is that, besides the questionable appropriateness of quantum corrections within a semiclassical framework (the complex physics described by quantum kinetic theories can hardly fit material-agnostic expressions), we are not even confident about some recombination and high-field transport properties of confined systems. For example, carrier leakage and Auger recombination imply the presence of nonequilibrium hot electrons, a concept that represents a major departure from conventional treatments based on DD approaches, which are implicitly based on the effective mass approximation. A legitimate question is therefore to what extent quantum-corrected DD models can be trusted, and what information/conclusions can be drawn from them.

Since all that may sound too abstract, let us consider a concrete example that also happens to divide the optoelectronic community, attracting journalistic attention. There is just no way that the droop debate, a lively controversy about the elusive nature of the nonthermal decline of the internal quantum efficiency in GaN-based visible light-emitting diodes (LEDs) at high current densities [84–86], could have taken place, had transport and recombination properties in nanostructured materials been precisely known.[†] Although most of the recent contributions on droop belong to the camps of proponents either of Auger recombination or of carrier leakage, alternative explanations have also been formulated, the lack of direct experimental information on critical quantities and the ambiguity of the parameters involved in ABC(D) recombination models allowing almost any interpretation. In other words, the debate clearly demonstrates that reproducing droop with a semiclassical model and an appropriate set of fitting parameters does not *eo ipso* imply any model validation.

"Fitting" is, of course, the really sensitive term here. Part of why fitting is so unwelcome in the modeling community is that we are just starting to realize how much we are still forced to rely on untested models, which is threatening to the very concept of numerical simulation as a predictive tool for device design and optimization. Gradually emerging is the awareness that the LED droop problem is too complex and entangled for overly simplified finger-pointing (e.g., an Auger versus leakage dichotomy is simplistic and wrong, because Auger recombination would promote carriers at energies well beyond the heterobarriers and would therefore provide an important contribution to leakage) and that several axiomatic assumptions of electronic transport in optoelectronic devices could lead us to wrong conclusions. The relevant axioms here are the separation of bound and unbound carriers, usually forced by defining an arbitrary energetic boundary, the assumption that carrier confinement is strong enough so that a fully coherent quantum-mechanical description is appropriate for confined carriers, and an incoherent description may be adopted for unbound carriers.

[†] The discussion concerning the loss mechanisms in visible LEDs involves nearly every aspect of carrier transport, including carrier leakage, carrier delocalization, compositional alloy fluctuations, and Auger and defect-assisted recombination. Auger recombination was first proposed as the origin of droop on the basis of the Auger-like density-dependence of the observed losses [87], a proposal supported by first-principles theoretical calculations [88–90] and by recent experiments [91–93]. The main contra-Auger arguments are the ubiquitous use by the industry of electron-blocking layers to prevent electron escape from the active region of UV and visible LEDs, which suggests electron leakage as an important droop-inducing mechanism (see [85,94] and references therein), and the anomalous temperature dependence that the putative Auger coefficients should have to *fit* the efficiency curve [95].

Moving from quantum-corrected semiclassical models to genuine quantum-kinetic descriptions, one has to face the fact that two closely related aspects of the problem, namely carrier transport and optical generation, are traditionally considered from different perspectives. In the emerging field of semiconductor quantum optics, where the focus is on ultrafast transients, the DM approach is usually applied to analyze the correlations' dynamics in the coupled quantized light-semiconductor system[96], while the nonequilibrium Green's function (NEGF) approach [97] is generally considered more suited to carrier transport in the steady-state regime. Although the two formalisms may be shown to be equivalent in the equal-time limit, it is difficult to establish a connection between NEGF and DM in the general case (a discussion on this interesting topic may be found in Refs. [98,99]); and in fact it is not even possible to compare their accuracy, as the strategy to abridge the infinite hierarchy of equations arising in the many-body problem is profoundly different in the two cases (the cluster expansion method in DM theories [96], the summation over some classes of diagrams in NEGF formulations [100]). NEGF and DM are presented in the literature in different contexts probably because the one-time DM approach is admittedly more transparent for time-domain problems, while the two-times NEGF method provides a more device-oriented viewpoint of carrier transport, which seems more appropriate to the analysis of devices seen as open systems, i.e., as systems that exchange current with reservoirs.

2.4.1 Contour-Ordered NEGF

The central idea is that the analysis of the motion of strongly interacting carriers in nanostructured materials does not necessarily require detailed information about each particle in the system, i.e., the many-body wavefunction (that would be a formidable problem), but rather the *average* behavior of zero-, one- or more-particle assemblies, which may be obtained from field theoretical quantities such as no-particle (vacuum amplitude), one-particle, two-particle, and more-particles Green's functions, also known as "propagators." Examples of one-particle quantities of interest are, e.g., the local density of states (the excitation spectrum of the nanostructure), the spectral charge density, which describes how these states are occupied, and the spectral current density, which describes whether the propagating carriers emit/lose energy or tunnel through the structure. But the exact evaluation of the relevant propagators is as difficult as the solution of the many-body Schrödinger's equation. A systematic way to approximate propagators is based on the perturbative expansion of the Green's function written in the interaction picture [101] in powers of the perturbation (interacting) Hamiltonian. As propagators obey wave equations with a singular inhomogeneity, i.e., a delta function in the variables (see the following discussion), they may be considered a generalized version of classical Green's functions.

NEGF theory was developed as an extension of zero- and finite-temperature Green's function techniques, probably inspired by their common diagrammatic formulation. The zero-temperature formalism is based on the Gell–Mann and Low theorem that relates the (unknown) interacting ground state of the system to the noninteracting state defined at $t \to -\infty$, before the interactions are adiabatically switched on. The time evolution operator that generates the series evolves on the real-time axis from $t \to -\infty$ to $t \to +\infty$, when the system returns to the noninteracting ground state [101]. In the finite-temperature formalism, introduced by Matsubara (1955), the ground state average appearing in the definition of the

Green's function is replaced by an average over a grand canonical ensemble, as the system is statistically distributed over all its excited levels. Interpreting the Boltzmann weight factor that appears in the definition of the statistical average as the evolution of the system for imaginary times, the resulting Green's function can be expanded diagramatically exactly as in the zero temperature case, but with time integrations restricted to a vertical path in the complex plane [101–103]. Although well suited for the description of reservoirs responsible for injecting carriers in a device, Matsubara Green's functions are not applicable to carrier transport, since a system driven out of equilibrium by an external (possibly time-dependent) field may not return to its initial state for asymptotically large times, i.e., the so-called "vacuum stability condition," a crucial assumption in equilibrium Green's function (EGF) theory, does not hold. A general NEGF formalism (beyond linear response theory) was established independently[†] by Kadanoff and Baym [104] (1962) and Keldysh [105] (1965) under the influence of Schwinger's school [106]. The expectation value of a given observable was reinterpreted as the system evolution on a contour along the real-time axis running from the remote past ($t = -\infty$), when the system was at equilibrium, to the highest relevant time of the evaluated Green's function, and back to the remote past. The round-trip or contour approach was the key to avoid the ill-defined state at $t = +\infty$, leading to contour-ordered Green's functions that have the usual diagrammatic expansion [107].

We summarize here only the relevant equations of the NEGF formulation, restricting our attention to steady-state conditions, and assuming that the reader is familiar with the occupation number formalism (second quantization),[†] our objective being to give a glimpse of the general assumptions underlying NEGF, and, whenever possible, to make connections with the drift-diffusion model presented in Section 2.2.

Consider a system evolving under the Hamiltonian operator $\hat{\mathcal{H}} = \hat{H} + \hat{H}^{\text{ext}}(t)$. The time-independent Hamiltonian $\hat{H} = \hat{H}_0 + \hat{H}^i$ describes the isolated system, where $\hat{H}_0$ is the exactly solvable part that contains the lattice and the electrostatic potential, and $\hat{H}^i$ is the complicated part that contains the many-body aspects of the problem, e.g., carrier–carrier and carrier–phonon interactions. $\hat{H}^{\text{ext}}(t)$ is the explicitly time-dependent perturbation that drives the system out of equilibrium. The perturbation $\hat{H}^{\text{ext}}(t)$ is applied at times $t > t_0$, and could be, e.g., an electric field, a light excitation pulse, or an imbalance of the chemical potentials of the reservoirs coupled to the system.

The task is to calculate the expectation value of a given observable, to which one associates a quantum mechanical operator $\hat{O}$, for times $t \geq t_0$. The nonequilibrium ensemble average with respect to the grand-canonical ensemble is defined as $\langle \hat{O}(t) \rangle = \text{Tr}[\rho(\hat{H})\hat{O}_{\mathcal{H}}(t)]$. The subscript $\mathcal{H}$ indicates that the time-dependence is governed by the full Hamiltonian, i.e., $\hat{O}$ is written in the Heisenberg picture. The thermal equilibrium DM $\rho(\hat{H}) = \mathcal{Z}^{-1} e^{-\beta \hat{H}}$ describes the system before the perturbation is turned on. Here $\mathcal{Z} = \text{Tr}(e^{-\beta \hat{H}})$ is the grand partition function, $\beta = 1/k_{\text{B}}T$ is the inverse temperature measured in energy units, and Tr stands for trace. We can simplify the time dependence of any operator moving from the Heisenberg picture to the interaction picture with respect to the noninteracting Hamiltonian $\hat{H}_0$

$$\hat{O}_{\mathcal{H}}(t) = \hat{S}^{\dagger}(t, t_0)\hat{O}_{H_0}(t)\hat{S}(t, t_0). \tag{2.66}$$

[†] A brief scientific/historical review of early Russian works on the Green's functions applications to many-body theory is presented in [108].

[†] We refer the reader to the excellent textbooks by Fetter and Walecka [102], Mahan [101], and Mattuck [103] for a detailed description of EGF theory. The NEGF formalism is described in [104,107,109–111]. Although EGF and NEGF theories are closely related, they are usually presented as separate topics, following the chronological/historical order as they originally appeared in the literature. A unified (nonchronological) description of EGF and NEGF formalisms and of their connection is presented by Stefanucci [111] and Rammer [110]. A notable attempt to make the formalism more accessible to readers not familiar with advanced many-body perturbation theory may be found in Datta [112].

The S-matrix operator evolving the interaction-picture wavefunction from t_0 to t (t_0 being the time instant at which the Heisenberg and interaction representations coincide) can be formally written as

$$\hat{S}(t,t_0) = \hat{T}\left\{ \exp\left[-\frac{\mathrm{i}}{\hbar}\int_{t_0}^{t} \mathrm{d}\tau \left(\hat{H}^{i}_{H_0}(\tau) + \hat{H}^{\mathrm{ext}}_{H_0}(\tau)\right)\right]\right\}, \tag{2.67}$$

with $\hat{H}^{i,\mathrm{ext}}_{H_0} = e^{\mathrm{i}\hat{H}_0(t-t_0)}\hat{H}^{i,\mathrm{ext}}e^{-\mathrm{i}\hat{H}_0(t-t_0)}$. When applied to a product of operators, the chronological time-ordering operator $\hat{T}$ orders them with the earliest times appearing on the right and the latest on the left. Considering that operators commute under operator ordering, we may also write

$$\hat{S}(t,t_0) = \hat{T}\left\{ \exp\left[-\frac{\mathrm{i}}{\hbar}\int_{t_0}^{t} \mathrm{d}\tau \hat{H}^{i}_{H_0}(\tau)\right] \exp\left[-\frac{\mathrm{i}}{\hbar}\int_{t_0}^{t} \mathrm{d}\tau \hat{H}^{\mathrm{ext}}_{H_0}(\tau)\right]\right\}. \tag{2.68}$$

It is understood that the exponential expressions in Equation 2.68 are a shorthand for series of time-ordered products of operators

$$\begin{aligned} &\hat{T}\left\{ \exp\left[-\frac{\mathrm{i}}{\hbar}\int_{t_0}^{t} \mathrm{d}\tau \hat{H}^{i}_{H_0}(\tau)\right]\right\} \\ &= \sum_{n=0}^{\infty}\left(\frac{-\mathrm{i}}{\hbar}\right)^{n}\frac{1}{n!}\int_{t_0}^{t}\mathrm{d}t_1\int_{t_0}^{t}\mathrm{d}t_2\cdots\int_{t_0}^{t}\mathrm{d}t_n\, \hat{T}\left\{\hat{H}^{i}_{H_0}(t_1)\hat{H}^{i}_{H_0}(t_2)\cdots\hat{H}^{i}_{H_0}(t_n)\right\}, \end{aligned} \tag{2.69}$$

which represents the starting point of the many-body perturbation expansion. In Equation 2.69, the term $n = 0$ is the unit operator and the $\hat{T}$ operator indicates that the products of the Hamiltonians at different times in the series expansion that defines the exponential must be kept in the correct chronological order (see [101, p. 67]). It is convenient to merge the two transformation integrals in (2.66) into the contour integral

$$\hat{O}_{H}(t) = \hat{T}_C\left\{ \exp\left[-\frac{\mathrm{i}}{\hbar}\int_{C} \mathrm{d}\tau \hat{H}^{i}_{H_0}(\tau)\right] \exp\left[-\frac{\mathrm{i}}{\hbar}\int_{C} \mathrm{d}\tau \hat{H}^{\mathrm{ext}}_{H_0}(\tau)\right]\hat{O}_{H_0}(t)\right\}, \tag{2.70}$$

where the time-ordering operator $\hat{T}_C$ orders the operators along the contour C of Figure 2.3, running from t_0 to t (chronological branch) and then back to t_0 (antichronological branch), so that operators with time labels that occur later on the contour stand to the left of operators that occur earlier; in other words, $\hat{T}_C$ reduces to the chronological or antichronological ordering operator for the part of the contour running

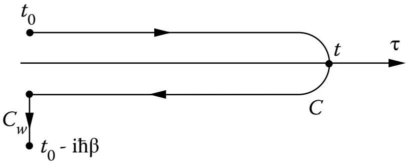


FIGURE 2.3 The contour $\tilde{C} = C \cup C_w$ for the transformation to the interaction picture where the time evolution is governed by the noninteracting Hamiltonian H_0. The contour C stretches from t_0 to t and then back to t_0. The small shift from the real axis is only meant to distinguish the two branches and to clarify the integration direction. The appendix C_w stretching into the lower complex half-plane from t_0 to $t_0 - \mathrm{i}\hbar\beta$ is required when initial correlations are of interest; in steady-state conditions, we can neglect the contribution of the imaginary appendix C_w by taking the limit $t_0 \to -\infty$. (From J. Rammer., *Quantum Field Theory of Non-equilibrium States,* Cambridge University Press, Cambridge, 2007.)

forward or backward in time, respectively. The nonequilibrium statistical average of an observable $\hat{O}_{\mathcal{H}}(t)$ may now be written as

$$\begin{aligned}\langle \hat{O}_{\mathcal{H}}(t)\rangle &= \mathrm{Tr}(\rho(\hat{H})\hat{O}_{\mathcal{H}}(t))\\ &= \frac{\mathrm{Tr}\left(e^{-\beta\hat{H}}\hat{T}_C\left\{\exp\left[-\frac{\mathrm{i}}{\hbar}\int_C \mathrm{d}\tau \hat{H}^i_{H_0}(\tau)\right]\exp\left[-\frac{\mathrm{i}}{\hbar}\int_C \mathrm{d}\tau \hat{H}^{\mathrm{ext}}_{H_0}(\tau)\right]\hat{O}_{H_0}(t)\right\}\right)}{\mathrm{Tr}(e^{-\beta\hat{H}})}.\end{aligned} \tag{2.71}$$

The Boltzmann weight factor can be expressed in terms of $\hat{H}_0$ by means of an imaginary time-evolution operator

$$e^{-\beta\hat{H}} = e^{-\beta\hat{H}_0}\,\hat{T}_{C_w}\left\{\exp\left[-\frac{\mathrm{i}}{\hbar}\int_{t_0}^{t_0-\mathrm{i}\hbar\beta}\mathrm{d}\tau\hat{H}^i_{H_0}(\tau)\right]\right\}, \tag{2.72}$$

where $\hat{T}_{C_w}$ orders operators along the vertical track stretching into the lower complex half-plane from t_0 to $t_0 - \mathrm{i}\hbar\beta$; see Figure 2.3. Upon substitution of Equation 2.72 in Equation 2.71, $\hat{T}_{C_w}$ and $\hat{T}_C$ can be merged into one contour operator $\hat{T}_{\tilde{C}}$ defined on $\tilde{C} = C \cup C_w$, which gives

$$\langle \hat{O}_{\mathcal{H}}(t)\rangle = \frac{\mathrm{Tr}\left(e^{-\beta\hat{H}_0}\hat{T}_{\tilde{C}}\left\{\exp\left[-\frac{\mathrm{i}}{\hbar}\int_{\tilde{C}}\mathrm{d}\tau\hat{H}^i_{H_0}(\tau)\right]\exp\left[-\frac{\mathrm{i}}{\hbar}\int_C\mathrm{d}\tau\hat{H}^{\mathrm{ext}}_{H_0}(\tau)\right]\right\}\hat{O}_{H_0}(t)\right)}{\mathrm{Tr}\left(e^{-\beta\hat{H}_0}\hat{T}_{\tilde{C}}\left\{\exp\left[-\frac{\mathrm{i}}{\hbar}\int_{\tilde{C}}\mathrm{d}\tau\hat{H}^i_{H_0}(\tau)\right]\exp\left[-\frac{\mathrm{i}}{\hbar}\int_C\mathrm{d}\tau\hat{H}^{\mathrm{ext}}_{H_0}(\tau)\right]\right\}\right)} \tag{2.73}$$

$$= \frac{\langle\hat{T}_{\tilde{C}}\left\{\exp\left[-\frac{\mathrm{i}}{\hbar}\int_{\tilde{C}}\mathrm{d}\tau\hat{H}^i_{H_0}(\tau)\right]\exp\left[-\frac{\mathrm{i}}{\hbar}\int_C\mathrm{d}\tau\hat{H}^{\mathrm{ext}}_{H_0}(\tau)\right]\hat{O}_{H_0}(t)\right\}\rangle_0}{\langle\hat{T}_{\tilde{C}}\left\{\exp\left[-\frac{\mathrm{i}}{\hbar}\int_{\tilde{C}}\mathrm{d}\tau\hat{H}^i_{H_0}(\tau)\right]\exp\left[-\frac{\mathrm{i}}{\hbar}\int_C\mathrm{d}\tau\hat{H}^{\mathrm{ext}}_{H_0}(\tau)\right]\right\}\rangle_0}, \tag{2.74}$$

where all the time dependence is governed by the exactly solvable Hamiltonian $\hat{H}_0$, and $\langle\ldots\rangle_0$ stands for statistical average with respect to the noninteracting DM $\hat{\rho}_0$

$$\langle\ldots\rangle_0 = \frac{\mathrm{Tr}\left(\exp(-\beta\hat{H}_0)\ldots\right)}{\mathrm{Tr}\left(\exp(-\beta\hat{H}_0)\right)}. \tag{2.75}$$

In the denominator of Equation 2.73, we have introduced the identity evolution operator defined along a closed contour that connects the initial state to itself (no operators acting at intermediate times)

$$\hat{T}_C\left\{\exp\left[-\frac{\mathrm{i}}{\hbar}\int_C\mathrm{d}\tau(\hat{H}^i_{H_0}(\tau)+\hat{H}^{\mathrm{ext}}_{H_0}(\tau))\right]\right\} = \hat{1}, \tag{2.76}$$

as required by combinatorial/diagrammatic arguments to cancel the disconnected diagrams of the numerator. Notice that the external perturbation is zero on the imaginary part of contour C_w. There is no restriction concerning the choice of t_0. If we are not interested in transient phenomena, we can assume that all transients have died out by letting $t_0 \rightarrow -\infty$, and the contribution from the imaginary part of the contour C_w vanishes, meaning that $\tilde{C} = C$. Moreover, we can extend the contour to $t = \infty$, since, beyond the largest time, the forward and backward evolutions simply produce the identity operator. We then obtain the Schwinger–Keldysh contour, which starts at $t = -\infty$ and proceeds to $t = \infty$ and then back again to $t = -\infty$; see Figure 2.4.

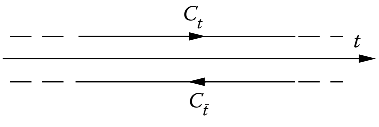


FIGURE 2.4 The Schwinger–Keldysh contour $C = C_t \cup C_{\tilde{t}}$. The small shift from the real axis of the chronological and antichronological branches $C_t = (-\infty, \infty)$ and $C_{\tilde{t}} = (\infty, -\infty)$ is only meant to distinguish the integration direction.

The previous considerations lead us to the analysis of the single-particle NEGF defined as the nonequilibrium ensemble average of contour-ordered field operators:

$$\begin{aligned} G(\underline{r}, t; \underline{r}', t') &= -\frac{\mathrm{i}}{\hbar}\langle \hat{T}_C\{\hat{\psi}_{\mathcal{H}}(\underline{r}, t)\hat{\psi}^{\dagger}_{\mathcal{H}}(\underline{r}', t')\}\rangle \\ &= -\frac{\mathrm{i}}{\hbar}\left(\theta_C(t, t')\langle \hat{\psi}_{\mathcal{H}}(\underline{r}, t)\hat{\psi}^{\dagger}_{\mathcal{H}}(\underline{r}', t')\rangle \mp \theta_C(t', t)\langle \hat{\psi}^{\dagger}_{\mathcal{H}}(\underline{r}', t')\hat{\psi}_{\mathcal{H}}(\underline{r}, t)\rangle\right). \end{aligned} \tag{2.77}$$

The function $\theta_C(t, t')$ is the Heaviside function defined on the contour, with $\theta_C(t, t') = 1$, if t is later on C than t', and zero otherwise. The upper (lower) sign in $\mp$ is for fermions (bosons); from now on, we restrict our attention to fermionic systems. The Heisenberg field operator $\hat{\psi}^{\dagger}_{\mathcal{H}}(\underline{r}', t')$ $(\hat{\psi}_{\mathcal{H}}(\underline{r}, t))$ creates (annihilates) a particle at position $\underline{r}'$ $(\underline{r})$ and at time t' (t). The NEGF $G(\underline{r}, t; \underline{r}', t')$ expresses a correlation between two times t and t' and two positions $\underline{r}$ and $\underline{r}'$. For t later on C than t', $G(\underline{r}, t; \underline{r}', t')$ can be interpreted as the reaction of a system to a particle created at position $\underline{r}'$ and time t', to the propagation of this perturbation to position $\underline{r}$ and time t, and finally to its annihilation there. In other words, we probe the interacting system by adding one particle, which propagates and interacts with the other particles until we remove it from the system. A similar interpretation holds if t' is later on C than t. Applying Equation 2.74 to the average in Equation 2.77 provides the starting point of the perturbation expansion of the NEGF

$$G(\underline{r}, t; \underline{r}', t') = \frac{\langle \hat{T}_{\tilde{C}}\left\{\exp\left[-\frac{\mathrm{i}}{\hbar}\int_{\tilde{C}} \mathrm{d}\tau \hat{H}^{i}_{H_0}(\tau)\right]\exp\left[-\frac{\mathrm{i}}{\hbar}\int_{C} \mathrm{d}\tau \hat{H}^{\mathrm{ext}}_{H_0}(\tau)\right]\hat{\psi}_{H_0}(\underline{r}, t)\hat{\psi}^{\dagger}_{H_0}(\underline{r}', t')\right\}\rangle_0}{\langle \hat{T}_{\tilde{C}}\left\{\exp\left[-\frac{\mathrm{i}}{\hbar}\int_{\tilde{C}} \mathrm{d}\tau \hat{H}^{i}_{H_0}(\tau)\right]\exp\left[-\frac{\mathrm{i}}{\hbar}\int_{C} \mathrm{d}\tau \hat{H}^{\mathrm{ext}}_{H_0}(\tau)\right]\right\}\rangle_0}. \tag{2.78}$$

The following discussion is based on the assumption that initial correlations are absent, in other words, the time evolution is on the Schwinger–Keldysh contour C. The time evolution of $G(\underline{r}, t; \underline{r}', t')$ with respect to the times t and t' is obtained from the corresponding equations of motion in the Heisenberg picture. A Hamiltonian operator that includes a carrier–carrier interaction $\hat{H}^i = \hat{V}$, and a single-particle potential $\hat{H}^{\mathrm{ext}} = \hat{U}$, reads in second quantization

$$\begin{aligned} \hat{\mathcal{H}}(t) &= \int \mathrm{d}\underline{r}\, \hat{\psi}^{\dagger}_{\mathcal{H}}(\underline{r}, t)[H_0(\underline{r}) + U(\underline{r}, t)]\hat{\psi}_{\mathcal{H}}(\underline{r}, t) \\ &+ \frac{1}{2}\int \mathrm{d}\underline{r} \int \mathrm{d}\underline{r}'\hat{\psi}^{\dagger}_{\mathcal{H}}(\underline{r}, t)\hat{\psi}^{\dagger}_{\mathcal{H}}(\underline{r}', t)V(\underline{r} - \underline{r}')\hat{\psi}_{\mathcal{H}}(\underline{r}', t)\hat{\psi}_{\mathcal{H}}(\underline{r}, t). \end{aligned} \tag{2.79}$$

The equation of motion describing the time evolution of the NEGF can be determined from the derivative of the contour-ordered pair of Heisenberg field operators

$$\begin{aligned} \frac{\partial}{\partial t}\langle \hat{T}_C\{\hat{\psi}_{\mathcal{H}}(\underline{r}, t)\hat{\psi}^{\dagger}_{\mathcal{H}}(\underline{r}', t')\}\rangle &= \left\langle \hat{T}_C\left\{\left[\frac{\partial}{\partial t}\hat{\psi}_{\mathcal{H}}(\underline{r}, t)\right]\hat{\psi}^{\dagger}_{\mathcal{H}}(\underline{r}', t')\right\}\right\rangle + \delta_C(t, t')\left[\hat{\psi}_{\mathcal{H}}(\underline{r}, t)\hat{\psi}^{\dagger}_{\mathcal{H}}(\underline{r}', t')\right]_{+} \\ &= \left\langle \hat{T}_C\left\{\left[\frac{\partial}{\partial t}\hat{\psi}_{\mathcal{H}}(\underline{r}, t)\right]\hat{\psi}^{\dagger}_{\mathcal{H}}(\underline{r}', t')\right\}\right\rangle + \delta_C(t, t')\delta(\underline{r} - \underline{r}'). \end{aligned} \tag{2.80}$$

In the last line of Equation 2.80 the equal-time anticommutation property $\left[\hat{\psi}_{\mathcal{H}}(\underline{r},t)\hat{\psi}^{\dagger}_{\mathcal{H}}(\underline{r}',t)\right]_{+} = \delta(\underline{r}-\underline{r}')$ was used. To evaluate the first RHS term in Equation 2.80, it is necessary to know the time evolution of the annihilation operator, whose time derivative can be calculated from the corresponding Heisenberg equation of motion

$$\begin{aligned}
\mathrm{i}\hbar\frac{\partial}{\partial t}\hat{\psi}_{\mathcal{H}}(\underline{r},t) &= \left[\hat{\psi}_{\mathcal{H}}(\underline{r},t),\hat{\mathcal{H}}(t)\right]_{-} \\
&= [H_0(\underline{r}) + U(\underline{r},t)]\hat{\psi}_{\mathcal{H}}(\underline{r},t) + \int \mathrm{d}\underline{r}'V(\underline{r}-\underline{r}')\hat{\psi}^{\dagger}_{\mathcal{H}}(\underline{r}',t)\hat{\psi}_{\mathcal{H}}(\underline{r}',t)\hat{\psi}_{\mathcal{H}}(\underline{r},t).
\end{aligned} \tag{2.81}$$

Replacing in Equation 2.80 the time-derivative Equation 2.81 gives the equation of motion of $G(\underline{r}_1,t_1;\underline{r}_{1'},t_{1'}) = G(11')$ relative to t_1 (a similar procedure gives the derivative relative to $t_{1'}$)

$$\begin{aligned}
\left[\mathrm{i}\hbar\frac{\partial}{\partial t_1} - H_0(\underline{r}_1) - U(1)\right]G(11') &= \delta(11') - \mathrm{i}\hbar\int_C \mathrm{d}2V(1-2)G^{(2)}(121'2^{+}) \\
\left[-\mathrm{i}\hbar\frac{\partial}{\partial t_{1'}} - H_0(\underline{r}_{1'}) - U(1')\right]G(11') &= \delta(11') - \mathrm{i}\hbar\int_C \mathrm{d}2V(1'-2)G^{(2)}(12^{-}1'2),
\end{aligned} \tag{2.82}$$

where the two-particle Green's function is defined as

$$G^{(2)}(121'2') = \left(\frac{-\mathrm{i}}{\hbar}\right)^2\left\langle \hat{T}_C\{\hat{\psi}_{\mathcal{H}}(1)\hat{\psi}_{\mathcal{H}}(2)\hat{\psi}^{\dagger}_{\mathcal{H}}(2')\hat{\psi}^{\dagger}_{\mathcal{H}}(1')\}\right\rangle. \tag{2.83}$$

The notation $2^{\pm}$ in Equation 2.82 is intended as a remainder that the time argument of $\hat{\psi}^{\dagger}_{\mathcal{H}}(2)$ must be chosen to be infinitesimally larger or smaller than t_2 [104]. In writing Equation 2.82, we have used the shorthand notation $\int_C \mathrm{d}1 = \int_C \mathrm{d}t_1 \int \mathrm{d}\underline{r}_1$ and the definitions

$$\delta(12) = \delta_C(t_1,t_2)\delta(\underline{r}_1-\underline{r}_2) \tag{2.84}$$

$$V(1-2) = V(\underline{r}_1-\underline{r}_2)\delta(t_1-t_2). \tag{2.85}$$

Equation 2.82 is not closed because the interaction couples the one-particle GF to the two-particle GF, etc., leading to an infinite hierarchy of coupled equations involving GFs of ever-increasing order. Rather than adopting the hierarchical approach employed in Section 2.2 for the solution of the BTE, we can formally decouple the Martin–Schwinger hierarchy by approximating the two-particle Green's function by means of the self-energy $\Sigma(11')$ encoding interaction effects of all particles on the single-particle dynamics

$$\left[\mathrm{i}\hbar\frac{\partial}{\partial t_1} - H_0(\underline{r}_1) - U(1)\right]G(11') = \delta(11') + \int_C \mathrm{d}3\Sigma(13)G(31') \tag{2.86}$$

$$\left[-\mathrm{i}\hbar\frac{\partial}{\partial t_{1'}} - H_0(\underline{r}_{1'}) - U(1')\right]G(11') = \delta(11') + \int_C \mathrm{d}3G(13)\Sigma(31'). \tag{2.87}$$

The self-energy may be obtained to different levels of approximations (HF, self-consistent Born approximation [SCBA], T-matrix, etc.) using the Feynman diagrammatic expansion or functional derivatives. Considering that the noninteracting GF $G_0(11')$ is the solution of Equation 2.82 without the two-particle GF

$$G_0(11') = \left[\mathrm{i}\hbar\frac{\partial}{\partial t_1} - H_0(\underline{r}_1) - U(1)\right]^{-1}\delta(11'), \tag{2.88}$$

and integrating over one of the times to eliminate the δ-functions, one obtains the two equivalent forms of the Dyson's equations in integral form

$$G(11') = G_0(11') + \int_C \mathrm{d}2 \int_C \mathrm{d}3\, G_0(12)\Sigma(23)G(31') \tag{2.89}$$

$$G(11') = G_0(11') + \int_C \mathrm{d}2 \int_C \mathrm{d}3\, G(12)\Sigma(23)G_0(31'). \tag{2.90}$$

The equations of motion for the NEGF $G(11')$ include integrals over a complex time contour $\int_C$. Since it is cumbersome to keep track of the time branch in the evaluation of contour integrals, four new Green's functions with real-time arguments are introduced,

$$G(11') = \begin{cases} G^t(11') & t_1, t_{1'} \in C_t \\ G^{\bar{t}}(11') & t_1, t_{1'} \in C_{\bar{t}} \\ G^<(11') & t_1 \in C_t, t_{1'} \in C_{\bar{t}} \\ G^>(11') & t_1 \in C_{\bar{t}}, t_{1'} \in C_t, \end{cases} \tag{2.91}$$

which are named the chronological, antichronological, lesser and greater Green's functions, respectively, and are defined by

$$G^t(11') = -\frac{\mathrm{i}}{\hbar}\langle \hat{T}_t\{\hat{\psi}_{\mathcal{H}}(1)\hat{\psi}^\dagger_{\mathcal{H}}(1')\}\rangle \tag{2.92}$$

$$G^{\bar{t}}(11') = -\frac{\mathrm{i}}{\hbar}\langle \hat{T}_{\bar{t}}\{\hat{\psi}_{\mathcal{H}}(1)\hat{\psi}^\dagger_{\mathcal{H}}(1')\}\rangle \tag{2.93}$$

$$G^<(11') = +\frac{\mathrm{i}}{\hbar}\langle \hat{\psi}^\dagger_{\mathcal{H}}(1')\hat{\psi}_{\mathcal{H}}(1)\rangle \tag{2.94}$$

$$G^>(11') = -\frac{\mathrm{i}}{\hbar}\langle \hat{\psi}_{\mathcal{H}}(1)\hat{\psi}^\dagger_{\mathcal{H}}(1')\rangle. \tag{2.95}$$

The chronological and antichronological Green's functions are usually replaced by the retarded and advanced Green's functions defined by

$$G^R(11') = -\frac{\mathrm{i}}{\hbar}\theta(t-t')\langle[\hat{\psi}_{\mathcal{H}}(1), \hat{\psi}^\dagger_{\mathcal{H}}(1')]_+\rangle = \theta(t-t')[G^>(11') - G^<(11')] \tag{2.96}$$

$$G^A(11') = \frac{\mathrm{i}}{\hbar}\theta(t'-t)\langle[\hat{\psi}_{\mathcal{H}}(1), \hat{\psi}^\dagger_{\mathcal{H}}(1')]_+\rangle = -\theta(t'-t)[G^>(11') - G^<(11')]. \tag{2.97}$$

Subtracting the last two equations, we have

$$G^R(11') - G^A(11') = G^>(11') - G^<(11'), \tag{2.98}$$

which shows that only three of the Green's functions are linearly independent. The contour integrations in Equation 2.90 can be decomposed into real-time integrations by means of the Langreth rules [107]. Neglecting initial correlations, which are relevant only for the transient behavior, we obtain the equations of motion of the real-time Green's functions

$$G^{R(A)}(11') = G_0^{R(A)}(11') + \int \mathrm{d}2 \int \mathrm{d}3 G_0^{R(A)}(12)\Sigma^{R(A)}(23)G^{R(A)}(31') \tag{2.99}$$

$$G^{\lessgtr}(11') = \int \mathrm{d}2 \int \mathrm{d}3 G^R(12)\Sigma^{\lessgtr}(23)G^A(31'), \tag{2.100}$$

with $\int \mathrm{d}1 = \int \mathrm{d}\underline{r}_1 \int_{-\infty}^{+\infty} \mathrm{d}t_1$. For steady-state calculations, only the time difference $\tau = t' - t$ is meaningful. Fourier transforming $O^\alpha = G^\alpha, \Sigma^\alpha$ $(\alpha = R, A, <, >)$ to energy coordinates

$$O^\alpha(\underline{r}_1, \underline{r}_2, E) = \int_{-\infty}^{+\infty} \mathrm{d}\tau e^{\frac{\mathrm{i}}{\hbar}E\tau} O^\alpha(\underline{r}_1, \underline{r}_2, \tau) \tag{2.101}$$

$$O^\alpha(\underline{r}_1, \underline{r}_2, \tau) = \frac{1}{2\pi\hbar} \int_{-\infty}^{+\infty} \mathrm{d}E e^{-\frac{\mathrm{i}}{\hbar}E\tau} O^\alpha(\underline{r}_1, \underline{r}_2, E), \tag{2.102}$$

we may rewrite Equations 2.99 and 2.100 as

$$\begin{aligned} G^{R(A)}(\underline{r}_1, \underline{r}_{1'}, E) = {} & G_0^{R(A)}(\underline{r}_1, \underline{r}_{1'}, E) \\ & + \int \mathrm{d}\underline{r}_2 \int \mathrm{d}\underline{r}_3 G_0^{R(A)}(\underline{r}_1, \underline{r}_2, E) \Sigma^{R(A)}(\underline{r}_2, \underline{r}_3, E) G^{R(A)}(\underline{r}_3, \underline{r}_{1'}, E) \end{aligned} \tag{2.103}$$

$$G^{\lessgtr}(\underline{r}_1, \underline{r}_{1'}, E) = \int \mathrm{d}\underline{r}_2 \int \mathrm{d}\underline{r}_3 G^R(\underline{r}_1, \underline{r}_2, E) \Sigma^{\lessgtr}(\underline{r}_2, \underline{r}_3, E) G^A(\underline{r}_3, \underline{r}_{1'}, E) \tag{2.104}$$

with the noninteracting Green function given by

$$G_0^{R(A)}(\underline{r}_1, \underline{r}_{1'}, E) = [E + (-)\mathrm{i}\eta - H_0(\underline{r}_1)]^{-1} \delta(\underline{r}_1 - \underline{r}_{1'}). \tag{2.105}$$

The small $\eta \to 0^+$ parameter provides the correct analytical properties. Time-reversal symmetry implies $G^R(\underline{r}_1, \underline{r}_{1'}, E) = [G^A(\underline{r}_1, \underline{r}_{1'}, E)]^\dagger$, and $G^{\lessgtr}(\underline{r}_1, \underline{r}_{1'}, E) = -[G^{\lessgtr}(\underline{r}_1, \underline{r}_{1'}, E)]^\dagger$, which reduces the number of independent Green's functions to two.

The Green's functions are directly related to physical observables. In particular, the expressions for the electron and hole densities in steady state follow directly from the definition of the density operator [107]

$$n(\underline{r}) = -\mathrm{i} \int \frac{\mathrm{d}E}{2\pi} G^<(\underline{r}, \underline{r}, E) \tag{2.106}$$

$$p(\underline{r}) = \mathrm{i} \int \frac{\mathrm{d}E}{2\pi} G^>(\underline{r}, \underline{r}, E). \tag{2.107}$$

Taking the difference of the equations of motion (2.86) and (2.87), in the limit $1' \to 1$, gives

$$\begin{aligned} & \underbrace{\lim_{1' \to 1} \left\{ \mathrm{i}\hbar \left(\frac{\partial}{\partial t_1} + \frac{\partial}{\partial t_{1'}} \right) G(11') \right\}}_{\frac{\partial \rho}{\partial t}(\underline{r}, \mathrm{t})} - \underbrace{\lim_{1' \to 1} \left\{ [H_0(\underline{r}_1) - H_0(\underline{r}_{1'})] G(11') \right\}}_{\nabla \cdot \mathrm{J}(\underline{r}, \mathrm{t})} = \\ & \underbrace{\lim_{1' \to 1} \int \mathrm{d}2\, [\Sigma(12) G(21') - G(12) \Sigma(21')]}_{-\mathrm{U}_\rho(\underline{r}, \mathrm{t})}, \end{aligned} \tag{2.108}$$

which may be identified with the DD continuity equation 2.21. In steady state, Equation 2.108 reads [113]

$$\begin{aligned} \nabla \cdot \underline{J}(\underline{r}) = -\int \frac{\mathrm{d}E}{2\pi\hbar} \int \mathrm{d}\underline{r}' [& \Sigma^R(\underline{r}, \underline{r}', E) G^<(\underline{r}', \underline{r}, E) + \Sigma^<(\underline{r}, \underline{r}', E) G^A(\underline{r}', \underline{r}, E) \\ & - G^R(\underline{r}, \underline{r}', E) \Sigma^<(\underline{r}', \underline{r}, E) - G^<(\underline{r}, \underline{r}', E) \Sigma^A(\underline{r}', \underline{r}, E)]. \end{aligned} \tag{2.109}$$

The RHS of Equation 2.109 should vanish when it is integrated over all bands, if the self-energies are properly chosen to ensure current conservation.[†] Restricting the integration to the conduction (valence) band, Equation 2.109 can be identified as the NEGF version of the semiclassical steady-state electron (hole) continuity equation, with the RHS of Equation 2.109 corresponding to the net electron (hole) generation rate. Because the self-energies are addictive, one can separate the recombination rates due to different mechanisms.

For the numerical evaluation of the Dyson's equations, we need to express the Green's functions defined on the continuous space variables $\underline{r}$ and $\underline{r}'$ in the discrete basis adopted in Section 2.3, which consists of a product of fast-varying (over the crystal primitive cell) zone-center Bloch functions u_{m0}, plane waves in the homogenous transverse direction $\underline{\rho}$, and shape functions N_i slowly varying in the crystal growth direction z

$$\psi_{\alpha\underline{k}}(\underline{r}) = \frac{1}{\sqrt{\mathcal{A}}} N_i(z) e^{i\underline{k}\cdot\underline{\rho}} u_{m0}(\underline{r}), \tag{2.110}$$

where the compound index $\alpha = (i, m)$ combines indices i for space and m for band, and $\mathcal{A}$ is the normalization cross-section area. Expanding the field operators on this basis

$$\hat{\psi}_{\mathcal{H}}(\underline{r}, t) = \sum_{\alpha,\underline{k}} \psi_{\alpha\underline{k}}(\underline{r}) \hat{c}_{\alpha,\underline{k}}(t) \tag{2.111}$$

$$\hat{\psi}^{\dagger}_{\mathcal{H}}(\underline{r}, t) = \sum_{\alpha,\underline{k}} \psi^{*}_{\alpha\underline{k}}(\underline{r}) \hat{c}^{\dagger}_{\alpha,\underline{k}}(t) \tag{2.112}$$

leads to the contravariant and covariant matrix representations $O_{\alpha\beta}(\underline{k}, E)$, $\tilde{O}_{\alpha\beta}(\underline{k}, E)$ of the steady-state fields $O = G, \Sigma$[†]

$$O(\underline{r}, \underline{r}', E) = \frac{1}{\mathcal{A}} \sum_{\underline{k}} \sum_{\alpha\beta} \Psi_{\alpha\underline{k}}(\underline{r}) O_{\alpha\beta}(\underline{k}, E) \Psi^{*}_{\beta\underline{k}}(\underline{r}') \tag{2.113}$$

$$\tilde{O}_{\alpha\beta}(\underline{k}, E) = \int \mathrm{d}\underline{r}\, \mathrm{d}\underline{r}'\; \Psi^{*}_{\alpha\underline{k}}(\underline{r}) O(\underline{r}, \underline{r}', E) \Psi_{\beta\underline{k}}(\underline{r}'). \tag{2.114}$$

Using a contravariant representation for the Green's functions and a covariant representation for the self-energies, one obtains a coupled set of linear equations that in matrix form read [115, Chapter 3]

$$[G^{R}(\underline{k}, E)] = (E[M] - [H(\underline{k})] - [\Sigma^{R}(\underline{k}, E)])^{-1} \tag{2.115}$$

$$[G^{A}(\underline{k}, E)] = [G^{R}(\underline{k}, E)]^{\dagger} \tag{2.116}$$

$$[G^{\lessgtr}(\underline{k}, E)] = [G^{R}(\underline{k}, E)][\Sigma^{\lessgtr}(\underline{k}, E)][G^{A}(\underline{k}, E)], \tag{2.117}$$

with $M_{\alpha\beta} = \int \mathrm{d}\underline{r}\Psi^{*}_{\alpha\underline{k}}(\underline{r})\Psi_{\beta\underline{k}}(\underline{r}) \approx M_{ij}\delta_{mn}$ [69].

[†] Typical approximations of the self-energy (e.g., HF, Born, T-matrix, and random-phase-approximation) ensure the conservation for particle number, momentum, and energy. Conservation laws may be violated when an arbitrary diagram is included in the self-energy [114], or when self-energies are interpolated on an energy grid [69].

[†] The choice of an orthonormal basis suppresses the complication of having two different representations, since in that case contravariant and covariant components coincide.

2.4.2 Boundary Self-Energies

Having partitioned the nanostructure into device and reservoir regions, boundary self-energies accounting for the openness of the system may be evaluated by augmenting Dyson's equation with open BCs relating the normal derivative of the solution at the boundary to its boundary value [116]. Consider first the simple case of the effective mass approximation, which allows us to invert the energy dispersion relation analytically. As in Section 2.3, let us assume a 1D system, which occupies the simulation region along the z-axis with left and right contacts (or reservoirs) occupying the regions $z < z_1$ and $z > z_N$, respectively, defined as fully absorbing elements attempting to maintain thermal equilibrium and charge neutrality at the z_1 and z_N boundaries. Assuming that the electrostatic potential is constant in the semi-infinite leads, the wavefunction attempting to propagate (or decay) from the device to the leads can be written in terms of plane waves. For $z < z_1$, the outgoing wave is $\zeta_{out}(z) = A^- e^{ik_z^- z}$, where A^- is the unknown outgoing coefficient, $k_z^- = -\sqrt{2m^* E}/\hbar$ is the wave number associated with the (longitudinal) energy E, and m^* is the effective mass. Integrating by parts the second-order term in the Hamiltonian, Equation 2.26 gives the boundary term $-\frac{\hbar^2}{2m^*}\,\partial_z\zeta(z)\big|_{z=z_1}$. The continuity of the wavefunction and its derivative across the boundary gives the contribution $-\frac{\hbar^2}{2m^*}ik_z^-$ to the matrix element $[H]_{11}$, which is the only term that describes the coupling of the device to the contact, and therefore defines the upper-left corner of the otherwise zero boundary self-energy matrix of the left contact: $[\Sigma_{11}^{RB}(\underline{k}, E)] = -\frac{\hbar^2}{2m^*}ik_z^-$. Similarly, at the right reservoir, we express the outgoing wave as $\zeta_{out}(z) = A^+ e^{ik_z^+ z}$, propagating or decaying to the right at energy E and wave number $k^+ = \sqrt{2m^* E}/\hbar$, which gives the lower-right corner of the otherwise zero boundary self-energy matrix of the right contact: $[\Sigma_{NN}^{RB}(\underline{k}, E)] = -\frac{\hbar^2}{2m^*}ik_z^+$.

Moving from the effective mass approximation to a multiband $k \cdot p$ description of the electronic structure, the task of describing the coupling of the device to a semi-infinite domain on either side of the region of interest becomes a more complicated affair, as the outgoing waves may occur into any of the (propagating or evanescent) states k_z and bands n such that $E_n(\underline{k}, k_z) = E$. For example, the outgoing waves at the left reservoir are found by selecting only real k_z's such that the group velocity $v_g(k_z) < 0$ (reflected waves propagating to the left) and those complex k_z's for which $\mathrm{Im}(k_z) < 0$ (reflected waves decaying to the left). A similar argument holds for the right contact. The problem of finding the component k_z of the wave vector along the confining direction having defined the kinetic energy and the transverse wave vector is what is known as the *complex band structure problem*. In the conventional band structure problem, one fixes $(\underline{k}, k_z)$ and finds possible values of E by solving the eigenvalue problem $H(\underline{k}, k_z)\psi = E\psi$. In the complex band structure problem, one fixes the in-plane wave vector $\underline{k}$ and the energy E and finds possible values for k_z.[†] Recalling the decomposition of the bulk Hamiltonian (Equation 2.26), this procedure leads to a quadratic eigenvalue problem that can be cast into a linear one [117],

$$\begin{bmatrix} 0 & \mathbb{1} \\ H^{(0)} - E\mathbb{1} & H^{(1)} \end{bmatrix} \begin{pmatrix} \psi \\ k_z\psi \end{pmatrix} = k_z \begin{bmatrix} \mathbb{1} & 0 \\ 0 & -H^{(2)} \end{bmatrix} \begin{pmatrix} \psi \\ k_z\psi \end{pmatrix}. \qquad (2.118)$$

Complex eigenvalues occur in pairs, i.e., running (evanescent) states traverse the nanostructure with the same real (imaginary) part and opposite signs. If we classify the $2M$ solutions of the complex band structure problem at each lead according to their direction of motion and organize them in the columns of matrices $[\psi^{\pm}]$, we obtain a boundary self-energy matrix with all zero entries, save for the upper-left and lower-right $M \times M$ blocks

---

[†] The following discussion is valid whenever the Hamiltonian has a polynomial dependence on k_z as in Equation 2.26; it does not apply to nonlocal plane-wave pseudopotential Hamiltonians; see [118] for a more general approach to the calculation of *ab initio* complex band structures.

$$[\Sigma_{11}^{RB}(\underline{k}, E)] = -[H^{(2)}][\psi^{-}][\mathrm{diag}(\mathrm{i}k_{z1}^{-}, \mathrm{i}k_{z2}^{-}, \ldots, \mathrm{i}k_{zM}^{-})][\psi^{-}]^{-1} + [\mathrm{i}H_{L}^{(1)}] \tag{2.119a}$$

$$[\Sigma_{NN}^{RB}(\underline{k}, E)] = -[H^{(2)}][\psi^{+}][\mathrm{diag}(\mathrm{i}k_{z1}^{+}, \mathrm{i}k_{z2}^{+}, \ldots, \mathrm{i}k_{zM}^{+})][\psi^{+}]^{-1} + [\mathrm{i}H_{L}^{(1)}]. \tag{2.119b}$$

The expressions for $[\Sigma^{B\lessgtr}(\underline{k}, E)]$ follow from the fluctuation–dissipation theorem [101]. Assuming that the left contact is at Fermi level E_{FL}, we have

$$[\Sigma_{11}^{<B}(\underline{k}, E)] = \mathrm{i}f_{\mathrm{FD}}(E, E_{\mathrm{FL}})[\Gamma_{1}^{B}(\underline{k}, E)] \tag{2.120a}$$

$$[\Sigma_{11}^{>B}(\underline{k}, E)] = -\mathrm{i}(1 - f_{\mathrm{FD}}(E, E_{\mathrm{FL}}))[\Gamma_{1}^{B}(\underline{k}, E)], \tag{2.120b}$$

where $[\Gamma_{1}^{B}(\underline{k}, E)] = \mathrm{i}([\Sigma_{11}^{RB}(\underline{k}, E)] - [\Sigma_{11}^{RB}(\underline{k}, E)]^{\dagger})$ is the broadening function that quantifies the level broadening in the open system due to the connection to the reservoir. Similar expressions hold for the right contact.

2.4.3 Scattering Self-Energies

While boundary self-energies can be evaluated exactly within an FE scheme, scattering self-energies have to be computed perturbatively. Most of the complexity of the NEGF formalism and accordingly most of the numerical burden originate from the modeling of scattering self-energies, and it is for this reason that these quantities are usually described in NEGF studies with phenomenological models (e.g., Büttiker probes, phase- and momentum-relaxing Golizadeh self-energies) or approximated as local in space[†] and/or decoupled from Green's functions. Rather than using simplified models, we adopt the rigorous approach based on the SCBA.

The starting point is the perturbative expansion of Green's function (Equation 2.78), which includes time-ordered products of interaction-picture field operators averaged over the initial equilibrium ensemble. The equilibrium statistical average of time-ordered products can be decomposed in equilibrium statistical averages of all the possible time-ordered pairs of the field operators (Wick's theorem); the resulting terms of the series, which contain the unperturbed Green's function (i.e., Green's function related to the unperturbed, exactly solvable, Hamiltonian), can be conveniently represented in a pictorial fashion by Feynman diagrams. Introduced by Richard P. Feynman around 1948 as a graphic shorthand in his study of quantum electrodynamics, these diagrams have become an essential tool for performing calculations in modern quantum field theory.

It helps to conceive of Feynman diagrams in terms of words. Although the single diagrams may be given only a quasiphysical[‡] interpretation, the whole perturbation series forms a sentence that does have a real physical meaning [103]. By partially summing over selected[§] sets of diagrams, one obtains a method that goes beyond ordinary perturbation theory, which is just as well, since most of the relevant diagrams represent divergent terms, and one has to sum them up to the infinite order to obtain a convergent answer. When performing sums over diagrams, one needs to consider only topologically distinct diagrams (with an appropriate multiplication factor). Moreover, the denominator in Equation 2.78 cancels the disconnected diagrams of the numerator (i.e., those containing subunits that are not connected to the rest of the diagram by any lines). Diagrammatically speaking, Green's function is just the sum of all topologically distinct and connected diagrams. Context is essential for the interpretation of Feynman's diagrammatic language. The translation of the semantic units of the diagrams (propagation lines, interaction lines, and vertices) into mathematical expressions requires different dictionaries (also known as Feynman rules) depending on the types of quasiparticles involved, e.g., charge carriers, phonons, photons, and on the actual formalism employed, e.g., the zero- or finite-temperature formalism of systems at equilibrium (EGF) or the (finite-temperature) contour-ordered formalism of out-of-equilibrium systems (NEGF).

[†] Most scattering potentials do not vanish for finite distances $z - z'$. In addition, Green's functions in weakly confined systems decay relatively slowly in space, so that the effective range of the scattering self-energies may be relatively large. Although unwarranted, the local approximation (i.e., spatially diagonal self-energies) seems the only viable approach to 3D NEGF simulation [122].

[‡] Because the single terms in the expansion represent virtual rather than real scattering processes, conservative textbooks warn the reader not to give the diagrams any physical interpretation, and recommend regarding them merely as *calculational* devices, i.e., as tools designed to keep track of complicated mathematical expressions arising in a perturbative calculation. Part of the success of Feynman diagrams over alternative diagrammatic methods is related to their similarity to photographs of *real* particle tracks, although this similarity was considered ambiguous, if not misleading, since it incorrectly suggested trajectories, in contrast to the fundamental principles of quantum mechanics [119]. Made of dots, lines, and wiggles, these drawings are so charmingly expressive that it is really tempting to consider them as *representational* devices. Acknowledging the "quasi-physical" nature of the diagrams is actually the attempt by some textbooks at a compromise between calculational

It is beyond the scope of this chapter to develop the full diagrammatic formulation of many-body perturbation theory. Here we just present the SCBA expressions of the carrier–phonon self-energies, see [115, Appendix E] and [69, Appendix C] for a derivation. Assuming that the phonon population N_Q is at equilibrium, and restricting our attention to bulk phonon modes (in other words, the effects of confinements enter only through the electronic part of the interaction), the carrier–phonon (Fock) self-energy in our nonorthogonal basis is

$$\Sigma^{\gtrless}_{\alpha\beta}(\underline{k},E) = \sum_{\underline{Q}} |U_Q|^2 e^{\mathrm{i}q_z(z_i-z_j)} \left[M(N_Q G^{\gtrless}(\underline{k}-\underline{q},E\pm\hbar\omega_Q) + (N_Q+1)G^{\gtrless}(\underline{k}-\underline{q},E\mp\hbar\omega_Q))M\right]_{\alpha\beta} \tag{2.121}$$

$$\Sigma^R(\underline{k},E) = \frac{1}{2}(\Sigma^{>}(\underline{k},E) - \Sigma^{<}(\underline{k},E)) - \mathrm{i}\mathcal{P}\int \frac{\mathrm{d}E'}{2\pi}\frac{\Sigma^{>}(\underline{k},E) - \Sigma^{<}(\underline{k},E)}{E-E'}, \tag{2.122}$$

where $\underline{Q} = (\underline{q}, q_z)$, U_Q is the scattering strength and M is the mass matrix in Equation 2.115.[†] The principal part is usually neglected (it just leads to a negligible energy renormalization), which gives

$$\Sigma^R(\underline{k},E) \approx \frac{1}{2}(\Sigma^{>}(\underline{k},E) - \Sigma^{<}(\underline{k},E)). \tag{2.123}$$

A common deformation-potential model used in acoustic scattering is a linear phonon dispersion $\omega_Q = Qu_l$ and $U_Q = \sqrt{\frac{\hbar D_a^2}{2\mathcal{V}\rho u_l}Q}$, where u_l is the longitudinal sound velocity in the material, D_a is the deformation potential, ρ is the semiconductor mass density, and $\mathcal{V} = \mathcal{A}\mathcal{L}$ is the normalization volume. Since $\hbar\omega_Q$ is small compared to $k_\mathrm{B}T$, the equipartition approximation holds, i.e., $N_Q \approx \frac{k_\mathrm{B}T}{\hbar\omega_Q} \approx N_Q + 1$, i.e.,

---

and representational interpretations, as, e.g., "it seems a bit extreme to completely reject any sort of physical interpretation whatsoever" in [103, p. 88]. Admittedly, the intuitive appeal of Feynman diagrams seems unmatched in modern physics.

[§] Selective summation means that the perturbation series is evaluated approximately by summing to infinite order certain types of diagrams. Example: the sum over open oyster and bubble diagrams yields the HF approximation [103, p. 120].

[†] Since photons and phonons are both bosons, Equation 2.121 also holds for carrier–photon interaction, the only difference being the expression of the photon number $N_{\lambda,q}$, which depends on the type of the radiation field considered (e.g., thermal radiation governed by the Bose–Einstein distribution, or monochromatic illumination). The carrier–photon self-energy is, of course, an important ingredient in the NEGF analysis of optoelectronic nanodevices, because it enables us to treat carrier dynamics and recombination processes on equal footing.

the acoustic self-energy is approximately elastic, which gives

$$\Sigma^{\gtrless R}_{\alpha\beta}(E) = \frac{1}{\Delta_{ij}} \frac{D_a^2 k_B T}{\mathcal{V}\rho u_l^2} \sum_{\underline{q}} \left[M G^{\gtrless R}(\underline{k} - \underline{q}, E) M \right]_{\alpha\beta}, \tag{2.124}$$

where α, β are compound indices that combine indices i, j for space and m, n for band, and

$$\frac{1}{\Delta_{ij}} = \sum_{q_z} e^{\mathrm{i}q_z(z_i - z_j)} = \frac{\mathcal{L}}{2\pi} \int_{-\pi/a}^{\pi/a} \mathrm{d}q_z \, e^{\mathrm{i}q_z(z_i - z_j)} = \frac{\mathcal{L}}{\pi} \frac{\sin \frac{\pi}{a}(z_i - z_j)}{z_i - z_j}. \tag{2.125}$$

The Fröhlich theory of polar optical scattering assumes a dispersion-less longitudinal phonon with energy $\hbar\omega_{\mathrm{LO}}$ and an interaction strength $U_q = \sqrt{\frac{e^2 \hbar\omega_{\mathrm{LO}}}{2\epsilon_p \mathcal{V}} \frac{Q}{Q^2 + q_0^2}}$, where $\epsilon_p^{-1} = \epsilon_\infty^{-1} - \epsilon_s^{-1}$, ϵ_s and ϵ_∞ are the static and optical dielectric constants of the material, and $q_0 = \sqrt{\frac{e^2 n}{\epsilon_s k_B T}}$ is the inverse Debye–Hückel screening length in the nondegenerate limit. In the axial approximation, the self-energy is

$$\begin{aligned} \Sigma^{\gtrless}_{\alpha\beta}(\underline{k}, E) ={}& \frac{e^2 \hbar\omega_{\mathrm{LO}}}{2\mathcal{V}\epsilon_p} \frac{\mathcal{V}}{(2\pi)^3} \int \mathrm{d}\underline{q}\, \underline{q} \int \mathrm{d}q_z \int_0^{2\pi} \mathrm{d}\phi \, e^{\mathrm{i}q_z(z_i - z_j)} \frac{q_z^2 + q^2 + k^2 + 2kq\cos\phi}{(q_z^2 + q^2 + k^2 + q_0^2 + 2kq\cos\phi)^2} \\ &\times \left[M(N_{\mathrm{LO}} G^{\gtrless}(\underline{q}, E \pm \hbar\omega_{\mathrm{LO}}) + (N_{\mathrm{LO}} + 1) G^{\lessgtr}(\underline{q}, E \mp \hbar\omega_{\mathrm{LO}}))M \right]_{\alpha\beta}, \end{aligned} \tag{2.126}$$

with the phonon occupation number $N_{\mathrm{LO}} = \left(e^{\frac{\hbar\omega_{\mathrm{LO}}}{kT}} - 1 \right)^{-1}$. Lengthy mathematical manipulations [120] lead to the final form of the lesser and greater self-energies

$$\begin{aligned} \Sigma^{\gtrless}_{\alpha\beta}(\underline{k}, E) ={}& \frac{e^2 \hbar\omega_{\mathrm{LO}}}{4\pi^2 \epsilon_p} \int \mathrm{d}\underline{q}\, \underline{q} F(q, s_{ij}, k, q_0) \\ &\times \left[M(N_{\mathrm{LO}} G^{\gtrless}(\underline{q}, E \pm \hbar\omega_{\mathrm{LO}}) + (N_{\mathrm{LO}} + 1) G^{\lessgtr}(\underline{q}, E \mp \hbar\omega_{\mathrm{LO}}))M \right]_{\alpha\beta} \end{aligned} \tag{2.127}$$

with $s_{ij} = z_i - z_j$ and

$$\begin{aligned} F(q, s_{ij}, k, q_0) ={}& \int_0^{\pi/a} \mathrm{d}q_z \cos(q_z s_{ij}) \\ &\times \left[\frac{1}{\sqrt{(q_z^2 + q^2 + q_0^2 + k^2)^2 - 4k^2 q^2}} - q_0^2 \frac{q_z^2 + q^2 + q_0^2 + k^2}{((q_z^2 + q^2 + q_0^2 + k^2)^2 - 4k^2 q^2)^{3/2}} \right]. \end{aligned} \tag{2.128}$$

Equations 2.115 through 2.117, together with the expressions for the boundary self-energies accounting for the coupling to the reservoir regions, and the scattering self-energies Equations 2.124 and 2.127 describing the interactions with acoustic and optical phonons, form a closed set of equations that have to be solved self-consistently with the macroscopic Poisson's equation. The self-consistent solution of Poisson's and Dyson's equations is known as the *outer loop* in NEGF theory, the *inner loop* being the one that connects Green's functions with the scattering self-energies.[†]

Summarizing, the NEGF algorithm for the calculation of physical quantities consists of the following steps (see Figure 2.5):

1. The retarded boundary self-energies $\Sigma^{RB}_{L,R}$ due to the coupling to the contacts are evaluated as in Equation 2.119; the corresponding lesser and greater quantities $\Sigma^{\lessgtr B}_{L,R}$ follow from the equilibrium relations Equation 2.120. Dyson's Equation 2.115 is solved for the retarded Green's function G^R. Conjugate transpose of G^R gives the advanced Green's function G^A.
2. The Keldysh Equation 2.117 gives the correlation functions $G^{\lessgtr}$ from the retarded Green's function G^R and the self-energies $\Sigma^{\lessgtr} = \Sigma^{\lessgtr B} + \Sigma^{\lessgtr S}$.
3. The scattering self-energies $\Sigma^{\lessgtr S}$ are updated using the previously calculated Green's functions. If the change is smaller than a given threshold, the inner loop stops; otherwise the previous steps (Dyson, Keldysh, and self-energy equations) are repeated.
4. Upon convergence of the inner loop, physical quantities (e.g., carrier densities and currents) can be computed from $G^{\lessgtr}$.
5. Given the NEGF carrier densities, space-charge effects are accounted for by an additional self-consistency loop (outer loop), in which Poisson's equation is solved for the Hartree potential U.

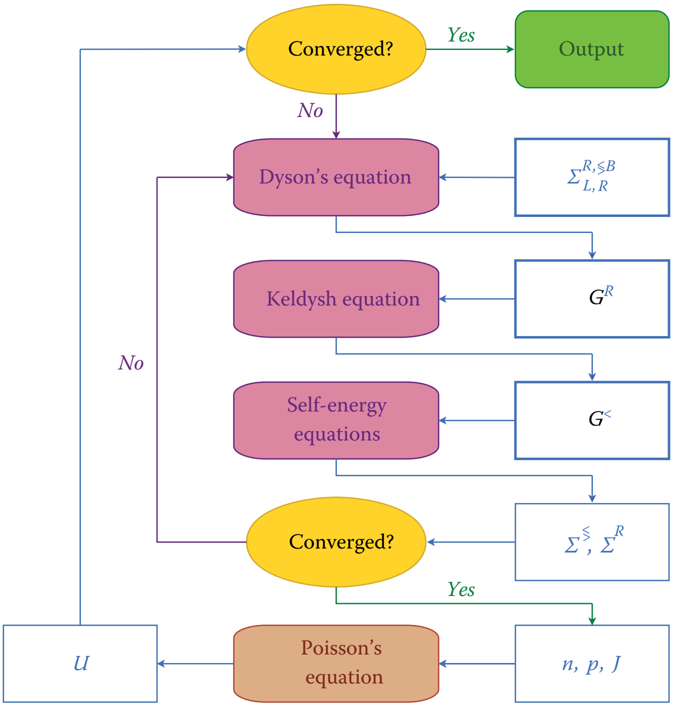


FIGURE 2.5 Flow chart of simulation. The inner self-consistency loop connects Green's functions with self-energies, while the outer loop provides the update of the Hartree potential from the solution of Poisson's equation.

[†] The outer loop can be traced back to the Hartree part of carrier–carrier interactions.

2.4.4 Conclusions

Our choice to discuss NEGF in more detail with respect to other genuine quantum approaches (e.g., the DM method) has less to do with general considerations on the merits of different formulations than with the authors' perspective on the simulation of a specific class of optoelectronic devices. In principle, a self-consistent description treating both carrier transport and optical generation on equal footing may be accomplished within an NEGF or a DM formalism. Approaching the problem from a DM perspective means solving the dynamics of diagonal (carrier populations) as well as off-diagonal (interband polarizations) elements of the reduced single-particle DM; current models usually formulated in terms of the bound states of the system have to be extended to describe their connection to the scattering states of the system. Approaching the problem from an NEGF perspective means addressing the staggering computational cost required by the calculation of Green's functions, as conventional recursive techniques are not viable when the scattering self-energies are fully nonlocal. Theoretical and numerical considerations may guide us in choosing which is the most appropriate method. As for carrier transport in GaN-based LEDs, we speculate that NEGF is probably the best choice if the objective is to clear away the uncertainty still surrounding some aspects of the droop issue, since the method gives access to energy-resolved quantities that enable us to separate different transport mechanisms. NEGF calculations provide a vivid pictorial representation of tunneling through the polarization-induced barrier, ballistic transport over the electron-blocking layer, and capture processes beyond what one-time techniques can do. As an example, Figure 2.6 shows the spectral current density in a simple LED structure that includes a single QW. Carrier–photon and carrier–photon self-energies were computed within the SCBA (carrier–carrier interactions are included at the Hartree level through coupling to Poisson's equation). Tunneling clearly plays an important role in delivering carriers to the QW.

The prevalence of FD techniques in the discretization of NEGF models of carrier transport seems, at first, simply a matter of preference/background of the NEGF demographic (a notable exception is [69]). A closer look at the problem reveals that there are actually numerical reasons motivating this seemingly arbitrary choice. For example, the use of an FE basis leads to nondiagonal terms in the self-energies, even if the

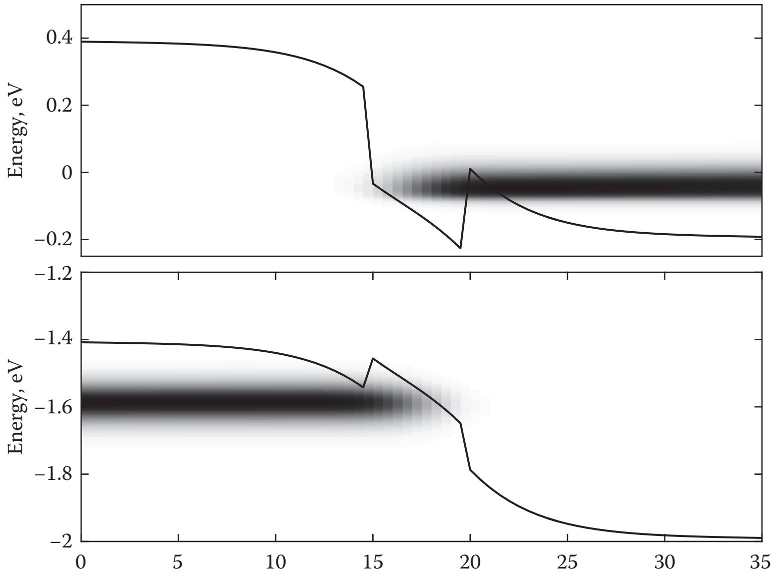


FIGURE 2.6 Spatially and energetically resolved current in a forward-biased $Al_{0.3}Ga_{0.7}As/GaAs$ *p-i-n* junction.

corresponding interactions are approximated as local in space.[†] Factor in that the use of nonorthogonal basis functions is more involved since it leads to different (contravariant and covariant) representations of Green's functions and self-energies. Factor in, too, that multiband envelope-function models widely employed in the design of optoelectronic devices may lead to spurious, unphysical solutions playing havoc in subband dispersions, especially in FE implementations. Notwithstanding the difficulties above, there are very good reasons to start using FE in the NEGF simulation of optoelectronic devices. Besides the geometrical flexibility afforded by the FE method, what makes the extra complication worthwhile is the rigorous treatment of material discontinuities in Schrödinger(-like) problems, where other approaches have to rely on special BCs to ensure current continuity. Higher-order polynomials (p-refinement) may be used to reduce the number of basis functions. More importantly, the FE approach allows us to unambiguously determine operator ordering and band parameters from nonlocal full-Brillouin-zone calculations, leading to well-posed, numerically stable envelope equations that provide realistic subband dispersions of quantum systems [58], while remedies proposed to eliminate spurious solutions in FD approaches have not yet found wide acceptance; the issue is also related to operator ordering and the choice of the band parameters*.

The staggering computational cost remains the main limitation of NEGF techniques. Reduced-order techniques [121], mode-space representations [122], and algorithms exploiting advanced matrix partitioning strategies (e.g., FIND [123], based on nested dissection, and SelInv [124] based on selected inversion) for the efficient calculation of Green's functions could make NEGF simulations maybe possible for realistic optoelectronic devices. Notable DM and NEGF contributions along these lines in the context of silicon nanowire field-effect transistors[125], solar cells [126] and LEDs [127,128] must be mentioned alongside promising projects such as ANGEL [69,129] and NEMO5 [130,131].

References

1. U. Rau. Reciprocity relation between photovoltaic quantum efficiency and electroluminescent emission of solar cells. *Phys. Rev. B.*, 76:085303, August 2007.
2. G. Ghione. *Semiconductor Devices for High-Speed Optoelectronics*. Cambridge: Cambridge University Press, 2009.
3. C. Cercignani and E. Gabetta, editors. *Transport Phenomena and Kinetic Theory: Applications to Gases, Semiconductors, Photons, and Biological Systems*. Boston, MA: Birkhäuser, 2007.
4. D. Vasileska, S. M. Goodnick, and G. Klimeck. *Computational Electronics. Semiclassical and Quantum Device Modeling and Simulation*. Boca Raton, FL: CRC Press, 2010.
5. C. Jacoboni. *Theory of Electron Transport in Semiconductors. A Pathway from Elementary Physics to Nonequilibrium Green Functions*. Berlin: Springer-Verlag, 2010.
6. F. Rossi and T. Kuhn. Theory of ultrafast phenomena in photoexcited semi-conductors. *Rev. Modern Phys.*, 74:895–950, August 2002.
7. T. Kuhn. Density matrix theory of coherent ultrafast dynamics. In E. Schöll, ed., *Theory of Transport Properties of Semiconductor Nanostructures*, chapter 6, pp. 173–214. Dordrecht: Springer Science+Business Media, 1998.
8. M. Fischetti and W. G. Vandenberghe. *Advanced Physics of Electron Transport in Semiconductors and Nanostructures*. Berlin: Springer-Verlag, 2016.

[†] We should explain here that the determination of the full GF would provide information on propagation and correlations between any two points of the device and thus would include all nonlocal phenomena, but in order to reduce the computational burden, a limited number of nonzero off-diagonal GF elements is usually considered. This approximation of which allows the use of efficient recursive techniques to compute selected elements of the inverse matrix (Green's function) needed for the evaluation of energy and spectrally resolved single-particle quantities.

9. F. Bertazzi, M. Moresco, and E. Bellotti. Theory of high field carrier transport and impact ionization in wurtzite GaN. Part I: A full band Monte Carlo model. *J. Appl. Phys.*, 106(6):063718, September 2009.
10. F. Bertazzi, F. Cappelluti, S. Donati Guerrieri, F. Bonani, and G. Ghione. Self-consistent coupled carrier transport full-wave EM analysis of semiconductor traveling-wave devices. *IEEE Trans. Microwave Theory Tech.*, 54(4):1611–1618, April 2006.
11. C. Jungemann and B. Meinerzhagen. *Hierarchical Device Simulation. The Monte-Carlo Perspective. Computational Microelectronics*. Vienna: Springer-Verlag, 2003.
12. C. Jacoboni and P. Lugli. *The Monte Carlo Method for Semiconductor Device Simulation. Computational Microelectronics*. Vienna: Springer-Verlag, 1989.
13. C. Moglestue. *Monte Carlo Simulation of Semiconductor Devices*. London: Chapman & Hall, 1993.
14. F. Rossi. *Theory of Semiconductor Quantum Devices. Microscopic Modeling and Simulation Strategies*. Berlin: Springer-Verlag, 2011.
15. I. Ezhov and C. Jirauschek. Influence of screening on longitudinal-optical phonon scattering in quantum cascade lasers. *J. Appl. Phys.*, 119(3):033102, 2016.
16. R. Claudia Iotti, F. Rossi, M. Serena Vitiello, G. Scamarcio, L. Mahler, and A. Tredicucci. Impact of nonequilibrium phonons on the electron dynamics in terahertz quantum cascade lasers. *Appl. Phys. Lett.*, 97(3):033110, 2010.
17. M. Moresco, F. Bertazzi, and E. Bellotti. Theory of high field carrier transport and impact ionization in wurtzite GaN. Part II: Application to avalanche photodetectors. *J. Appl. Phys.*, 106(6):063719, September 2009.
18. M. Moresco, F. Bertazzi, and E. Bellotti. GaN avalanche photodectors: A full band Monte Carlo study of gain, noise and bandwidth. *IEEE J. Quant. Electron.*, 47(4):447–454, April 2011.
19. E. Bellotti, F. Bertazzi, and M. Goano. Alloy scattering in AlGaN and InGaN: A numerical study. *J. Appl. Phys.*, 101(12):123706, 2007.
20. M. Goano, E. Bellotti, E. Ghillino, G. Ghione, and K. F. Brennan. Band structure nonlocal pseudopotential calculation of the III-nitride wurtzite phase materials system. Part I. Binary compounds GaN, AlN, and InN. *J. Appl. Phys.*, 88(11):6467–6475, December 2000.
21. M. V. Fischetti. Monte Carlo simulation of transport in technologically significant semi-conductors of the diamond and zinc-blende structures—Part I: Homogeneous trans-port. *IEEE Trans. Electron Dev.*, 38(3):634–649, March 1991.
22. M. V. Fischetti and S. E. Laux. Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects. *Phys. Rev. B*, 38(14):9721–9745, November 1988.
23. T. Kunikiyo, M. Takenaka, Y. Kamakura, M. Yamaji, H. Mizuno, M. Morifuji, K. Taniguchi, and C. Hamaguchi. A Monte Carlo simulation of anisotropic electron transport in silicon including full band structure and anisotropic impact-ionization model. *J. Appl. Phys.*, 75(1):297–312, January 1994.
24. H.-E. Nilsson, A. Martinez, E. Ghillino, U. Sannemo, E. Bellotti, and M. Goano. Numerical modeling of hole interband tunneling in wurtzite GaN and SiC. *J. Appl. Phys.*, 90(6):2847–2852, September 2001.
25. K. Hess. *Advanced Theory of Semiconductor Devices*. Englewood Cliffs, NJ: Prentice-Hall, 1988.
26. R. Stratton. Diffusion of hot and cold electrons in semiconductor barriers. *Phys. Rev.*, 126(6):2002–2014, June 1962.
27. T. Grasser, T.-W. Tang, H. Kosina, and S. Selberherr. A review of hydrodynamic and energy-transport models for semiconductor device simulation. *Proc. IEEE*, 91(2):251–274, February 2003.
28. G. Ghione and A. Benvenuti. Discretization schemes for high-frequency semiconductor device models. *IEEE Trans. Antennas Propag.*, 45(3):443–456, March 1997.

29. M. Rudan, M. Lorenzini, and R. Brunetti. Hydrodynamic simulation of semiconductor devices. In E. Schöll, ed., *Theory of Transport Properties of Semiconductor Nanostructures*, chapter 2, pp. 27–57. Dordrecht: Springer Science+Business Media, 1998.
30. E. M. Azoff. Closed-form method for solving the steady-state generalised energy-momentum conservation equations. *COMPEL*, 6(1):25–30, 1987.
31. E. M. Azoff. Energy transport numerical simulation of graded AlGaAs/GaAs hetero-junction bipolar transistors. *IEEE Trans. Electron. Dev.*, 36(4):609–616, 1989.
32. E. Bellotti and F. Bertazzi. Transport parameters for electrons and holes. In J. Piprek, ed., *Nitride Semiconductor Devices: Principles and Simulation*, chapter 4, pp. 69–93. Weinheim: Wiley-VCH Verlag, 2007.
33. E. Furno, F. Bertazzi, M. Goano, G. Ghione, and E. Bellotti. Hydrodynamic transport parameters of wurtzite ZnO from analytic- and full-band Monte Carlo simulation. *Solid-State Electron.*, 52(11):1796–1801, 2008.
34. V. Camarchia, E. Bellotti, M. Goano, and G. Ghione. Physics based modeling of submicron GaN permeable base transistors. *IEEE Electron. Device Lett.*, EDL-23(6):303–305, June 2002.
35. B. Carnez, A. Cappy, A. Kaszynski, E. Constant, and G. Salmar. Modeling of a sub-micrometer gate field-effect transistor including effects of nonstationary electron dynamics. *J. Appl. Phys.*, 51(1):784–790, January 1980.
36. M. Streiff, A. Witzig, M. Pfeiffer, P. Royo, and W. Fichtner. A comprehensive VCSEL device simulator. *IEEE J. Sel. Top. Quantum Electron.*, 9(3):879–891, 2003.
37. S. Selberherr. MOS device modeling at 77 K. *IEEE Trans. Electron. Dev.*, 36(8):1464–1474, August 1989.
38. J. S. Blakemore. Approximation for Fermi-Dirac integrals, especially the function $\mathcal{F}_{1/2}(\eta)$ used to describe electron density in a semiconductor. *Solid-State Electron.*, 25(11):1067–1076, 1982.
39. M. Goano. Series expansion of the Fermi-Dirac integral $\mathcal{F}_j(x)$ over the entire domain of real j and x. *Solid-State Electron.*, 36(2):217–221, 1993.
40. D. Schroeder. *Modelling of Interface Carrier Transport for Device Simulation. Computational Microelectronics*. Vienna: Springer-Verlag, 1994.
41. F. Bonani and G. Ghione. *Noise in Semiconductor Devices. Modeling and Simulation.* Berlin: Springer-Verlag, 2001.
42. J. S. Blakemore. *Semiconductor Statistics*, Volume 3 of International Series of Monographs on Semiconductors. New York, NY: Pergamon Press, 1962.
43. P. T. Landsberg. *Recombination in Semiconductors*. Cambridge: Cambridge University Press, 1991.
44. S. M. Sze. *Physics of Semiconductor Devices*, 2nd edition. New York, NY: John Wiley & Sons, 1981.
45. S. Selberherr. *Analysis and Simulation of Semiconductor Devices*. Vienna: Springer-Verlag, 1984.
46. F. Bonani, S. Donati Guerrieri, G. Ghione, and M. Pirola. A TCAD approach to the physics-based modeling of frequency conversion and noise in semi-conductor devices under large-signal forced operation. *IEEE Trans. Electron. Dev.*, 48(5):966–977, May 2001.
47. D. L. Scharfetter and H. K. Gummel. Large-signal analysis of a silicon read diode transistor. *IEEE Trans. Electron. Dev.*, ED-16(1):64–77, January 1969.
48. M. S. Adler, V. A. K. Temple, and R. C. Rustay. Theoretical basis for field calculations on multi-dimensional reverse biased semiconductor devices. *Solid-State Electron.*, 25(12):1179–1186, 1982.
49. A. Deinega and S. John. Finite difference discretization of semiconductor drift-diffusion equations for nanowire solar cells. *Comp. Phys. Comm.*, 183(10):2128–2135, 2012.
50. G. A. Franz, A. F. Franz, S. Selberherr, and P. Markowich. A quasi three dimensional semiconductor device simulation using cylindrical coordinates. In *Proceedings of the NASECODE III Conference*, pp. 122–127, Dublin: Boole Press, 1983.
51. K. Matsumoto, I. Takayanagi, T. Nakamura, and R. Ohta. The operation mechanism of a charge modulation device (CMD) image sensor. *IEEE Trans. Electron. Dev.*, 38(5):989–998, 1991.

52. M. Spevak and T. Grasser. Discretization of macroscopic transport equations on non-Cartesian coordinate systems. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.*,, 26(8):1408–1416, 2007.
53. M. Sharma and G. F. Carey. Semiconductor device modeling using flux upwind finite elements. *COMPEL*, 8(4):219–224, 1989.
54. M. Calciati, A. Tibaldi, F. Bertazzi, M. Goano, and P. Debernardi. Many-valley electron transport in AlGaAs VCSELs. Semicond. Sci. Tech., 32(5):055007, 2017. doi:10.1088/1361-6641/aa66bb
55. M. G. Ancona and G. J. Iafrate. Quantum correction to the equation of state of an electron gas in a semiconductor. *Phys. Rev. B*, 39:9536–9540, May 1989.
56. S. L. Chuang and C. S. Chang. $k{\cdot}p$ method for strained wurtzite semiconductors. *Phys. Rev. B*, 54(4):2491–2504, July 1996.
57. R. G. Veprek, S. Steiger, and B. Witzigmann. Ellipticity and the spurious solution problem of $k{\cdot}p$ envelope equations. *Phys. Rev. B*, 76(16):165320, 2007.
58. X. Zhou, F. Bertazzi, M. Goano, G. Ghione, and E. Bellotti. Deriving $k{\cdot}p$ parameters from full-Brillouin-zone descriptions: A finite-element envelope function model for quantum-confined wurtzite nanostructures. *J. Appl. Phys.*, 116(3):033709, July 2014.
59. R. G. Veprek. *Computational Modeling of Semiconductor Nanostructures for Optoelectronics*. PhD thesis, Swiss Federal Institute of Technology, Zürich, 2009.
60. F. Mireles and S. E. Ulloa. Ordered Hamiltonian and matching conditions for heterojunctions with wurtzite symmetry: GaN/AlxGa1$-x$N quantum wells. *Phys. Rev. B*, 60(19):13659–13667, November 1999.
61. R. G. Veprek, S. Steiger, and B. Witzigmann. Operator ordering, ellipticity and spurious solutions in $k \cdot p$ calculations of III-nitride nanostructures. *Opt. Quant. Electron.*, 40(14–15):1169–1174, November 2008.
62. L.-W. Wang and A. Zunger. Linear combination of bulk bands method for largescale electronic structure calculations on strained nanostructures. *Phys. Rev. B*, 59(24):15806–15818, June 1999.
63. B. A. Foreman. Quadratic response theory for spin–orbit coupling in semiconductor heterostructures. *Phys. Rev. B.*, 72:165344, October 2005.
64. A. Di Carlo. Microscopic theory of nanostructured semiconductor devices: Beyond the envelope-function approximation. *Semiconductor Sci. Tech.*, 18(1):R1–R31, 2003.
65. A. Trellakis, A. T. Galick, A. Pacelli, and U. Ravaioli. Iteration scheme for the solution of the two-dimensional Schrödinger–Poisson equations in quantum structures. *J. Appl. Phys.*, 81(12):7880–7884, June 1997.
66. M. V. Fischetti and S. E. Laux. Monte Carlo study of electron transport in silicon inversion layers. *Phys. Rev. B*, 48(4):2244–2274, July 1993.
67. B. Witzigmann, A. Witzig, and W. Fichtner. A multidimensional laser simulator for edge-emitters including quantum carrier capture. *IEEE Trans. Electron. Dev.*, 47(10):1926–1934, October 2000.
68. M. Grupen, G. Kosinovsky, and K. Hess. The effect of carrier capture on the modulation bandwidth of quantum well lasers. In *Proceedings of the IEDM*, pp. 23.6–23.6.4, Washington, DC, 1993.
69. S. Steiger. *Modelling Nano-LEDs*. PhD thesis, Swiss Federal Institute of Technology, Zürich, 2009.
70. W. W. Chow, H. C. Schneider, S. W. Koch, C.-H. Chang, L. Chrostowski, and C. J. Chang-Hasnain. Nonequilibrium model for semiconductor laser modulation response. *IEEE J. Quant. Electron.*, 38(4):402–409, April 2002.
71. E. F. Schubert. *Physical Foundations of Solid-State Devices*. Rensselaer Polytechnic Institute, Troy, NY, 2015.
72. M. Vallone. Quantum well electron scattering rates through longitudinal optic-phonon dynamical screened interaction: An analytic approach. *J. Appl. Phys.*, 114(5):053704, August 2013.
73. G. R. Hadley, K. L. Lear, M. E. Warren, K. D. Choquette, J. W. Scott, and S. W. Corzine. Comprehensive numerical modeling of vertical-cavity surface-emitting lasers. *IEEE J. Quant. Electron.*, 32(4):607–616, 1996.

74. O. Conradi, S. Helfert, and R. Pregla. Comprehensive modeling of vertical-cavity laser diodes by the method of lines. *IEEE J. Quantum Electron.*, 37(7):928–935, 2001.
75. M. Dems, T. Czyszanowski, and K. Panajotov. Plane-wave and cylindrical-wave admittance method for simulation of classical and photonic-crystal-based VCSELs. In *Proceedings of the SPIE, Vol. 6182, Photonic Crystal Materials and Devices III*, pp. 618219–1–618219–8, Bellingham, WA: SPIE, 2006.
76. M. Dems, R. Kotynski, and K. Panajotov. Planewave admittance method—A novel approach for determining the electromagnetic modes in photonic structures. *Opt. Express.*, 13(9):3196–3207, 2005.
77. M. Dems. Modelling of high-contrast grating mirrors. The impact of imperfections on their performance in VCSELs. *Opto-Electron. Rev.*, 19(3):340–345, 2011.
78. G. P. Bava, P. Debernardi, and L. Fratta. Three-dimensional model for vectorial fields in vertical-cavity surface-emitting lasers. *Phys. Rev. A*, 63(2):23816–1–23816–13, 2001.
79. P. Debernardi and G. P. Bava. Coupled mode theory: A powerful tool for analyzing complex VCSELs and designing advanced devices features. *IEEE J. Sel. Top. Quantum Electron.*, 9(3):905–917, 2003.
80. R. Michalzik. *VCSELs: Fundamentals, Technology and Applications of Vertical-Cavity Surface-Emitting Lasers*. Berlin: Springer-Verlag, 2013.
81. G. P. Bava, P. Debernardi, and L. Fratta. Quantum noise in vertical-cavity surface-emitting lasers with polarization competition. *J. Opt. Soc. Am. B*, 16(11):2147–2157, 1999.
82. E. Malic and A. Knorr. *Graphene and Carbon Nanotubes. Ultrafast Relaxation Dynamics and Optics*. Weinheim: VCH, 2013.
83. W. W. Chow and S. W. Koch. *Semiconductor-Laser Fundamentals. Physics of the Gain Materials*. Berlin: Springer-Verlag, 1999.
84. J. Piprek. Efficiency droop in nitride-based light-emitting diodes. *Phys. Status Solidi A*, 207(10):2217–2225, October 2010.
85. G. Verzellesi, D. Saguatti, M. Meneghini, F. Bertazzi, M. Goano, G. Meneghesso, and E. Zanoni. Efficiency droop in In-GaN/GaN blue light-emitting diodes: Physical mechanisms and remedies. *J. Appl. Phys.*, 114(7):071101, August 2013.
86. C. Weisbuch, M. Piccardo, L. Martinelli, J. Iveland, J. Peretti, and J. S. Speck. The efficiency challenge of nitride light-emitting diodes for lighting. *Phys. Status Solidi A*, 212(5):899–913, 2015.
87. Y. C. Shen, G. O. Mueller, S. Watanabe, N. F. Gardner, A. Munkholm, and M. R. Krames. Auger recombination in InGaN measured by photoluminescence. *Appl. Phys. Lett.*, 91(14):141101, 2007.
88. K. T. Delaney, P. Rinke, and C. G. Van de Walle. Auger recombination rates in nitrides from first principles. *Appl. Phys. Lett.*, 94:191109, 2009.
89. E. Kioupakis, D. Steiauf, P. Rinke, K. T. Delaney, and C. G. Van de Walle. First-principles calculations of indirect Auger recombination in nitride semiconductors. 2014. http://arxiv.org/abs/1412.7555.
90. E. Kioupakis, Q. Yan, and C. G. Van de Walle. Interplay of polarization fields and Auger recombination in the efficiency droop of nitride light-emitting diodes. *Appl. Phys. Lett.*, 101(23):231107, 2012.
91. M. Binder, A. Nirschl, R. Zeisel, T. Hager, H.-J. Lugauer, M. Sabathil, D. Bougeard, J. Wagner, and B. Galler. Identification of *nnp* and *npp* Auger recombination as significant contributor to the efficiency droop in (GaIn) N quantum wells by visualization of hot carriers in photoluminescence. *Appl. Phys. Lett.*, 103(7):071108, August 2013.
92. J. Iveland, L. Martinelli, J. Peretti, J. S. Speck, and C. Weisbuch. Direct measurement of Auger electrons emitted from a semiconductor light-emitting diode under electrical injection: Identification of the dominant mechanism for efficiency droop. *Phys. Rev. Lett.*, 110:177406, April 2013.
93. A. Nirschl, M. Binder, M. Schmid, I. Pietzonka, H.-J. Lugauer, R. Zeisel, M. Sabathil, D. Bougeard, and B. Galler. Towards quantification of the crucial impact of Auger recombination for the efficiency droop in (AlInGa) N quantum well structures. *Opt. Express.*, 24(3):2971–2980, February 2016.
94. V. Avrutin, S. A. Hafiz, F. Zhang, M. Zgr, E. Bellotti, F. Bertazzi, M. Goano, A. Matulionis, A. T. Roberts, H. O. Everitt, and H. Morko. Saga of efficiency degradation at high injection in InGaN light emitting diodes. *Turk. J. Phys.*, 38(3):269–313, November 2014.

95. J. Hader, J. V. Moloney, and S. W. Koch. Beyond the *ABC*: Carrier recombination in semiconductor lasers. In *SPIE Photonics West, Physics and Simulation of Optoelectronic Devices XIV, Vol. 6115, Proceedings of the SPIE*, pp. 61151T. Bellingham, WA: SPIE, January 2006.
96. M. Kira and S. W. Koch. Many-body correlations and excitonic effects in semiconductor spectroscopy. *Prog. Quant. Electron.*, 30(5):155–296, 2006.
97. P. Vogl and T. Kubis. The non-equilibrium Green's function method: An introduction. *J. Comp. Electron.*, 9(3):237–242, 2010.
98. W. Pötz. Microscopic theory of coherent carrier dynamics and phase breaking in semi-conductors. *Phys. Rev. B*, 54(8):5647–5664, August 1996.
99. A. Wacker. Coherence and spatial resolution of transport in quantum cascade lasers. *Phys. Status Solidi C*, 5(1):215–220, 2008.
100. U. Aeberhard. Theory and simulation of quantum photovoltaic devices based on the non-equilibrium Green's function formalism. *J. Comp. Electron.*, 10:394–413, October 2011.
101. G. D. Mahan. *Many-Particle Physics*, 3rd edition. New York, NY: Kluwer Academic Publishers, 2000.
102. A. L. Fetter and J. D. Walecka. *Quantum Theory of Many-Particle Physics*, 3rd edition. New York, NY: Dover Publications, 2003.
103. R. D. Mattuck. *A Guide to Feynman Diagrams in the Many-Body Problem*, 2nd edition. New York, NY: McGraw-Hill, 1976.
104. L. P. Kadanoff and G. Baym. *Quantum Statistical Mechanics. Green's Function Methods in Equilibrium and Nonequilibrium Problems*. New York, NY: W. A. Benjamin, 1962.
105. L. V. Keldysh. Diagram technique for nonequilibrium processes. *Sov. Phys. JETP*, 20(4):1018–1026, April 1965.
106. P. C. Martin and J. Schwinger. Theory of many-particle systems. I. *Phys. Rev.*, 115(6):1342–1026, September 1959.
107. H. Haug and A.-P. Jauho. *Quantum Kinetics in Transport and Optics of Semiconductors*, 2nd edition. Berlin: Springer, 2008.
108. L. V. Keldysh. Real-time nonequilibrium Green's functions. In M. Bonitz and D. Semkat, editors, *Progress in Nonequilibrium Green's Functions*, pp. 4–17. Singapore: World Scientific, 2003.
109. M. Bonitz. *Quantum Kinetic Theory*, 2nd edition. Cham, Switzerland: Springer, 2016.
110. J. Rammer. *Quantum Field Theory of Non-Equilibrium States*. Cambridge: Cambridge University Press, 2007.
111. G. Stefanucci and R. van Leeuwen. *Nonequilibrium Many-Body Theory of Quantum Systems. A Modern Introduction*. Cambridge: Cambridge University Press, 2013.
112. S. Datta. *Quantum Transport: Atom to Transistor*. Cambridge: Cambridge University Press, 2005.
113. U. Aeberhard. Quantum-kinetic theory of photocurrent generation via direct and phonon-mediated optical transitions. *Phys. Rev. B.*, 84:035454, July 2011.
114. P. Danielewicz. Quantum theory of nonequilibrium processes. I. *Ann. Phys.*, 152:239–304, 1984.
115. U. Aeberhard. *A Microscopic Theory of Quantum Well Photovoltaics*. PhD thesis, Swiss Federal Institute of Technology, Zürich, 2008.
116. E. Polizzi and S. Datta. Multidimensional nanoscale device modeling: The finite element method applied to the non-equilibrium Green's function formalism. In *3rd IEEE Conference on Nanotechnology (IEEE-NANO 2003)*, Vol. 2, Piscataway, NJ: IEEE pp. 40–43, August 2003.
117. D. L. Smith and C. Mailhiot. Theory of semiconductor superlattice electronic structure. *Rev. Modern Phys.*, 62(1):173–234, January 1990.
118. M. G. Vergniory, C. Yang, A. Garcia-Lekue, and L.-W. Wang. Calculation of complex band structure for plane-wave nonlocal pseudopotential Hamiltonian. *Comp. Mater. Sci.*, 48(3):544–550, 2010.
119. A. Wüthrich. *The Genesis of Feynman Diagrams*. Dordrecht: Springer, 2010.
120. R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic. Single and multiband modeling of quantum electron transport through layered semiconductor devices. *J. Appl. Phys.*, 81(12):7845–7869, February 1997.

121. J. Z. Huang, L. Zhang, P. Long, M. Povolotskyi, and G. Klimeck. Quantum transport simulation of III-V TFETs with reduced-order $k \cdot p$ method. 2015. http://arxiv.org/abs/1511.02516v1.
122. M. Luisier, A. Schenk, and W. Fichtner. Quantum transport in two-and three-dimensional nanoscale transistors: Coupled mode effects in the nonequi-librium Green's function formalism. *J. Appl. Phys.*, 100(4):043713, 2006.
123. S. Li, S. Ahmed, G. Klimeck, and E. Darve. Computing entries of the inverse of a sparse matrix using the FIND algorithm. *J. Comp. Phys.*, 227(22):9408–9427, November 2008.
124. L. Lin, J. Lu, L. Ying, R. Car, and E. Weinan. Fast algorithm for extracting the diagonal of the inverse matrix with application to the electronic structure analysis of metallic systems. *Commun. Math. Sci.*, 7(3):755–777, 2009.
125. J. Fang, W. G. Vandenberghe, B. Fu, and M. V. Fischetti. Pseudopotential-based electron quantum transport: Theoretical formulation and application to nanometer-scale silicon nanowire transistors. *J. Appl. Phys.*, 119(3):035701, 2016.
126. U. Aeberhard. Simulation of nanostructure-based and ultra-thin film solar cell devices beyond the classical picture. *J. Photon. Energy*, 4(1):042099, 2014.
127. A. Shedbalkar, Z. Andreev, and B. Witzigmann. Simulation of an indium gallium nitride quantum well light-emitting diode with the non-equilibrium Green's function method. *Phys. Status Solidi B*, 253(1):158–163, 2016.
128. R. Wang, Y. Zhang, F. Bi, G. Chen, and C. Yam. Quantum mechanical modeling of nanoscale light emitting diodes. 2015. http://arxiv.org/abs/1512.08498v1.
129. S. Steiger. ANGEL–A nonequilibrium Green function solver for LEDs. 2015. https://nanohub.org/resources/8136.
130. NEMO5. https://engineering.purdue.edu/gekcogrp/software-projects/nemo5/, 2010.
131. S. Steiger, M. Povolotskyi, H.-H. Park, T. Kubis, and G. Klimeck. NEMO5: A parallel multiscale nanoelectronics modeling tool. *IEEE Trans. Nanotech.*, 10(6):1464–1474, November 2011.

3

Electron–Photon Interaction

3.1 Introduction 81
3.2 Two-Level System 82
Optical Processes in the Two-Level System—A Discussion of the Einstein Coefficients • The Two-Level System in the Schrödinger Picture • The Two-Level System in the Heisenberg Picture
3.3 The Solid-State System 90
3.4 Quantization of the Light Field 97
3.5 Role of the Electromagnetic Environment 99

Angela Thränhardt

3.1 Introduction

The devices introduced in this book involve light–matter interaction. Light appears to be a trivial phenomenon known to everyone; however, its mathematical description is quite intricate and has been a major point of discussion in the physics community. Historically, this was manifested in a longstanding dispute on the nature of light, featuring Isaac Newton[†] as the advocate of a corpuscular model and Christiaan Huygens[‡] as the proponent of its wave nature.

Today, more than 300 years later, both perceptions have been reconciled and the description of light in terms of a wave–particle dualism is well established in the physics community. A full quantum-electrodynamic description, relieving the apparent contradiction between wave and particle nature, is often not required, i.e., in many cases, light can be well described as either a wave or a particle. This will be mirrored in this chapter, where both the classical and the quantum perspectives are introduced.

As often stated, the general guideline is that light may be treated as a wave when propagating in vacuum, yet it has to be quantized in the case of the interaction with matter. However, this is not always true, as, e.g., the behavior of light in a waveguide can well be understood in terms of the wave picture. In fact, quantum-optical effects only started to be discussed intensively in the 1950s and 1960s. Their investigation is closely related to the development of the laser and the requirement of solid theoretical foundations in this field. Simply speaking, the corpuscular theory of light states that it is distributed in quantized packages of energy $W = hf = \hbar\omega$, where h is Planck's constant, f is the light frequency, $\hbar = h/(2\pi)$, and ω is the angular frequency $\omega = 2\pi f$. A particle nature is thus attributed to light, where the light particle is known as a photon. The photon may be described quantum mechanically in the framework of second quantization, a formalism in which the light field and other physical field quantities are written as operators. This method was first developed by Paul Dirac[§] in 1927 in analogy to the first quantization where physical quantities such as position and momentum become operators; see Ref. [1]. Further information may be found in textbooks on quantum mechanics.

[†] Sir Isaac Newton, English natural philosopher (physicist and mathematician), 1643–1727.

[‡] Christiaan Huygens, Dutch physicist, astronomer, and mathematician, 1629–1695.

[§] Paul Dirac, British physicist, 1902–1984.

In this chapter, we consider electron–photon interaction, i.e., light–matter coupling. There are different electron–photon interaction mechanisms, e.g., the photoelectric effect, Compton scattering, and pair production. In the energy range of interest in optoelectronics, the photoelectric effect is the only relevant effect and will thus be treated exclusively.

The chapter briefly introduces the theoretical treatment of electron–photon interaction for different and mostly strongly simplified cases. A large part is devoted to the two-level system in different formalisms. This allows us to concentrate on the differences and commonalities of the various approaches. We look at the two-level system in the Schrödinger and the Heisenberg representation (Section 3.2) and then move on to a solid-state system (Section 3.3). Section 3.4 deals with the quantum-optical treatment of the two-level system, while Section 3.5 is devoted to the treatment of photonic structures.

Before we enter the detailed mathematics of a quantum mechanical system, some basics will be discussed. In any ensemble of atoms or molecules in interaction with an electromagnetic field, there are three processes occurring: spontaneous emission, absorption, and stimulated emission. Although spontaneous emission is quite common in everyday life—e.g., most light sources like light bulbs or light-emitting diodes (LEDs) operate based on spontaneous emission—it is difficult to treat theoretically. The existence of spontaneous emission, however, was already postulated by Albert Einstein in 1916/1917 with his argument based on thermodynamic equilibrium. For a handwaving explanation, consider an ensemble of atoms placed inside a closed cavity whose walls are held at a constant temperature. The walls, furthermore, are assumed to be perfect absorbers and emitters of radiation. After the system has reached thermal equilibrium, the cavity will be filled with the so-called "blackbody radiation" according to the temperature T of the system. The condition of thermal equilibrium then requires that for each transition, the spontaneous emission matches the absorption of blackbody radiation, necessitating the existence of spontaneous emission.

3.2 Two-Level System

3.2.1 Optical Processes in the Two-Level System—A Discussion of the Einstein Coefficients

For a more detailed treatment, we now move to the simplest existing system, the two-level system. It includes one electron that can occupy either one of two nondegenerate states. The two levels of the model may represent, e.g., the highest occupied and lowest unoccupied level in an atom, molecule, or quantum structure. We devote this entire section to the two-level system, including a discussion of the Einstein coefficients as well as calculations in the Schrödinger (Section 3.2.2) and the Heisenberg formalism (Section 3.2.3). In Section 3.4, we again return to the two-level system for a quantum-optical discussion. For an even more detailed treatment of two-level systems, the reader is referred to the book by Allen and Eberly [2].

A schematic of the two-level system is shown in Figure 3.1. The system consists of the ground state, level 1 and one excited state, level 2, with the respective energies ϵ_1 and ϵ_2. One of the levels is occupied by an electron, described by the eigenstates $|1\rangle$ or $|2\rangle$. We consider the interaction with an incident light field with photon density Φ_{ω_0}, where ω_0 refers to the angular light frequency corresponding to the energy difference of the two-level system, $\hbar\omega_0 = \epsilon_2 - \epsilon_1$. Let us now think of an ensemble of identical two-level systems, where the number of electrons in state 1 is N_1 and the number of electrons in state 2 is N_2. The following processes indicated in the schematic may occur in any of them:

1. *Spontaneous emission*, as described earlier, lowering an electron from the excited into the ground state and emitting a photon. The rate of spontaneous emission ρ_{spont} must therefore be proportional to the number of electrons in the excited state

$$\rho_{\text{spont}} = AN_2. \tag{3.1}$$

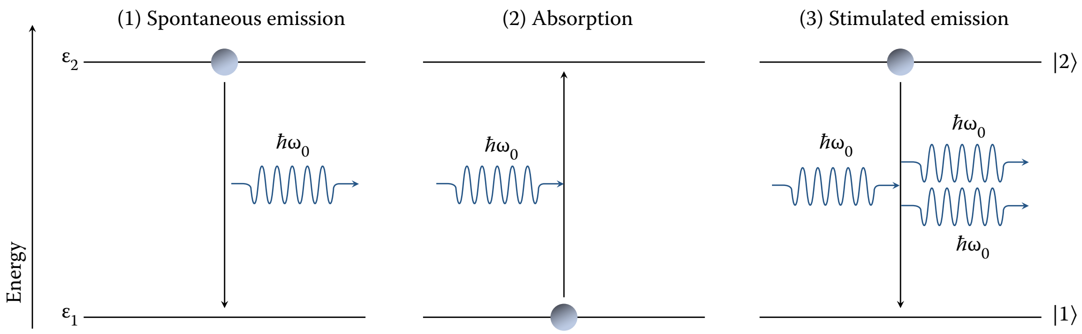


FIGURE 3.1 Relevant electron–photon interaction processes in a two-level system.

2. *Absorption*, raising an electron from the ground state into the excited state by absorbing an existing photon. The rate of absorption ρ_{abs} must be proportional to the number of electrons in the ground state and the photon density at the resonance frequency:

$$\rho_{\mathrm{abs}} = B_{12} N_1 \Phi_{\omega_0}. \tag{3.2}$$

3. *Stimulated emission*, where an electron is stimulated by a photon to emit a second photon and drop to the ground state in consequence. Both photons agree in frequency and phase and travel in the same direction. The rate of stimulated emission ρ_{stim} is proportional to the number of electrons in the excited state and the photon density at resonance

$$\rho_{\mathrm{stim}} = B_{21} N_2 \Phi_{\omega_0}. \tag{3.3}$$

A, B_{12}, and B_{21} are the *Einstein coefficients*, named after Albert Einstein (1879–1955), who first introduced them in 1916; see Ref. [3]. Considering all three processes, we get an expression for the overall rate of change of the occupancies of the two levels:

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = -\frac{\mathrm{d}N_2}{\mathrm{d}t} = -N_1 B_{12} \Phi_{\omega_0} + N_2 B_{21} \Phi_{\omega_0} + N_2 A. \tag{3.4}$$

Thinking of a thermal equilibrium situation, we demand

$$\frac{\mathrm{d}N_1}{\mathrm{d}t} = \frac{\mathrm{d}N_2}{\mathrm{d}t} = 0, \tag{3.5}$$

which, after substituting Equation 3.4 into 3.5 and making some algebraic conversions, renders

$$\frac{N_2}{N_1} = \frac{B_{12}}{B_{21} \Phi_{\omega_0} + A}. \tag{3.6}$$

On the other hand, thermodynamics tells us that the equilibrium distribution between ground and excited state is governed by the Boltzmann distribution,

$$\frac{N_2}{N_1} = \frac{e^{-\epsilon_2/(k_{\mathrm{B}} T)}}{e^{-\epsilon_1/(k_{\mathrm{B}} T)}} = e^{-\hbar\omega_0/(k_{\mathrm{B}} T)}, \tag{3.7}$$

where k_{B} is the Boltzmann constant. Equating Equation 3.6 with Equation 3.7 and solving for the photon density Φ_{ω_0} yields

$$\Phi_{\omega_0} = \frac{A}{B_{21}} \frac{1}{\frac{B_{12}}{B_{21}} e^{\hbar\omega_0/(k_{\mathrm{B}}T)} - 1}. \tag{3.8}$$

Now consider Planck's law,

$$\Phi_{\omega_0} = \frac{2\hbar{\omega_0}^3}{\pi c^3} \frac{1}{e^{\hbar\omega_0/(k_{\mathrm{B}}T)} - 1}, \tag{3.9}$$

where c is the velocity of light. Equations 3.8 and 3.9 are very similar in form. A comparison of the two yields

$$B_{12} = B_{21} \equiv B, \tag{3.10}$$

$$A = B\frac{2\hbar{\omega_0}^3}{\pi c^3}. \tag{3.11}$$

Thus, the coefficients of absorption and stimulated emission are equal. This also implies that the relative strength of absorption and stimulated emission is determined exclusively by the occupation of ground and excited state. The coefficient of spontaneous emission, A, is also directly and independently of temperature related to the other two Einstein coefficients.

3.2.2 The Two-Level System in the Schrödinger Picture

We now proceed to a quantum-mechanical treatment of the electron in the two-level system. At this point, we stick to a semiclassical treatment, i.e., a classical treatment of light. As we see, this takes us a long way toward understanding absorption and stimulated emission as well as deriving the optical Bloch equations. The reader is referred to the quantum-optical treatment presented in Section 3.4 to deal with spontaneous emission as it cannot be understood semiclassically.

The two-level system is described by the Hamiltonian of the undisturbed system, H_0, and the Hamiltonian of an optical field coupling to the electronic transition dipole moment, $\hat{H}_{\mathrm{L}}$:

$$\hat{H} = \hat{H}_0 + \hat{H}_{\mathrm{L}} \tag{3.12}$$

with

$$\begin{aligned} \hat{H}_0 &= \epsilon_1|1\rangle\langle 1| + \epsilon_2|2\rangle\langle 2|, \\ \hat{H}_{\mathrm{L}} &= -\mathbf{E}(t)\cdot\left\{\mathbf{d}_{12}|1\rangle\langle 2| + \mathbf{d}_{21}|2\rangle\langle 1|\right\}, \end{aligned} \tag{3.13}$$

using the Schrödinger wave functions of state 1 and state 2, $|1\rangle$ and $|2\rangle$, respectively. The light field couples to the transition dipole moment of the system, see also Ref. [4], using the transition dipole matrix elements $\mathbf{d}_{12} = -e\langle 1|\hat{\mathbf{r}}|2\rangle$ and $\mathbf{d}_{21} = \mathbf{d}_{12}^* = -e\langle 2|\hat{\mathbf{r}}|1\rangle$. The Schrödinger equation for the system wave function $|\psi\rangle$ reads

$$i\hbar\frac{\mathrm{d}}{\mathrm{d}t}|\psi\rangle = \left[\hat{H}_0 + \hat{H}_{\mathrm{L}}\right]|\psi\rangle. \tag{3.14}$$

The wave function $|\psi\rangle$ may be expanded in the eigenfunctions of the system,

$$|\psi\rangle = a_1(t)e^{-i\omega_1 t}|1\rangle + a_2(t)e^{-i\omega_2 t}|2\rangle, \tag{3.15}$$

using $\omega_1 = \epsilon_1/\hbar$ and $\omega_2 = \epsilon_2/\hbar$. Inserting this expansion into Equation 3.14 and multiplying with $\langle\psi|$ from the left-hand side yields equations for the coefficients $a_1(t)$ and $a_2(t)$,

$$\begin{aligned}\dot{a}_1(t) &= a_2(t)e^{-i(\omega_2-\omega_1)t}\langle 1|\hat{H}_{\mathrm{L}}|2\rangle,\\ \dot{a}_2(t) &= a_1(t)e^{-i(\omega_1-\omega_2)t}\langle 2|\hat{H}_{\mathrm{L}}|1\rangle.\end{aligned} \tag{3.16}$$

Applying a monochromatic incident field, $\mathbf{E}(t) = \frac{1}{2}\mathbf{E}_0(e^{-i\omega t} + e^{i\omega t})$, we derive

$$\begin{aligned}\dot{a}_1(t) &= \frac{a_2(t)}{2}\, e^{-i(\omega_2-\omega_1)t}\mathbf{E}_0 \cdot \mathbf{d}_{21}(e^{-i\omega t} + e^{i\omega t}),\\ \dot{a}_2(t) &= \frac{a_1(t)}{2}\, e^{-i(\omega_1-\omega_2)t}\mathbf{E}_0 \cdot \mathbf{d}_{12}(e^{-i\omega t} + e^{i\omega t}).\end{aligned} \tag{3.17}$$

Considering an optical field near the resonance frequency of the system ω_0, one of the addends in either equation is almost time independent and the other is oscillating quickly. In the rotating wave approximation (RWA), the latter term is ignored due to averaging out and we are left with

$$\begin{aligned}\dot{a}_1(t) &= \frac{a_2(t)}{2}\, e^{-i(\omega_0-\omega)t}\mathbf{E}_0 \cdot \mathbf{d}_{12},\\ \dot{a}_2(t) &= \frac{a_1(t)}{2}\, e^{i(\omega_0-\omega)t}\mathbf{E}_0 \cdot \mathbf{d}_{21}.\end{aligned} \tag{3.18}$$

For resonance conditions $\omega = \omega_0$, this system of coupled equations may easily be solved by time-differentiating the first equation and inserting the second, yielding

$$\frac{\mathrm{d}^2}{\mathrm{d}t^2}\, a_1(t) = -\frac{\left|\mathbf{E}_0 \cdot \mathbf{d}_{12}\right|^2}{4} a_1(t). \tag{3.19}$$

Defining the Rabi frequency[†]

$$\Omega = \frac{\left|\mathbf{E}_0 \cdot \mathbf{d}_{12}\right|}{\hbar}, \tag{3.20}$$

yields $a_1(t) = a_1(0)e^{\pm i\Omega t}$ as a solution to Equation 3.19, an equivalent expression for $a_2(t)$, and after resubstitution into Equation 3.15,

$$|\psi\rangle = a_1(0)e^{\pm i\Omega t}e^{-i\omega_1 t}|1\rangle + a_2(0)e^{\pm i\Omega t}e^{-i\omega_2 t}|2\rangle. \tag{3.21}$$

Starting with $a_1 = 0, a_2 = 1$, i.e., an electron in the excited state, we observe that the electron oscillates between ground and excited states with a cosine dependence and frequency Ω, describing the interplay between absorption and stimulated emission. These so-called *Rabi flops* are typical for any two-level system driven by an oscillatory field, including atoms in a light field as discussed here, but also particles in an oscillating magnetic field as investigated in nuclear magnetic resonance.

Looking at the optical processes possible in a two-level system with or without the incidence of light, we note the following: In the absence of the external perturbation represented by the laser pulse, the system is described by the Hamiltonian $\hat{H}_0$. If, at an initial time $t = 0$, it is in an eigenstate, it will remain in this state for any length of time. This applies to the ground state $|1\rangle$ as well as to the excited state $|2\rangle$.

[†] Named after American physicist Isidor Isaac Rabi (1898–1988).

It is, however, obviously not a realistic description, the reason for this being the absence of spontaneous emission in the semiclassical picture we used here. We see later that spontaneous emission enters the model with a quantization of light.

Now let us briefly consider the example of a laser system. Laser operation is basically as follows: An initial photon, usually produced by spontaneous emission, is confined to an active medium using a resonator and amplified by stimulated emission. Obviously, this will only work if stimulated emission exceeds absorption. Going back to Equation 3.10, we know that the Einstein coefficients for absorption and stimulated emission are equal and thus the occupancies of the ground and excited states govern the relative strength. Therefore, inversion has to be achieved for laser operation, i.e., the number of carriers in the excited state has to exceed the number of carriers in the ground state. Under certain conditions, quantum theory also allows lasing without inversion; since this requires intensive tailoring of the system, it will not be discussed here, but the reader is referred to the original literature, see, e.g., Refs. [5,6].

Unfortunately, inversion can, in general, not be achieved in a two-level system as there is no efficient pumping mechanism. Thus, a laser requires at least a three-level system for operation. The relevant equations for the three-level system may be derived in analogy to the case of the two-level system discussed here.

3.2.3 The Two-Level System in the Heisenberg Picture

We now introduce yet another description of the two-level system, changing from the Schrödinger picture to the Heisenberg picture. The passage from the Schrödinger picture to the Heisenberg picture is made by transferring the time dependence from the wave functions to the operator; for details, see any see any textbook on quantum mechanics. The formalism in this subsection is similar to the one in Section 3.3, which treats the many-body case. In this subsection, all calculations are presented in detail so the reader can reproduce them, while in Section 3.3, several steps are omitted for the sake of clarity and brevity.

The time evolution of the system in the Heisenberg picture is not governed by the Schrödinger equation any more, but by the Heisenberg equation of motion

$$\dot{\hat{\mathcal{O}}} = \frac{i}{\hbar}\left[\hat{H},\hat{\mathcal{O}}\right], \tag{3.22}$$

where [.] denotes the commutator,

$$\left[\hat{H},\hat{\mathcal{O}}\right] = \hat{H}\hat{\mathcal{O}} - \hat{\mathcal{O}}\hat{H}, \tag{3.23}$$

and $\hat{\mathcal{O}}$ is any operator.

The Heisenberg equation allows us to calculate the equation of motion for arbitrary operators. Here, we obviously only consider processes that conserve the number of carriers in the system. We revert to the formalism of second quantization mentioned in the introduction and introduce operators describing the electron. In the following, $\hat{c}_1$ ($\hat{c}_1^\dagger$) and $\hat{c}_2$ ($\hat{c}_2^\dagger$) describe annihilation (creation) operators for electrons in levels 1 and 2, respectively. The relevant particle-conserving two-operator quantities are thus $\hat{c}_1^\dagger\hat{c}_2$, $\hat{c}_2^\dagger\hat{c}_1$, $\hat{c}_1^\dagger\hat{c}_1$, and $\hat{c}_2^\dagger\hat{c}_2$. The first two quantities are the microscopic polarizations; it will be illustrated in Section 3.3 that the polarization is the relevant quantity when considering light–matter interaction. The latter two quantities $\hat{c}_1^\dagger\hat{c}_1$ and $\hat{c}_2^\dagger\hat{c}_2$ describe the carrier densities in levels 1 and 2, respectively. We observe that the equations of motion for these four quantities are coupled to each other, but form a closed system of equations.

The Hamiltonian $\hat{H} = \hat{H}_0 + \hat{H}_\mathrm{L}$ of the two-level system may be written as

$$\hat{H}_0 = \epsilon_1\hat{c}_1^\dagger\hat{c}_1 + \epsilon_2\hat{c}_2^\dagger\hat{c}_2,$$

$$\hat{H}_\mathrm{L} = -\mathbf{E}(t)\cdot(\mathbf{d}_{12}\hat{c}_1^\dagger\hat{c}_2 + \mathbf{d}_{21}\hat{c}_2^\dagger\hat{c}_1).$$

Inserting this Hamiltonian into the Heisenberg equation for the polarization $\hat{p} = \hat{c}_1^\dagger \hat{c}_2$, we derive the equation of motion for this quantity:

$$\begin{aligned}
\dot{\hat{p}} &= \frac{i}{\hbar}\left[\hat{H},\hat{p}\right] = \frac{i}{\hbar}\left(\hat{H}\hat{p} - \hat{p}\hat{H}\right) \\
&= \frac{i}{\hbar}\left\{\epsilon_1 \hat{c}_1^\dagger \hat{c}_1 + \epsilon_2 \hat{c}_2^\dagger \hat{c}_2 - \mathbf{E}(t)\cdot\left(\mathbf{d}_{12}\hat{c}_1^\dagger \hat{c}_2 + \mathbf{d}_{21}\hat{c}_2^\dagger \hat{c}_1\right)\right\} \hat{c}_1^\dagger \hat{c}_2 \\
&\quad - \hat{c}_1^\dagger \hat{c}_2 \left\{\epsilon_1 \hat{c}_1^\dagger \hat{c}_1 + \epsilon_2 \hat{c}_2^\dagger \hat{c}_2 - \mathbf{E}(t)\cdot\left(\mathbf{d}_{12}\hat{c}_1^\dagger \hat{c}_2 + \mathbf{d}_{21}\hat{c}_2^\dagger \hat{c}_1\right)\right\}.
\end{aligned} \tag{3.24}$$

The goal now is to relate the time derivative of the polarization to known or accessible quantities. This may also be the polarization itself, enabling an analytical or a numerical solution of the resulting differential equation. To achieve this, we need to interchange the operators, e.g., bring them into normal ordering in every term.

First, we remind ourselves of the Fermionic anticommutators to be used here

$$\left\{\hat{c}_i^\dagger,\, \hat{c}_j\right\}_+ = \hat{c}_i^\dagger \hat{c}_j + \hat{c}_j \hat{c}_i^\dagger = \delta_{ij}\,, \tag{3.25}$$

$$\left\{\hat{c}_i^\dagger,\, \hat{c}_j^\dagger\right\}_+ = 0\,,\; \left\{\hat{c}_i,\, \hat{c}_j\right\}_+ = 0, \tag{3.26}$$

the rules for interchanging two operators thus being

$$\begin{aligned}
&\hat{c}_i \hat{c}_j = -\hat{c}_j \hat{c}_i\,, \\
&\hat{c}_i^\dagger \hat{c}_j^\dagger = -\hat{c}_j^\dagger \hat{c}_i^\dagger\,, \\
&i \neq j \;\Rightarrow\; \hat{c}_i^\dagger \hat{c}_j = -\hat{c}_j \hat{c}_i^\dagger\,, \\
&i = j \;\Rightarrow\; \hat{c}_i^\dagger \hat{c}_i = \hat{1} - \hat{c}_i \hat{c}_i^\dagger
\end{aligned}$$

using the identity operator $\hat{1}$. The first two relations also show that the square of any Fermionic operator equals zero,

$$\hat{c}_i^\dagger \hat{c}_i^\dagger = \hat{c}_i \hat{c}_i = 0, \tag{3.27}$$

resulting from the Pauli principle[†] stating that no two fermions may occupy the same state.

From Equation 3.24, we calculate

$$\begin{aligned}
-i\hbar\dot{\hat{p}} &= \left(\epsilon_1 \hat{c}_1^\dagger \hat{c}_1 \hat{c}_1^\dagger \hat{c}_2 + \epsilon_2 \hat{c}_2^\dagger \hat{c}_2 \hat{c}_1^\dagger \hat{c}_2 - \mathbf{E}(t)\cdot\left(\mathbf{d}_{12}\hat{c}_1^\dagger \hat{c}_2 \hat{c}_1^\dagger \hat{c}_2 + \mathbf{d}_{21}\hat{c}_2^\dagger \hat{c}_1 \hat{c}_1^\dagger \hat{c}_2\right)\right) \\
&\quad - \left(\epsilon_1 \hat{c}_1^\dagger \hat{c}_2 \hat{c}_1^\dagger \hat{c}_1 + \epsilon_2 \hat{c}_1^\dagger \hat{c}_2 \hat{c}_2^\dagger \hat{c}_2 - \mathbf{E}(t)\cdot\left(\mathbf{d}_{12}\hat{c}_1^\dagger \hat{c}_2 \hat{c}_1^\dagger \hat{c}_2 + \mathbf{d}_{21}\hat{c}_1^\dagger \hat{c}_2 \hat{c}_2^\dagger \hat{c}_1\right)\right) \\
&= \left(\epsilon_1 \hat{c}_1^\dagger \hat{c}_1 \hat{c}_1^\dagger \hat{c}_2 - \epsilon_1 \hat{c}_1^\dagger \hat{c}_2 \hat{c}_1^\dagger \hat{c}_1 + \epsilon_2 \hat{c}_2^\dagger \hat{c}_2 \hat{c}_1^\dagger \hat{c}_2 - \epsilon_2 \hat{c}_1^\dagger \hat{c}_2 \hat{c}_2^\dagger \hat{c}_2\right) \\
&\quad - \mathbf{E}(t)\cdot\left(\mathbf{d}_{21}\hat{c}_2^\dagger \hat{c}_1 \hat{c}_1^\dagger \hat{c}_2 - \mathbf{d}_{21}\hat{c}_1^\dagger \hat{c}_2 \hat{c}_2^\dagger \hat{c}_1\right),
\end{aligned}$$

[†] Named after Austrian-born Swiss and American physicist Wolfgang Pauli (1900–1958).

and, commuting the terms into normal order by using the Fermionic anticommutators,

$$\begin{aligned}-i\hbar\dot{\hat{p}} &= \left(\epsilon_1\hat{c}_1^\dagger(\hat{1}-\hat{c}_1^\dagger\hat{c}_1)\hat{c}_2+\epsilon_1\hat{c}_1^\dagger\hat{c}_1^\dagger\hat{c}_2\hat{c}_1-\epsilon_2\hat{c}_2^\dagger\hat{c}_1^\dagger\hat{c}_2\hat{c}_2-\epsilon_2\hat{c}_1^\dagger(\hat{1}-\hat{c}_2^\dagger\hat{c}_2)\hat{c}_2\right)\\ &\quad-\mathbf{E}(t)\cdot\mathbf{d}_{21}\left(\hat{c}_2^\dagger(\hat{1}-\hat{c}_1^\dagger\hat{c}_1)\hat{c}_2-\hat{c}_1^\dagger(\hat{1}-\hat{c}_2^\dagger\hat{c}_2)\hat{c}_1\right).\end{aligned} \tag{3.28}$$

The four-operator terms either cancel or disappear due to Equation 3.27 and we are left with

$$\begin{aligned}-i\hbar\dot{\hat{p}} &= \left(\epsilon_1\hat{c}_1^\dagger\hat{c}_2-\epsilon_2\hat{c}_1^\dagger\hat{c}_2\right)-\mathbf{E}(t)\cdot\mathbf{d}_{21}\left(\hat{c}_2^\dagger\hat{c}_2-\hat{c}_1^\dagger\hat{c}_1\right)\\ &= \left(\epsilon_1-\epsilon_2\right)\hat{p}+\mathbf{E}(t)\cdot\mathbf{d}_{21}\left(\hat{n}_1-\hat{n}_2\right).\end{aligned} \tag{3.29}$$

Obviously, this equation couples to the equations for the occupation of the two levels, $\hat{n}_1=\hat{c}_1^\dagger\hat{c}_1, \hat{n}_2=\hat{c}_2^\dagger\hat{c}_2$. Deriving the equations of motion for these quantities in the same way, defining the energy difference in the system as $\hbar\omega_0=\epsilon_2-\epsilon_1$ as before and taking the expectation values lead to the optical Bloch equations (also called Maxwell–Bloch equations)

$$\begin{aligned}-i\hbar\dot{p}+\hbar\omega_0 p &= \mathbf{E}(t)\cdot\mathbf{d}_{21}(n_1-n_2),\\ -i\hbar\dot{n}_1 &= \mathbf{E}(t)\cdot(\mathbf{d}_{12}p-\mathbf{d}_{21}p^*),\\ -i\hbar\dot{n}_2 &= \mathbf{E}(t)\cdot(\mathbf{d}_{21}p^*-\mathbf{d}_{12}p).\end{aligned}$$

These equations are completely analogous to the Bloch equations in spin-$\frac{1}{2}$-systems (where a two-level system in this chapter is not to be understood as a two-level system containing electrons *with* spin but rather as a model system whose Hilbert space is isomorphic to a spin-$\frac{1}{2}$-system).

Let us now briefly discuss the electron–hole picture at this level. Starting from the ground state where the electrons are in level 1, often only a small fraction of the electrons is lifted to the upper level. It is then frequently advantageous to count the empty spaces in the ground state rather than the electrons. These "holes" carry positive charge because the missing electron implies lack of compensation of nuclear charge. We define

$$\begin{aligned}n_2 &= n_e &&\text{upper level}\\ n_1 &= 1-n_h &&\text{lower level}\\ n_1+n_2 &= 1 &&\Rightarrow n_e=n_h\equiv n\end{aligned}$$

and rewrite the optical Bloch equations

$$\begin{aligned}\dot{p} &= -i\omega_0 p &&+\tfrac{i}{\hbar}\mathbf{E}(t)\cdot\mathbf{d}_{21}(1-2n),\\ \dot{n} &= &&\tfrac{i}{\hbar}\mathbf{E}(t)\cdot(\mathbf{d}_{21}p-\mathbf{d}_{12}p^*).\end{aligned}$$

For low excitation powers, where the predominant part of the carriers stays in the lower level, n may be neglected (linear case, $n\approx 0$) and a single equation of motion for the polarization remains

$$\dot{p}=-i\omega_0 p+\frac{i}{\hbar}\mathbf{E}(t)\cdot\mathbf{d}_{21}. \tag{3.30}$$

Solving this for a specific light field, one discovers that even after the end of excitation, the polarization remains forever, the reason being that relaxation processes have been neglected thus far. This may be remedied by introducing a phenomenological damping into the equations of motion

$$\begin{aligned}\dot{p} &= -i\omega_0 p &&+\tfrac{i}{\hbar}\mathbf{E}(t)\cdot\mathbf{d}_{21}(1-2n)-\tfrac{p}{T_2},\\ \dot{n} &= &&\tfrac{i}{\hbar}\mathbf{E}(t)\cdot(\mathbf{d}_{21}p-\mathbf{d}_{12}p^*)-\tfrac{n}{T_1},\end{aligned}$$

where T_1 and T_2 are relaxation times, which are usually derived by fits to experiment. The inclusion of relaxation times results in a broadening of the spectrum, which is called homogeneous because all emitters experience the same type of broadening mechanism; see Ref. [2]. The counterpart to this mechanism is inhomogeneous broadening, which is caused by disorder; see also Section 3.3; here, different emitters experience different environmental influences.

We now solve the polarization equation of motion in the relaxation time approach to illustrate that it leads to a Lorentzian lineshape of the polarization and absorption spectrum. Defining $\Gamma = 1/T_2$, we calculate

$$\frac{\mathrm{d}}{\mathrm{d}t}\left\{pe^{(i\omega_0+\Gamma)t}\right\} = \frac{i}{\hbar}\mathbf{E}(t)\cdot\mathbf{d}_{21}(1-2n(t))e^{(i\omega_0+\Gamma)t} \tag{3.31}$$

and formally integrate this equation, deriving

$$p(t) = \frac{i}{\hbar}\int_{-\infty}^{t}\mathrm{d}t'\,e^{(i\omega_0+\Gamma)(t'-t)}\mathbf{E}(t')\cdot\mathbf{d}_{21}(1-2n(t')). \tag{3.32}$$

We now insert a monochromatic incident field

$$\mathbf{E}(t) = \frac{1}{2}\mathbf{E}_0\left\{e^{-i\omega t} + e^{i\omega t}\right\} \tag{3.33}$$

and assume a quasistationary or stationary state of the system, i.e., the carrier density changes little on the timescale of T_2. We may then replace $n(t')$ by $n(t)$ and get

$$p(t) = \frac{i}{2\hbar}\mathbf{E}_0\cdot\mathbf{d}_{21}(1-2n(t))\int_{-\infty}^{t}\mathrm{d}t'\,e^{(i\omega_0+\Gamma)(t'-t)}\left\{e^{-i\omega t'} + e^{i\omega t'}\right\}. \tag{3.34}$$

This replacement actually means that the prediction of the system behavior is made based only on the present state and not on the history of the system. This is a so-called "memoryless" process obtained using the Markov approximation.

The integral in Equation 3.34 may now be solved, yielding

$$\begin{aligned}
p(t) &= \frac{i}{2\hbar}\mathbf{E}_0\cdot\mathbf{d}_{21}(1-2n(t))e^{-(i\omega_0+\Gamma)t}\int_{-\infty}^{t}\mathrm{d}t'\left\{e^{(i(\omega_0-\omega)+\Gamma)t'} + e^{(i(\omega_0+\omega)+\Gamma)t'}\right\}\\
&= \frac{i}{2\hbar}\mathbf{E}_0\cdot\mathbf{d}_{21}(1-2n(t))e^{-(i\omega_0+\Gamma)t}\left[\frac{e^{(i(\omega_0-\omega)+\Gamma)t'}}{i(\omega_0-\omega)+\Gamma} + \frac{e^{(i(\omega_0+\omega)+\Gamma)t'}}{i(\omega_0+\omega)+\Gamma}\right]_{-\infty}^{t}\\
&= \frac{i}{2\hbar}\mathbf{E}_0\cdot\mathbf{d}_{21}(1-2n(t))e^{-(i\omega_0+\Gamma)t}\left\{\frac{e^{(i(\omega_0-\omega)+\Gamma)t}}{i(\omega_0-\omega)+\Gamma} + \frac{e^{(i(\omega_0+\omega)+\Gamma)t}}{i(\omega_0+\omega)+\Gamma}\right\}.
\end{aligned} \tag{3.35}$$

Considering resonant or near-resonant excitation, the denominator $i(\omega_0+\omega)+\Gamma$ can be considered as large and the second addend may be neglected. This corresponds to the RWA, which was already used in getting from Equation 3.17 to 3.18 in Section 3.2.2. We are left with

$$p(t) = \frac{i}{2\hbar}\mathbf{E}_0e^{-i\omega t}\cdot\mathbf{d}_{21}(1-2n(t))\frac{1}{i(\omega_0-\omega)+\Gamma}. \tag{3.36}$$

In the last fraction of Equation 3.36, $(i(\omega_0-\omega)+\Gamma)^{-1}$, the Lorentzian lineshape of the absorption spectrum becomes apparent, the absorption being approximately proportional to the imaginary part of the expression (see Ref. [7,8], Equation 3.53 in this chapter) and

$$\frac{1}{i(\omega_0 - \omega) + \Gamma} = \frac{\Gamma}{(\omega_0 - \omega)^2 + \Gamma^2} + i\frac{\omega_0 - \omega}{(\omega_0 - \omega)^2 + \Gamma^2}. \tag{3.37}$$

In Section 3.3, we discuss the drawbacks of the relaxation time approach. Generally, it constitutes a simple and easily comprehensible procedure, where, however, the relaxation times must be inserted by hand and are not calculated by the theory.

3.3 The Solid-State System

We now move on to a much more complex system, the semiconductor. Not only do we encounter a system of many atoms forming an energy landscape known as band structure here (see also Chapter 1), but, in addition, we also need to take into account a large number of electrons. However, a semiconductor is what the diodes, amplifiers, laser diodes, optical modulators, and solar cells discussed in this book are made out of. In order to get a flavor of a many-body treatment of electron–photon interaction, the derivation of the semiconductor Bloch equations will be discussed briefly here. For an in-depth treatment, the reader is referred to Ref. [7], the references therein, and a wealth of original literature. Since a many-body problem can only be solved approximately, for every case the equations have to be examined separately and the proper approximations made. This is far beyond the scope of this chapter.

In optoelectronics, we customarily deal with direct-bandgap semiconductors. Here, we consider the simplest possible case, a direct-gap two-band semiconductor and derive the relevant equations of motion. It comprises a valence and a conduction band, where in the ground state, all the electrons are in the valence band and fill up its states completely. Similar to the two-level system, they may be raised to the conduction band by optical excitation. Figure 3.2 shows a simplified schematic of the direct-gap two-band semiconductor. Equations of motion for more complicated systems may be obtained by expanding the indices in the derivation below to all the existing bands; a derivation for a multiband system is thus straightforward but more cumbersome.

Looking at the semiconductor–light interaction, we need to keep in mind that the size of an atom is in the range of a single Ångström, while the wavelength of visible light is approximately 400–700 nm, i.e., 4000–7000 Ångström. The electric field is thus approximately constant over the extension of a single atom, and all quantities in Maxwell's equations[†] are averaged over a large number of atoms.

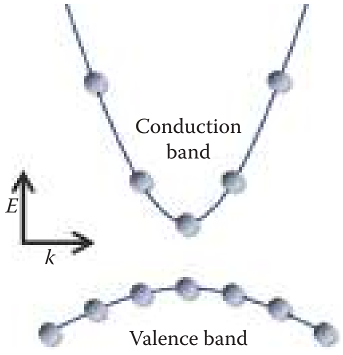


FIGURE 3.2 Schematic of a two-band direct-gap semiconductor.

[†] Named after Scottish physicist James Clerk Maxwell (1831–1879).

The Hamiltonian of the solid-state system is made up of three contributions,

$$\hat{H} = \hat{H}_{el} + \hat{H}_{\mathrm{L}} = (\hat{H}_0 + \hat{H}_C) + \hat{H}_{\mathrm{L}}, \tag{3.38}$$

where $\hat{H}_{el}$ is the Hamiltonian of the electronic system, consisting of the sum of the individual single particle contributions, $\hat{H}_0$, and the Coulomb interaction between the electrons, $\hat{H}_C$. As before, $\hat{H}_{\mathrm{L}}$ denotes the interaction with the light field.

The structure of an inorganic semiconductor is characterized by its crystallinity, i.e., periodicity. In other words, the environment of an electron is the same in position $\mathbf{r}$ and position $\mathbf{r}+\mathbf{R}_n$, where $\mathbf{R}_n$ is any lattice vector, i.e., a vector linking two identical sites in an infinite lattice. The periodicity is taken into account in the Bloch theorem[†] for the wave functions,

$$\psi_\lambda(\mathbf{k}, \mathbf{r}+\mathbf{R}_n) = e^{i\mathbf{k}\cdot\mathbf{R}_n}\psi_\lambda(\mathbf{k}, \mathbf{r}). \tag{3.39}$$

Consequently, the Hamiltonian is expanded in electronic Bloch wave functions,

$$\psi_\lambda(\mathbf{k}, \mathbf{r}) \approx \mathrm{e}^{i\mathbf{k}\cdot\mathbf{r}}\,\frac{u_\lambda(\mathbf{k}, \mathbf{r})}{L^{\frac{3}{2}}}. \tag{3.40}$$

L^3 is the volume of the crystal and λ denotes the band index, comprising the conduction band $\lambda = c$ and the valence band $\lambda = v$. Bloch functions take into account the lattice periodicity of the system by including a lattice periodic function $u_\lambda(\mathbf{k}, \mathbf{r})$,

$$u_\lambda(\mathbf{k}, \mathbf{r}+\mathbf{R}_n) = u_\lambda(\mathbf{k}, \mathbf{r}). \tag{3.41}$$

The introduction of the crystal volume L^3 necessitates further explanation, as Equation 3.39 is only valid in an infinite system. Customary crystals are so large that the assumption of infinity is quite fitting. Since for a proper normalization of the wave functions, one needs a finite crystal volume, we consider the (approximately infinite) crystal to be made up of mesoscopic cubes of size L^3 and demand periodic boundary conditions of the wave function, i.e., the wave function has to have the same value on either side of the cube.

For the single particle contribution, we derive

$$\begin{aligned}\hat{H}_0 &= \sum_{\mathbf{k},\lambda=v,c} \epsilon_\lambda(\mathbf{k})\,\hat{c}^\dagger_{\lambda,\mathbf{k}}\hat{c}_{\lambda,\mathbf{k}} \\ &= \sum_{\mathbf{k}} \epsilon_v(\mathbf{k})\,\hat{c}^\dagger_{v,\mathbf{k}}\hat{c}_{v,\mathbf{k}} + \sum_{\mathbf{k}} \epsilon_c(\mathbf{k})\,\hat{c}^\dagger_{c,\mathbf{k}}\hat{c}_{c,\mathbf{k}}.\end{aligned} \tag{3.42}$$

$\hat{c}_{\lambda,\mathbf{k}}$ ($\hat{c}^\dagger_{\lambda,\mathbf{k}}$) is the annihilation (creation) operator of an electron in state $\lambda, \mathbf{k}$.

Since the energies $\epsilon_\lambda(\mathbf{k})$ are rather difficult to assess, parabolic bands are often assumed, corresponding to the effective mass approximation. At the extrema of the band structure, the linear term in the Taylor expansion becomes zero, such that a parabolic approximation may be made. Formally, a constant called the “effective mass,” $m_{\lambda,\mathrm{eff}} = \frac{\hbar^2}{\left.\frac{\partial^2\epsilon_\lambda(\mathbf{k})}{\partial k^2}\right|_{k=0}}$, may then be introduced,

$$\begin{aligned}\epsilon_\lambda(\mathbf{k}) &= \epsilon_\lambda(\mathbf{k}=0) + \frac{1}{2}\frac{\partial^2\epsilon_\lambda(\mathbf{k})}{\partial k^2} + O(k^3) \\ &\approx \epsilon_\lambda(\mathbf{k}=0) + \frac{\hbar^2k^2}{2m_{\lambda,\mathrm{eff}}}.\end{aligned} \tag{3.43}$$

[†] Named after Swiss physicist Felix Bloch (1905–1983).

Here, $O(k^3)$ denotes higher-order terms. For the Bloch functions, this implies that the lattice-periodic part becomes momentum independent, i.e., $u_\lambda(\mathbf{k}, \mathbf{r}) \approx u_\lambda(\mathbf{k} = 0, \mathbf{r})$. This approximation is always valid in a certain range around an extremum of the band structure but is used extensively in calculations even far away from these points. In the effective-mass approximation, the electrons behave like free particles with the effective mass $m_{\lambda,\text{eff}}$ rather than the free-particle mass. $m_{\lambda,\text{eff}}$ may be positive or negative, examples being the conduction and valence band in the case of the two-band semiconductor we are looking at. When the sign is negative, this is often included into the energy by transferring to the electron–hole picture. Furthermore, the effective mass, in general, is not independent of direction and must be written as a tensor,

$$\left(\frac{1}{m_{\lambda,\text{eff}}}\right)_{\alpha\beta} = \frac{1}{\hbar^2} \left.\frac{\partial^2 \epsilon_\lambda(\mathbf{k})}{\partial k_\alpha k_\beta}\right|_{k=0}. \tag{3.44}$$

The direction dependence of the effective mass is discussed, e.g., in Refs. [9–11].

Returning to light–matter coupling in semiconductors, we have seen that light couples to electrical dipoles,

$$\hat{H}_\text{L} = -\mathbf{E}(t) \cdot (e\hat{\mathbf{r}}) = -\mathbf{E}(t) \cdot \hat{\mathbf{d}}. \tag{3.45}$$

Expanding the interaction Hamiltonian in these wave functions yields

$$\begin{aligned}
\hat{H}_\text{L} &= -e\,\mathbf{E}(t) \cdot \sum_{\mathbf{k},\lambda,\lambda'} \langle \lambda', \mathbf{k} | \, \mathbf{r} \, | \lambda, \mathbf{k} \rangle \, \hat{c}^\dagger_{\lambda',\mathbf{k}} \hat{c}_{\lambda,\mathbf{k}} \\
&= -\mathbf{E}(t) \cdot \left[\sum_{\mathbf{k},\lambda,\lambda',\lambda\neq\lambda'} \mathbf{d}_{\lambda',\lambda}(\mathbf{k}) \, \hat{c}^\dagger_{\lambda',\mathbf{k}} \hat{c}_{\lambda,\mathbf{k}} \right] \\
&= -\mathbf{E}(t) \cdot \sum_{\mathbf{k}} \left[\mathbf{d}_{cv}(\mathbf{k}) \, \hat{c}^\dagger_{c,\mathbf{k}} \hat{c}_{v,\mathbf{k}} + \mathbf{d}_{vc}(\mathbf{k}) \, \hat{c}^\dagger_{v,\mathbf{k}} \hat{c}_{c,\mathbf{k}} \right] \\
&= -\mathbf{E}(t) \cdot \hat{\mathbf{P}},
\end{aligned} \tag{3.46}$$

where $\hat{\mathbf{P}}$ is the optical interband polarization. $\mathbf{d}_{\lambda',\lambda}(\mathbf{k})$ are the band and momentum-dependent optical transition dipole matrix elements with in-plane momentum $\mathbf{k}$ and subband labels λ, λ', corresponding to the transition dipole matrix elements $\mathbf{d}_{12}, \mathbf{d}_{21}$ in the two-level system.

To obtain Equation 3.46, the following approximations are made: (1) the photon momentum is neglected and (2) intraband transitions are not considered. Rather than retracing the entire derivation,[†] let us discuss these approximations:

1. Neglect of the photon momentum: In this book, we concentrate on light in and around the visible range, i.e., with wavelengths roughly between 400 and 700 nm. For a central wavelength of 550 nm, the wave number is $k = 2\pi/\lambda = 11 \times 10^6\ \text{m}^{-1}$, which is very small compared to typical wave vectors in a semiconductor band structure. Photonic transitions are consequently vertical in a band diagram.
2. Neglect of intraband transitions: Here, only interband transitions will be considered, assuming an excitation at an energy near the band gap. Suppression of intraband transitions in a regular structure becomes obvious from the vanishing photon momentum as discussed earlier.

Looking at the system Hamiltonian, Equation 3.38, we still need to calculate the term $\hat{H}_C$ describing Coulomb interaction between the carriers. $\hat{H}_C$ is specified using the so-called jellium model, which substitutes the ionic background of the lattice with a smooth background. The derivation is not included here, but may be found in Ref. [7]. The final expression of the Hamiltonian is

---

[†] See Ref. [7] for details.

$$\begin{aligned}\hat{H}_C &= \sum_{\mathbf{k},\mathbf{k}',\mathbf{q}\neq 0,\lambda,\lambda'=v,c} V_q\, \hat{c}^\dagger_{\lambda,\mathbf{k}+\mathbf{q}} \hat{c}^\dagger_{\lambda',\mathbf{k}'-\mathbf{q}} \hat{c}_{\lambda',\mathbf{k}'} \hat{c}_{\lambda,\mathbf{k}} \\ &= \sum_{\mathbf{k},\mathbf{k}',\mathbf{q}\neq 0} V_q \left[\hat{c}^\dagger_{v,\mathbf{k}+\mathbf{q}} \hat{c}^\dagger_{v,\mathbf{k}'-\mathbf{q}} \hat{c}_{v,\mathbf{k}'} \hat{c}_{v,\mathbf{k}} + \hat{c}^\dagger_{c,\mathbf{k}+\mathbf{q}} \hat{c}^\dagger_{c,\mathbf{k}'-\mathbf{q}} \hat{c}_{c,\mathbf{k}'} \hat{c}_{c,\mathbf{k}} + 2\hat{c}^\dagger_{c,\mathbf{k}+\mathbf{q}} \hat{c}^\dagger_{v,\mathbf{k}'-\mathbf{q}} \hat{c}_{v,\mathbf{k}'} \hat{c}_{c,\mathbf{k}} \right],\end{aligned} \tag{3.47}$$

where the second line spells out the interaction for our two-band model. $\hat{H}_C$ thus describes the repulsion between electrons in the valence band, electrons in the conduction band and between valence and conduction band. V_q is the Fourier transform of the Coulomb potential,

$$V(\mathbf{r}) = \frac{e^2}{4\pi\mathcal{E}_0\mathcal{E}_r r}, \tag{3.48}$$

$$V_q = \int \frac{\mathrm{d}^3 r}{L^3} V(\mathbf{r}) \mathrm{e}^{i\mathbf{q}\cdot\mathbf{r}} = \frac{e^2}{\mathcal{E}_0\mathcal{E}_r L^3 q^2}, \tag{3.49}$$

where for the calculation a convergence generating factor needs to be used. $\mathcal{E}_0$ and $\mathcal{E}_r$ are the vacuum and relative material permittivity, respectively. The sum in Equation 3.47 omits the case $\mathbf{q} = 0$ where obviously the Coulomb matrix element has a singularity; this term is canceled by the ionic background charge in the jellium model.

The complete Hamiltonian is thus

$$\begin{aligned}\hat{H} &= \hat{H}_0 + \hat{H}_C + \hat{H}_\mathrm{L} \\ &= \sum_{\mathbf{k},\lambda=v,c} \epsilon_\lambda(\mathbf{k})\, \hat{c}^\dagger_{\lambda,\mathbf{k}} \hat{c}_{\lambda,\mathbf{k}} + \sum_{\mathbf{k},\mathbf{k}',\mathbf{q}\neq 0,\lambda,\lambda'=v,c} V_q\, \hat{c}^\dagger_{\lambda,\mathbf{k}+\mathbf{q}} \hat{c}^\dagger_{\lambda',\mathbf{k}'-\mathbf{q}} \hat{c}_{\lambda',\mathbf{k}'} \hat{c}_{\lambda,\mathbf{k}} \\ &\quad - \mathbf{E}(t) \cdot \left[\sum_{\mathbf{k},\lambda,\lambda',\lambda\neq\lambda'} \mathbf{d}_{\lambda',\lambda}(\mathbf{k},\mathbf{k})\, \hat{c}^\dagger_{\lambda',\mathbf{k}} \hat{c}_{\lambda,\mathbf{k}} \right].\end{aligned} \tag{3.50}$$

This Hamiltonian is now going to be used for the calculation of the optical polarization $\hat{\mathbf{P}}$,

$$\begin{aligned}\hat{\mathbf{P}} &= \sum_{\mathbf{k}} \left(\mathbf{d}_{vc}\, \hat{c}^\dagger_{v,\mathbf{k}} \hat{c}_{c,\mathbf{k}} + \mathbf{d}_{cv}\, \hat{c}^\dagger_{c,\mathbf{k}} \hat{c}_{v,\mathbf{k}} \right) \\ &= \sum_{\mathbf{k}} \left(\mathbf{d}_{vc}\, \hat{c}^\dagger_{v,\mathbf{k}} \hat{c}_{c,\mathbf{k}} + \text{h.c.} \right) = \sum_{\mathbf{k}} \left(\mathbf{d}_{vc}\, \hat{p}_{vc,\mathbf{k}} + \text{h.c.} \right),\end{aligned} \tag{3.51}$$

where $p_{vc,\mathbf{k}} = \langle \hat{p}_{vc,\mathbf{k}} \rangle$ are the microscopic interband coherences or polarizations and h.c. denotes the Hermitian conjugate. The optical polarization is a quantity that describes the reaction of matter to a light field, coupling light (electric field $\mathbf{E}(\mathbf{r}, t)$) to matter in the wave equation

$$\Delta\mathbf{E}(\mathbf{r}, t) - \left(\frac{n}{c}\right)^2 \frac{\partial^2}{\partial t^2} \mathbf{E}(\mathbf{r}, t) = \mu_0 \frac{\partial^2}{\partial t^2} \mathbf{P}(\mathbf{r}, t) \tag{3.52}$$

with the index of refraction n and the vacuum permeability μ_0. The optical polarization is a fundamental parameter from which numerous optical variables can be calculated, e.g., the absorption α or laser gain g being given as

$$\alpha(\omega) = -g(\omega) = \frac{\omega}{nc} \mathrm{Im} \left\{ \frac{P(\omega)}{L^3 E(\omega)} \right\}, \tag{3.53}$$

see Ref. [8]. In Equation 3.53, the index of refraction n was assumed to be constant in the region of interest. Calculation of the optical polarization thus gives us deep insight into the optical properties of matter and allows us to simulate optical devices such as the ones discussed in this book.

Returning to the actual calculation of the optical polarization, the next step is the insertion of the many-body Hamiltonian into the Heisenberg equation of motion for the microscopic polarizations

$$\frac{\mathrm{d}}{\mathrm{d}t}\hat{c}^{\dagger}_{v,\mathbf{k}}\hat{c}_{c,\mathbf{k}} = \frac{i}{\hbar}\left[\hat{H}, \hat{c}^{\dagger}_{v,\mathbf{k}}\hat{c}_{c,\mathbf{k}}\right]. \tag{3.54}$$

Calculating the appropriate commutators and taking the expectation values on both sides renders

$$\begin{aligned}\left[i\hbar\frac{\mathrm{d}}{\mathrm{d}t} - (\epsilon_c(\mathbf{k}) - \epsilon_v(\mathbf{k}))\right] p_{vc,\mathbf{k}} &= (n_{c,\mathbf{k}} - n_{v,\mathbf{k}})\mathbf{d}_{cv}\cdot\mathbf{E}(t)\\ &+ \sum_{\mathbf{k}',\mathbf{q}\neq 0} V_q\left[\langle\hat{c}^{\dagger}_{c,\mathbf{k}'+\mathbf{q}}\hat{c}^{\dagger}_{v,\mathbf{k}-\mathbf{q}}\hat{c}_{c,\mathbf{k}'}\hat{c}_{c,\mathbf{k}}\rangle + \langle\hat{c}^{\dagger}_{v,\mathbf{k}'+\mathbf{q}}\hat{c}^{\dagger}_{v,\mathbf{k}-\mathbf{q}}\hat{c}_{v,\mathbf{k}'}\hat{c}_{c,\mathbf{k}}\rangle\right.\\ &\left.+\langle\hat{c}^{\dagger}_{v,\mathbf{k}}\hat{c}^{\dagger}_{c,\mathbf{k}'-\mathbf{q}}\hat{c}_{c,\mathbf{k}'}\hat{c}_{c,\mathbf{k}-\mathbf{q}}\rangle + \langle\hat{c}^{\dagger}_{v,\mathbf{k}}\hat{c}^{\dagger}_{v,\mathbf{k}'-\mathbf{q}}\hat{c}_{v,\mathbf{k}'}\hat{c}_{c,\mathbf{k}-\mathbf{q}}\rangle\right],\end{aligned} \tag{3.55}$$

where $n_{c,\mathbf{k}} = \langle\hat{c}^{\dagger}_{c,\mathbf{k}}\hat{c}_{c,\mathbf{k}}\rangle$ and $n_{v,\mathbf{k}} = \langle\hat{c}^{\dagger}_{v,\mathbf{k}}\hat{c}_{v,\mathbf{k}}\rangle$ are the carrier densities in the conduction and valence band, respectively. The transition dipole matrix element was assumed to be real, $\mathbf{d}_{cv} = \mathbf{d}_{vc}$.

Obviously, the polarization as a two-operator quantity on the left-hand side of Equation 3.55 couples to four-operator quantities on the right side via the Coulomb interaction. Computing the equation of motion for these four-operator terms will in turn lead to a coupling to six-operator quantities, which will then couple to eight-operator quantities and so on, with ever-rising computational cost. This so-called *infinite hierarchy problem*, which is an expression of the many-body character of the electron system, necessitates a truncation at some level. A relatively simple approximation at two-operator level is the dynamic Hartree–Fock approximation.† Roughly speaking, it assumes that any particle experiences the impact of all other particles as a mean field created by them. Technically, in Hartree–Fock approximation, we split four-operator quantities into all possible combinations of two-operator quantities. For example, the first Coulomb term of Equation 3.55 yields

$$\begin{aligned}\langle\hat{c}^{\dagger}_{c,\mathbf{k}'+\mathbf{q}}\hat{c}^{\dagger}_{v,\mathbf{k}-\mathbf{q}}\hat{c}_{c,\mathbf{k}'}\hat{c}_{c,\mathbf{k}}\rangle &\approx \langle\hat{c}^{\dagger}_{c,\mathbf{k}'+\mathbf{q}}\hat{c}_{c,\mathbf{k}}\rangle\langle\hat{c}^{\dagger}_{v,\mathbf{k}-\mathbf{q}}\hat{c}_{c,\mathbf{k}'}\rangle - \langle\hat{c}^{\dagger}_{c,\mathbf{k}'+\mathbf{q}}\hat{c}_{c,\mathbf{k}'}\rangle\langle\hat{c}^{\dagger}_{v,\mathbf{k}-\mathbf{q}}\hat{c}_{c,\mathbf{k}}\rangle\\ &\approx \delta_{\mathbf{k},\mathbf{k}'+\mathbf{q}}\langle\hat{c}^{\dagger}_{c,\mathbf{k}}\hat{c}_{c,\mathbf{k}}\rangle\langle\hat{c}^{\dagger}_{v,\mathbf{k}-\mathbf{q}}\hat{c}_{c,\mathbf{k}-\mathbf{q}}\rangle - \delta_{\mathbf{k}',\mathbf{k}'-\mathbf{q}}\langle\hat{c}^{\dagger}_{c,\mathbf{k}'}\hat{c}_{c,\mathbf{k}'}\rangle\langle\hat{c}^{\dagger}_{v,\mathbf{k}}\hat{c}_{c,\mathbf{k}}\rangle\\ &= \delta_{\mathbf{k},\mathbf{k}'+\mathbf{q}}n_{c,\mathbf{k}}p_{vc,\mathbf{k}-\mathbf{q}}\,,\end{aligned}$$

using $\mathbf{q} \neq 0$. In the second and third lines, we have taken into account that only terms that are diagonal in momentum are excited in a spatially homogeneous system. Approximating the other terms likewise, the resulting equation of motion for $p_{vc,\mathbf{k}}$ is

$$\left[i\hbar\frac{\mathrm{d}}{\mathrm{d}t} - (e_c(\mathbf{k}) - e_v(\mathbf{k}))\right] p_{vc,\mathbf{k}} = [n_{c,\mathbf{k}} - n_{v,\mathbf{k}}]\left(\mathbf{d}_{cv}\cdot\mathbf{E}(t) + \sum_{\mathbf{q}\neq\mathbf{k}} V_{|\mathbf{k}-\mathbf{q}|}p_{vc,\mathbf{q}}\right), \tag{3.56}$$

---

† Named after English mathematician and physicist Douglas Rayner Hartree (1897–1958) and Soviet physicist Vladimir Aleksandrovich Fock (1898–1974).

with renormalized energies

$$e_\lambda(\mathbf{k}) = \epsilon_\lambda(\mathbf{k}) - \sum_{\mathbf{q}\neq\mathbf{k}} V_{|\mathbf{k}-\mathbf{q}|} n_{\lambda,\mathbf{q}} \equiv \epsilon_\lambda(\mathbf{k}) + \Sigma_{exc,\lambda}(\mathbf{k}), \tag{3.57}$$

where $\Sigma_{exc}(\mathbf{k})$ is the exchange self-energy.

Equation 3.56 is not yet closed, but couples to the carrier densities $n_{c,\mathbf{k}}$, $n_{v,\mathbf{k}}$. To close the system of equations, the equations of motion for these quantities are derived in analogy to the prior derivation of Equation 3.56, and the following equations are obtained:

$$\frac{\mathrm{d}}{\mathrm{d}t} n_{c,\mathbf{k}} = -\frac{\mathrm{d}}{\mathrm{d}t} n_{v,\mathbf{k}} = -\frac{2}{\hbar}\,\mathrm{Im}\left[(\mathbf{d}_{cv}\cdot\mathbf{E}(t) + \sum_{\mathbf{q}\neq\mathbf{k}} V_{|\mathbf{k}-\mathbf{q}|} p_{vc,\mathbf{q}}) p^*_{vc,\mathbf{k}}\right]. \tag{3.58}$$

We now have a closed system of equations for $p_{vc,\mathbf{k}}$, $n_{c,\mathbf{k}}$, and $n_{v,\mathbf{k}}$, which constitute the *semiconductor Bloch equations* in Hartree–Fock approximation. For a given electromagnetic light field, these equations may be solved numerically (starting with the unpolarized system before excitation and calculating the temporal evolution).

Let us now look at the interaction between the carriers, i.e., the Coulomb effects. These are marked by the Coulomb potential V_q and occur in the renormalized energies $e_\lambda(\mathbf{k})$ and in a renormalization to the Rabi frequency,

$$\Omega_{\mathrm{Coulomb},\mathbf{k}} = \frac{\left|\mathbf{d}_{cv}\cdot\mathbf{E}(t) + \sum_{\mathbf{q}\neq\mathbf{k}} V_{|\mathbf{k}-\mathbf{q}|} p_{vc,\mathbf{q}}\right|}{\hbar}. \tag{3.59}$$

A characteristic of the Hartree–Fock approximation is thus that Coulomb effects may be represented as renormalization terms.

Neglecting the Coulomb interaction, we regain the *optical Bloch equations*:

$$\left[i\hbar\frac{\mathrm{d}}{\mathrm{d}t} - (\epsilon_c(\mathbf{k}) - \epsilon_v(\mathbf{k}))\right] p_{vc,\mathbf{k}} = [n_{c,\mathbf{k}} - n_{v,\mathbf{k}}]\mathbf{d}_{cv}\cdot\mathbf{E}(t) p_{vc,\mathbf{q}}, \tag{3.60}$$

$$\frac{\mathrm{d}}{\mathrm{d}t} n_{c,\mathbf{k}} = -\frac{\mathrm{d}}{\mathrm{d}t} n_{v,\mathbf{k}} = -\frac{2}{\hbar}\,\mathrm{Im}\left[\mathbf{d}_{cv}\cdot\mathbf{E}(t) p^*_{vc,\mathbf{k}}\right]. \tag{3.61}$$

Here, electrons at different $\mathbf{k}$-values do not couple to each other, i.e., the semiconductor is described as an ensemble of uncoupled two-level systems. Only the Coulomb interaction couples the different wave vectors and makes the semiconductor a true many-body system.

Using the optical Bloch equations and an approach similar to the derivation of Equation 3.36 in Section 3.2 and Equation 3.53, one obtains an often-used formula for the semiconductor absorption or gain [12],

$$\alpha(\hbar\omega) = -g(\hbar\omega) = \frac{\pi e^2}{nc\mathcal{E}_0 m_e^2\omega} |\hat{\mathbf{e}}\cdot\mathbf{d}_{cv}|^2 \int_0^\infty \mathrm{d}E\,\rho_r(E) \frac{\Gamma/(2\pi)}{(E_g + E - \hbar\omega)^2 + (\Gamma/2)^2} \left\{f_v(E) - f_c(E)\right\}, \tag{3.62}$$

$\rho_r(E)$ is the density of states in a three-dimensional parabolic band structure,

$$\rho_r(E) = \frac{1}{2\pi^2}\left(\frac{2m_r}{\hbar^2}\right)^{3/2} E^{1/2}, \tag{3.63}$$

which occurs when transferring from an integral over equally spaced $\mathbf{k}$-values to an energy integral [8,12]. The effective-mass approximation has been used, and m_r is the reduced mass originating from the effective masses of hole and electron. $\hat{\mathbf{e}}$ is the unit vector pointing in the direction of the light field, and m_e is the free electron mass. The carrier densities n_v, n_c have been replaced by Fermi–Dirac distributions $f_v(E)$, $f_c(E)$, assuming that scattering is fast enough to keep the carrier distributions in thermal equilibrium. This is in general a good approximation especially in a high-density (laser) system; although hole burning or other deviations from an equilibrium distribution may occur [13,14], kinetic holes in the carrier distribution are often smoothed out by the energy integral in Equation 3.62.

Equation 3.62 is thus obtained making, apart from common assumptions such as an infinite crystal lattice, the following approximations:

- Two-band approximation
- Effective-mass approximation, including the assumption of direction-independence of the effective mass
- Neglect of Coulomb interaction between the carriers; many body effects are only included as a phenomenological broadening
- Assumption of a carrier distribution in thermal equilibrium
- Dephasing time $T_2 = 1/\Gamma$ for the decline of the polarization

Looking at laser gain, the spectrum is rather unstructured and deviations between theory and experiment are often compensated by fitting the free parameters of the theory, such as carrier density and broadening; usually, the bandgap is also obtained by fits to experiment. Many gain spectra may thus be fitted with Equation 3.62. A predictive theory, however, needs to take into account many-body effects and closely look at the impact of the other simplifications mentioned above for the structure and experimental conditions it simulates and has to go far beyond the approximations made in Equation 3.62. An excellent overview of the effects of different approximations is given in Ref. [8].

The Hartree–Fock approximation that was introduced above is a widespread method to include many-body effects while keeping computation time reasonable, yet there are many other possible approaches. These include, e.g., (1) the linear case, where the carrier densities are assumed to be zero. The polarization may then be calculated in Hartree–Fock approximation. (2) Calculations beyond Hartree–Fock, where the Coulomb terms are treated in a more elaborate way. In particular, the Coulomb terms also include carrier scattering, which is not mirrored in the Hartree–Fock approximation. Therefore, Hartree–Fock equations require an artificial relaxation to be included. Instead, scattering can be included in the second Born approximation[†] by deriving the equations of motion for four-operator terms and factorizing at this level. This avoids having to include an ad hoc-scattering time, which can only be determined experimentally, and additional changes in temperature, carrier density and other experimental parameters.

Another fact that has not been discussed so far is that realistic semiconductor structures are never perfect. There has been incredible progress in the growth of semiconductor structures since Wolfgang Pauli termed solid-state physics as "dirt physics" in 1931 [15], but there still remain some inhomogeneities and disturbances of the crystal lattice. As was mentioned previously (see Section 3.2.3), disorder introduces a different kind of broadening termed inhomogeneous, where the lineshape is often assumed to be Gaussian.

Thus, different effects are manifest in the lineshape and linewidth of resonances; in other words, the lineshape says a lot about the processes determining it. To illustrate this, Figure 3.3 [16] shows a calculation of a single absorption resonance including homogeneous and inhomogeneous broadening in differing amounts. Both lineshapes include a homogeneous phonon broadening corresponding to a temperature of 30 K; for the calculation of the solid line, an inhomogeneous broadening of 0.287 meV has been added, while for the dotted line only half the disorder strength was included and an additional phenomenological homogeneous broadening of 0.066 meV was introduced via a T_2-time approach. It becomes obvious that homogeneous broadening shifts oscillator strength to the tails of the resonance. In fact, the tails of a Lorentzian lineshape become so strong that unphysical effects may appear, e.g., absorption below the gain resonance of a laser spectrum.

[†] Named after German physicist and mathematician Max Born (1882–1970).

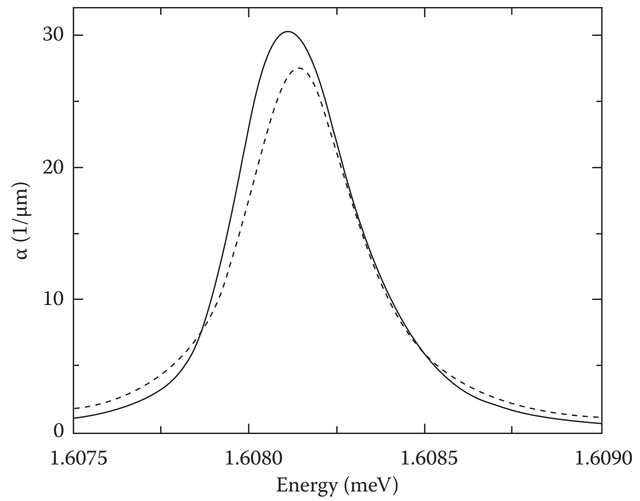


FIGURE 3.3 Absorption of a single resonance for two cases with equal total width, but differing amounts of homogeneous and inhomogeneous broadening. The solid line corresponds to phonon broadening for a temperature of 30 K and disorder broadening of 0.287 meV. For the dashed line, the disorder strength was halved and an additional phenomenological homogeneous broadening of 0.066 meV was introduced. As expected, greater homogeneous broadening results in a more symmetric line shape and stronger tails. (From A. Thränhardt et al., *Phys. Rev. B.*, 68, 035316, 2003.)

3.4 Quantization of the Light Field

Thus far, we have worked with semiclassical descriptions, i.e., including a classical light field. Yet it was shown in Section 3.2.1 that any system requires a spontaneous emission term even in a simple model, whereas in Section 3.2.2 it became apparent that the semiclassical model does not offer a term to that extent. Looking at the device aspect, we remark that the simplest optoelectronic device, the LED, emits light entirely due to spontaneous emission, and the laser diode depends on it for start-up. Quantum effects also play a role, e.g., in nanolasers and single quantum dots in resonators [17]. We thus briefly discuss quantization of the electromagnetic field. Due to the rich range of phenomena involved we leave the major discussion to the wealth of textbooks on the market, e.g., Refs. [18,19] among many others.

Consider again a two-level atom. The Hamiltonian of the fully quantized system consists of three parts,

$$\hat{H} = \hat{H}_{\text{atom}} + \hat{H}_{\text{light}} + \hat{H}_{\text{int}} , \tag{3.64}$$

where $\hat{H}_{\text{atom}}$ corresponds to the Hamiltonian of the undisturbed electron system (formerly $\hat{H}_0$), $\hat{H}_{\text{light}}$ is the Hamiltonian of the undisturbed light system, and $\hat{H}_{\text{int}}$ describes the interaction between the two.

Looking at a system in free space, we need to take into account a variety of photon modes characterized by their wave vectors $\mathbf{q}$ pointing in all directions. Spontaneous emission of a two-level atom can then be calculated. Starting from the excited two-level atom, the system wave function can be written as

$$|\Psi(t)\rangle = \hat{a}(t)e^{-i\omega_0 t}|e;0\rangle + \sum_{\mathbf{q},s} \hat{b}_{\mathbf{q},s}(t)e^{-i\omega_q t}|g;1_{\mathbf{q},s}\rangle , \tag{3.65}$$

i.e., either the system is in its initial state $|e;0\rangle$ where the atom is excited and there is no photon, or a photon with wave vector $\mathbf{q}$ and polarization s has been emitted and the atom thus returned to its ground state, yielding a wave vector $|g;1_{\mathbf{q},s}\rangle$. Here, the energy of the photon will correspond to the energy difference of the two-level system $\hbar\omega_0$, thus the absolute value of the photon wave vector $q = \omega_0/c$ is well defined, but not its direction. In 1930, Weisskopf[†] and Wigner[‡] found an irreversible exponential decay of the atom with a decay rate equal to Einstein's *A* coefficient (see Equation 3.11) [20].

We briefly discuss another case here, where there is only one photon mode at the energy $\hbar\omega_C$. Consider, e.g., a two-level atom in a cavity only allowing a specific photon mode, leading us to the Jaynes–Cummings model [21]. The undisturbed photon Hamiltonian is

$$\hat{H}_{\text{light}} = \hbar\omega_C \, \hat{a}^\dagger \hat{a} \, . \tag{3.66}$$

Here, $\hat{a}^\dagger$ ($\hat{a}$) is a photon creation (annihilation) operator. The atom Hamiltonian remains unchanged,

$$\hat{H}_{\text{atom}} = \epsilon_1 |1\rangle\langle 1| + \epsilon_2 |2\rangle\langle 2| = \hbar\omega_0 |2\rangle\langle 2|, \tag{3.67}$$

by setting $\epsilon_1 = 0$ and $\epsilon_2 = \hbar\omega_0$. We introduce the ladder operators

$$\begin{aligned} \hat{\sigma}^+ &= |2\rangle\langle 1|, \\ \hat{\sigma}^- &= |1\rangle\langle 2|, \end{aligned} \tag{3.68}$$

expressing the transfer of the electron from level 1 to level 2 and vice versa, respectively. Next, we rewrite $\hat{H}_{\text{atom}}$ as

$$\hat{H}_{\text{atom}} = \hbar\omega_0 \hat{\sigma}^+ \hat{\sigma}^- \tag{3.69}$$

and express the interaction Hamiltonian in RWA as

$$\hat{H}_{\text{int}} = \hbar g (\hat{a}^\dagger \hat{\sigma}^- + \hat{a} \hat{\sigma}^+), \tag{3.70}$$

where g is the coupling strength. The RWA expresses that, whenever a photon is created, the electron has to drop from the excited state to the ground state ($\hat{a}^\dagger \hat{\sigma}^-$) and when a photon is absorbed, the electron has to jump from the ground state into the excited state ($\hat{a}\hat{\sigma}^+$).

The Jaynes–Cummings Hamiltonian may then be solved by writing it in a matrix representation and diagonalizing this matrix. Looking at Equation 3.65, the wave functions must be linear combinations of $|2, n\rangle$ (the atom is in the excited state, state 2, and there are n photons in the system) and state $|1, n + 1\rangle$ (the atom is in the ground state, state 1, and there are $n + 1$ photons in the system). However, $|2, n\rangle$ and $|1, n + 1\rangle$ are not the eigenstates of the system any more. They are often called the bare states while the true eigenstates, which result from diagonalization of Hamiltonian (3.64), are referred to as dressed states. On resonance, they are found to be

$$\begin{aligned} |S\rangle &= \frac{|2, n\rangle + |1, n + 1\rangle}{\sqrt{2}}, \\ |A\rangle &= \frac{|2, n\rangle - |1, n + 1\rangle}{\sqrt{2}}. \end{aligned} \tag{3.71}$$

Similar to the semiclassical case, for a system initially in the upper state this results in an oscillation:

$$\begin{aligned} |c_{en}(t)|^2 &= \cos^2\left(g\sqrt{n + 1}t\right), \\ |c_{gn+1}(t)|^2 &= \sin^2\left(g\sqrt{n + 1}t\right). \end{aligned} \tag{3.72}$$

[†] Victor Frederick Weisskopf, Austrian-born American physicist (1908–2002).

[‡] Eugene Paul Wigner, Hungarian-American physicist (1902–1995).

There is, however, an important difference for the case $n = 0$, i.e., when no light field is present: While from Equation (3.20), we obtain $\Omega = 0$ for the semiclassical case (i.e., no Rabi oscillations occur), the quantum optical calculation yields the so-called vacuum Rabi oscillations with nonzero vacuum Rabi frequency. The excited atom periodically emits and reabsorbs a photon.

An interesting case also occurs if the two-level system interacts not with a single photon state $|n\rangle$, but with a coherent superposition of photon states. In this case, the probability of finding the system in the excited state will dephase due to the different Rabi frequencies of the various photon states; however, at certain times there will be a revival of the excited state because of constructive interference. This "quantum revival" was predicted theoretically in 1980 [22] and experimentally observed in 1987 [23].

3.5 Role of the Electromagnetic Environment

Thus far, we have not considered the structure of the photon wave function at all. However, for the Jaynes–Cummings model (see Section 3.4), we consider the interaction of a two-level system with a single photon mode. For the preparation of this system, a high-quality microcavity is required. As we have seen, the strength of the light field at the location of the electron determines the interaction strength between light and matter. Therefore, resonators are widely used to engineer electron–photon interaction processes, and the last section deals with the theoretical handling of these structures.

In resonator structures, calculation of the appropriate photon modes is required to allow access to the strength of electron–photon coupling. This will be clarified using the example of a microcavity. Modes can be established by a transfer-matrix method, which will be briefly discussed. For simplicity, we only look at a planar structure and perpendicular light incidence in this book chapter, while for in-depth information, the reader is referred to Refs. [24,25]. Information on the transfer matrix method may also be found in Ref. [26] and the references therein.

In free space or homogeneous matter, plane waves $\mathbf{E}_0 e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)}$ are solutions to Maxwell's equations [4]. Any light field may be expanded in plane waves. The transfer matrix method uses this to express the light field as a plane wave in every layer of a microstructure and appropriately match the conditions at the interfaces.

We consider a planar structure where the growth direction is taken as the z-axis. The index of refraction then only depends on z. The homogeneous wave equation reads

$$\left\{ \frac{1}{c^2}\frac{\partial}{\partial t^2} - \Delta \right\} \mathbf{u}_\mathbf{q}(z) = 0\,, \tag{3.73}$$

where $\mathbf{u}_\mathbf{q}(z)$ is the light mode we are looking for, indexed with its wave vector $\mathbf{q}$. Since we are looking at perpendicular incidence in a planar structure, the wave vector points in the z-direction. Due to the geometry of the problem, the two independent polarizations are equivalent to each other. The index of polarization is thus omitted here, and bearing in mind that we need to find two independent modes for each wave vector q. The polarization becomes important for incidence at an angle.

Consider first a single boundary between two layers L (left) and R (right). Generally, each mode $\mathbf{u}_\mathbf{q}(z)$ is a combination of plane waves traveling in forward and backward directions in each layer,

$$\begin{gathered} \mathbf{u}_\mathbf{q}(z) = A_j e^{iq_{zj}(z-z_j)}\mathbf{e}_j + B_j e^{-iq_{zj}(z-z_j)}\mathbf{e}'_j \\ \text{where} \quad q_{zj} = n(z_j)\,q\,. \end{gathered} \tag{3.74}$$

Maxwell's equations demand continuity for the tangential components of the electric and magnetic field and thus (in this geometry) for the entire mode $\mathbf{u}_\mathbf{q}(z)$ and its derivation $\mathrm{d}/\mathrm{d}z\mathbf{u}_\mathbf{q}(z)$. From this condition, transmission and reflection may be derived as

$$T_j^- = \frac{2q_{z,j}}{q_{z,j} + q_{z,j+1}}, \tag{3.75}$$

$$R_j^- = \frac{q_{z,j} - q_{z,j+1}}{q_{z,j} + q_{z,j+1}}, \tag{3.76}$$

$$T_j^+ = \frac{2q_{z,j+1}}{q_{z,j} + q_{z,j+1}}, \tag{3.77}$$

$$R_j^+ = \frac{q_{z,j+1} - q_{z,j}}{q_{z,j} + q_{z,j+1}}; \tag{3.78}$$

see Ref. [4] for more details. T_j^- and R_j^- (T_j^+ and R_j^+) are transmission and reflection coefficients for light incident from the left (right) side, respectively. This may be generalized to a matrix representation, relating neighboring heterostructure layers by

$$\begin{pmatrix} A_{j+1} \\ B_{j+1} \end{pmatrix} = \frac{1}{T_j^+} \begin{pmatrix} e^{iq_{z,j}d_j} & R_j^+ e^{-iq_{z,j}d_j} \\ -R_j^- e^{iq_{z,j}d_j} & e^{-iq_{z,j}d_j} \end{pmatrix} \begin{pmatrix} A_j \\ B_j \end{pmatrix} \equiv \mathbf{M}_j \begin{pmatrix} A_j \\ B_j \end{pmatrix}, \tag{3.79}$$

where d_j is the width of layer j, $d_j = z_j - z_{j+1}$, and $\mathbf{M}_j$ is the transfer matrix. Recursive application yields

$$\begin{pmatrix} A_{N+1} \\ B_{N+1} \end{pmatrix} = \mathbf{M}_N \mathbf{M}_{N-1} \dots \mathbf{M}_2 \mathbf{M}_1 \begin{pmatrix} A_1 \\ B_1 \end{pmatrix} \equiv \mathbf{M}_{tot} \begin{pmatrix} A_1 \\ B_1 \end{pmatrix} \tag{3.80}$$

and allows the calculation of modes $\mathbf{u}_\mathbf{q}^+(z)$ and $\mathbf{u}_\mathbf{q}^-(z)$ incident from the left and right of the microcavity. Figure 3.4 [27]. shows a resonant mode in a typical Galliumarsenide (GaAs)/Aluminiumarsenide (AlAs) microcavity structure with 10 layer pairs on top and 20 layer pairs on the substrate. It becomes obvious that

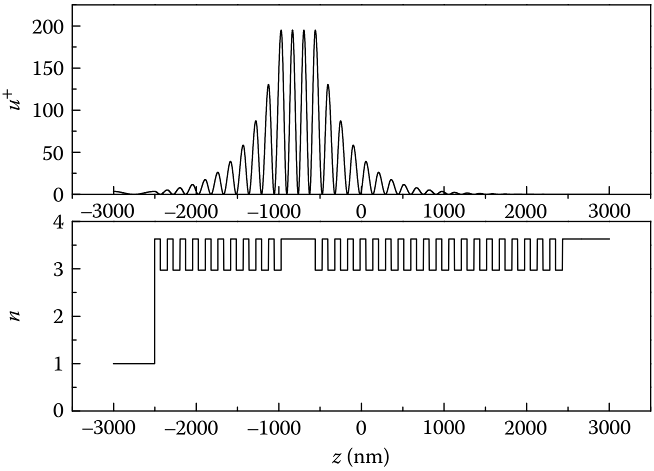


FIGURE 3.4 Top: Central mode of a GaAs/AlAs microcavity with 10 (on air) and 20 (on substrate) $\lambda/4$ layer pairs of GaAs and AlAs with background refractive indices of $n_{\text{GaAs}} = 3.63$ and $n_{\text{AlAs}} = 2.97$, respectively. Bottom: Refractive index of the microcavity along the growth direction z. (From A. Thränhardt, Equilibrium and Nonequilibrium Dynamics in Semiconductor Lasers, Habilitation Thesis, Marburg, 2006.)

the electric light field is enhanced by a factor of about 200 even in this simple structure in the peaks in the cavity. In this way, light–matter interaction may be reinforced by several orders of magnitude. Of course, photonic engineering also works for two- or three-dimensional photonic crystal structures where more elaborate theoretical approaches have to be used for the calculation of the modes, e.g., the finite difference time domain (FDTD) method [28,29], where electric and magnetic field components are calculated alternately.

References

1. P. A. M. Dirac, The quantum theory of the emission and absorption of radiation, *Proc. R. Soc. A* **114**, 243 (1927).
2. L. Allen, J. H. Eberly, *Optical Resonance and Two-Level Atoms, Dover Books on Physics*, New York, (1988) (revised edition).
3. A. Einstein, Zur Quantentheorie der Strahlung, *Phys. Zeitschrift* **18**, 121 (1917), first published in *Mitteilungen der Phys. Ges. Zürich* 18 (1916).
4. J. D. Jackson, *Classical Electrodynamics*, 3rd ed., New York: Wiley (1999).
5. J. Mompart, R. Corbalán, Lasing without inversion, *J. Opt. B: Quantum Semiclass. Opt*. **2**, R7 (2000).
6. M. D. Frogley, J. F. Dynes, M. Beck, J. Faist, C. C. Phillips, Gain without inversion in semiconductor nanostructures, *Nat. Mater.* **5**, 175 (2006).
7. H. Haug, S. W. Koch, *Quantum Theory of the Optical and Electronic Properties of Semiconductors*, 5th ed., Singapore: World Scientific (2009).
8. W. W. Chow, S. W. Koch, *Semiconductor–Laser Fundamentals*, Berlin/Heidelberg: Springer (1999).
9. S. Adachi, *GaAs and Related Materials: Bulk Semiconducting and Superlattice Properties*, Singapore: World Scientific (1994).
10. G. Franco Bassani, G. Pastori Parravicini, *Electronic States and Optical Transitions in Solids*, New York, NY: Pergamon Press (1975).
11. D. Dragoman, M. Dragoman, *Optical Characterization of Solids*, Berlin/Heidelberg: Springer (2010).
12. Shun Lien Chuang, *Physics of Optoelectronic Devices*, New York/Brisbane/Singapore: Wiley-InterScience (1995).
13. A. Thränhardt, S. Becker, C. Schlichenmaier, I. Kuznetsova, T. Meier, S. W. Koch, J. Hader, J. V. Moloney, W. W. Chow, Nonequilibrium gain in optically pumped GaInNAs laser structures, *Appl. Phys. Lett.* **85**, 5526 (2004).
14. E. Kühn, S. W. Koch, A. Thränhardt, J. Hader, J. V. Moloney, Microscopic simulation of nonequilibrium features in quantum-well pumped semiconductor disk lasers, *Appl. Phys. Lett.* **96**, 051116 (2010).
15. K. V. Meyenn (Ed.), *Wolfgang Pauli: Wissenschaftlicher Briefwechsel mit Bohr, Einstein, Heisenberg u.a. Band II: 1930–1939/Scientific Correspondence with Bohr, Einstein, Heisenberg a.o.*, Berlin/Heidelberg/New York, NY/Tokyo: Springer, Reprint (2014).
16. A. Thränhardt, C. Ell, S. Mosor, G. Rupper, G. Khitrova, H. M. Gibbs, S. W. Koch, Interplay of phonon and disorder scattering in semiconductor quantum wells, *Phys. Rev. B* **68**, 035316, 2003.
17. F. Jahnke (ed.), Q*uantum Optics with Semiconductor Nanostructures*, Oxford/Cambridge/Philadelphia/New Delhi: Woodhead Publishing Series in Electronic and Optical Materials (2012).
18. P. Meystre, M. Sargent III, *Elements of Quantum Optics*, 4th ed., Berlin/Heidelberg: Springer (2010).
19. M. Kira, S. W. Koch, *Semiconductor Quantum Optics*, Cambridge: Cambridge University Press (2011).
20. V. Weisskopf, E. Wigner, Berechnung der natürlichen Linienbreite auf Grund der Diracschen Lichttheorie, *Zeitschrift für Physik* **63**, 54 (1930).
21. E. T. Jaynes, F. W. Cummings, Comparison of quantum and semiclassical radiation theories with application to the beam maser, *Proc. IEEE* **51**, 89 (1963).

22. J. H. Eberly, N. B. Narozhny, J. J. Sanchez-Mondragon, Periodic spontaneous collapse and revival in a simple quantum model, *Phys. Rev. Lett.* **44**, 1323 (1980).
23. G. Rempe, H. Walther, N. Klein, Observation of quantum collapse and revival in a one-atom MASER, *Phys. Rev. Lett.* **58**, 353 (1987).
24. M. Born, E. Wolf, *Principles of Optics: Electromagnetic Theory of Propagation, Interference and Diffraction of Light*, 7th ed., Cambridge: Cambridge University Press (2002).
25. B. E. Sernelius, *Surface Modes in Physics*, Berlin: Wiley-VCH (2001).
26. G. Khitrova, H. M. Gibbs, F. Jahnke, M. Kira, S. W. Koch, Nonlinear optics of normal-mode-coupling semiconductor microcavities, *Rev. Mod. Phys.* **71** (5), 1591 (1999).
27. A. Thränhardt, Equilibrium and Nonequilibrium Dynamics in Semiconductor Lasers, Habilitation Thesis, Marburg (2006).
28. K. Yee, Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media, *IEEE Trans. Antennas. Propag.* **14**, 302 (1966).
29. D. M. Sullivan, *Electromagnetic Simulation Using the FDTD Method*, Hoboken/New Jersey: Wiley, 2013.

4

Optical Waveguiding

4.1 Introduction to Optical Waveguide Theory 103
4.2 Modes of Optical Waveguides 108
Slab Optical Waveguide • Circular Optical Waveguide • Index Guiding • Effective Index Method • Radiation Modes, Leaky Waves, Gain Guiding, and Supermodes
4.3 Beam Propagation Method 117
4.4 Tapered and Bent Waveguides 121
Tapered Waveguides • Bent Waveguides

Slawomir Sujecki

Optical waveguides are essential elements of many optoelectronic devices. Their primary role is shaping, confining, and guiding of an optical beam along a selected direction. The design and modeling of optical waveguides is based on the application of the macroscopic Maxwell's equations. This approach consists of approximating the microscopic electromagnetic fields that vary on the atomic scale with fields that are averaged over several atomic layers. Considering that an atom's diameter is up to several tenths of a nanometer while the optical waveguides contain structural details with dimensions smaller than 10 nm (e.g., quantum wells of a laser diode), the application of the macroscopic Maxwell equations to the design and modeling of optical waveguides is at its limit. This limitation should be borne in mind throughout this chapter.

Here, we provide an introduction to the theory of optical waveguides. In Section 4.1, we derive the wave equations and introduce a hierarchy of approximations relevant to optical waveguide modeling and design. In Section 4.2, we discuss the modal characteristics of infinitely long shift invariant optical waveguides. Since optical waveguides used in optoelectronic devices are not infinitely long and may not be straight, in Section 4.3, we outline the theory of beam propagation method, which is the standard numerical approach used for the analysis of wave propagation in optical waveguiding structures. Finally, we discuss the optical properties of bent and tapered waveguides.

4.1 Introduction to Optical Waveguide Theory

For isotropic, linear, dielectric media, the relations between the vectors of the electric and magnetic field and the electric and magnetic flux are given by Maxwell's equations

$$\nabla \times \vec{E}\left(\vec{r}, t\right) = -\frac{\partial \vec{B}\left(\vec{r}, t\right)}{\partial t} \tag{4.1}$$

$$\nabla \times \vec{H}\left(\vec{r}, t\right) = \frac{\partial \vec{D}\left(\vec{r}, t\right)}{\partial t}. \tag{4.2}$$

In the frequency domain, Equation 4.1 has the following form:

$$\nabla \times \vec{E}\left(\vec{r}\right) = -j\omega\vec{B}\left(\vec{r}\right) \tag{4.3}$$

$$\nabla \times \vec{H}\left(\vec{r}\right) = j\omega\vec{D}\left(\vec{r}\right), \tag{4.4}$$

where the angular frequency $\omega = 2\pi$ ft. Equations 4.3 and 4.4 imply that that the unknown fields depend on time according to $e^{j\omega t}$. The real field distributions of the electric and magnetic field and electric and magnetic flux $\vec{E}\left(\vec{r},t\right), \vec{H}\left(\vec{r},t\right), \vec{D}\left(\vec{r},t\right), \vec{B}\left(\vec{r},t\right)$ are obtained from the complex distributions $\vec{E}\left(\vec{r}\right), \vec{H}\left(\vec{r}\right), \vec{D}\left(\vec{r}\right), \vec{B}\left(\vec{r}\right)$ by multiplying a particular complex field distribution with $e^{j\omega t}$ term and extracting the real part of the product.

Equations 4.3 and 4.4 are complemented by the (Gauss's law) conditions

$$\nabla \cdot \vec{D}\left(\vec{r}\right) = 0 \tag{4.5}$$

$$\nabla \cdot \vec{B}\left(\vec{r}\right) = 0. \tag{4.6}$$

In addition, the dependence between the electric field and the electric flux vectors is given by

$$\vec{D}\left(\vec{r}\right) = \varepsilon\vec{E}\left(\vec{r}\right) = \varepsilon_0\varepsilon_r\vec{E}\left(\vec{r}\right), \tag{4.7}$$

where ε_0 is the electric permittivity of the free space and ε_r is the relative permittivity of the medium while the magnetic field and the magnetic flux vectors are given by the following equation:

$$\vec{B}\left(\vec{r}\right) = \mu_0\vec{H}\left(\vec{r}\right), \tag{4.8}$$

where μ_0 is the magnetic permeability of the free space. From the theoretical point of view, it is convenient to assume that optical waveguides are infinitely long and perfectly straight. Such fictitious optical waveguide structure is longitudinally shift invariant and hence the transverse and longitudinal space variables can be separated. In order to derive the wave equations, we introduce a rectangular coordinate system x, y, z and align the longitudinal waveguide dimension with the z-axis (Figure 4.1). Under such assumptions, ε does not depend on z and the spatial distribution of the unknown fields can be expressed as a product of an unknown function depending on the transverse coordinates only and a phase factor $\exp(-j\beta z)$,

$$\vec{\Phi}\left(\vec{r}\right) = \Phi\left(x,y\right)\exp(-j\beta z), \tag{4.9}$$

where $\vec{\Phi}\left(\vec{r}\right)$ stands for $\vec{E}\left(\vec{r}\right), \vec{H}\left(\vec{r}\right), \vec{D}\left(\vec{r}\right)$, or $\vec{B}\left(\vec{r}\right)$. The constant β is called the modal propagation constant while the functions $E(x,y)$, $H(x,y)$, $D(x,y)$, and $B(x,y)$ provide the field distributions for a given waveguide mode.

Substituting Equation 4.9 into Equations 4.3 through 4.8 while observing that ε does not depend on z, we obtain the following equations for the transverse components of the electric field vector

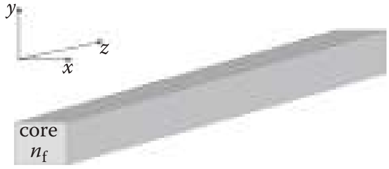


FIGURE 4.1 Schematic diagram of an optical waveguide in a rectangular coordinate system x, y, z.

$E\left(x,y\right)=\vec{i}_xE_x\left(x,y\right)+\vec{i}_yE_y\left(x,y\right)+\vec{i}_zE_z\left(x,y\right),$

$$\frac{\partial^2 E_x}{\partial y^2}+\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial\left(\varepsilon E_x\right)}{\partial x}\right)+\left(\omega^2\mu\varepsilon-\beta^2\right)E_x+\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial\left(\varepsilon E_y\right)}{\partial y}\right)-\frac{\partial^2 E_y}{\partial x\partial y}=0 \tag{4.10}$$

$$\frac{\partial^2 E_y}{\partial x^2}+\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial\left(\varepsilon E_y\right)}{\partial y}\right)+\left(\omega^2\mu\varepsilon-\beta^2\right)E_y+\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial\left(\varepsilon E_x\right)}{\partial x}\right)-\frac{\partial^2 E_x}{\partial y\partial x}=0, \tag{4.11}$$

where $\vec{\mathrm{E}}\left(\vec{r}\right)=E\left(x,y\right)\exp\left(-j\beta z\right)$. Similarly, for the magnetic field vector $H\left(x,y\right)=\vec{i}_xH_x\left(x,y\right)+\vec{i}_yH_y\left(x,y\right)+\vec{i}_zH_z\left(x,y\right)$ we obtain

$$\frac{\partial^2 H_x}{\partial x^2}+\varepsilon\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial H_x}{\partial y}\right)+\left(\omega^2\mu\varepsilon-\beta^2\right)H_x-\varepsilon\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial H_y}{\partial x}\right)+\frac{\partial^2 H_y}{\partial y\partial x}=0 \tag{4.12}$$

$$\frac{\partial^2 H_y}{\partial y^2}+\varepsilon\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial H_y}{\partial x}\right)+\left(\omega^2\mu\varepsilon-\beta^2\right)H_y-\varepsilon\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial H_x}{\partial y}\right)+\frac{\partial^2 H_x}{\partial x\partial y}=0. \tag{4.13}$$

Considering that the refractive index n is defined as the ratio of the speed of light in the medium to the speed of light in the free space, we have the following relationship between the refractive index and the dielectric permittivity: $n=\sqrt{\varepsilon/\varepsilon_0}$.

We note that the equations for the transverse vector components of the electric field, Equations 4.10 and 4.11, do not contain any coupling terms with the longitudinal ones. Furthermore, the equations for both transverse components are coupled but there is no coupling between the electric and magnetic fields. Hence, Equations 4.10 and 4.11 can be solved autonomously for E_x and E_y while the remaining electromagnetic field components can be calculated from

$$E_z=-\frac{1}{j\beta\varepsilon}\left(\frac{\partial\left(\varepsilon E_x\right)}{\partial x}+\frac{\partial\left(\varepsilon E_y\right)}{\partial y}\right) \tag{4.14}$$

$$H_x=-\frac{1}{j\omega\mu}\left(\frac{\partial E_z}{\partial y}+j\beta E_y\right) \tag{4.15}$$

$$H_y=\frac{1}{j\omega\mu}\left(\frac{\partial E_z}{\partial x}+j\beta E_x\right) \tag{4.16}$$

$$H_z=-\frac{1}{j\omega\mu}\left(\frac{\partial E_y}{\partial x}-\frac{\partial E_x}{\partial y}\right). \tag{4.17}$$

Alternatively, Equations 4.12 and 4.13 can be solved to obtain the distributions of H_x and H_y while the other field components distributions can be obtained from

$$H_z=-\frac{1}{j\beta}\left(\frac{\partial H_x}{\partial x}+\frac{\partial H_y}{\partial y}\right) \tag{4.18}$$

$$E_x=\frac{1}{j\omega\varepsilon}\left(\frac{\partial H_z}{\partial y}+j\beta H_y\right) \tag{4.19}$$

$$E_y = -\frac{1}{j\omega\varepsilon}\left(\frac{\partial H_z}{\partial x} + j\beta H_x\right) \tag{4.20}$$

$$E_z = \frac{1}{j\omega\varepsilon}\left(\frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y}\right). \tag{4.21}$$

Equations 4.10 and 4.11 form an eigenvalue problem, whereby β^2 is the eigenvalue and E_x and E_y form the corresponding eigenvector. Analogous observation applies to Equations 4.12 and 4.13. In the context of the optical waveguide analysis, β is referred to as a modal propagation constant while the corresponding eigenvector provides the modal field distribution.

The solution of the eigenvalue problem formed by either Equations 4.10 and 4.11 or Equations 4.12 and 4.13 allows obtaining the exact solution of the Maxwell equations under the assumption of the longitudinal shift invariance of the optical waveguide structure. In the literature relating to the optical waveguide modeling, such a solution of the Maxwell equations is referred to as a vectorial solution. Similarly, computer software–based tools for the design of optical waveguides that solve either Equations 4.10 and 4.11 or Equations 4.12 and 4.13 are referred to as vectorial mode solvers.

For completeness, we note that there are other ways of formulating the solution of Maxwell equations for an optical waveguide (Sujecki, 2014). However, the relevance of Equations 4.10 through 4.13 stems from the fact that most computer software tools for optical waveguide design rely on the formulation of the problem in terms of either the transverse magnetic or electric field components. A particular advantage of such approach is the elimination of spurious solutions. Magnetic field distribution is smoother than that of the electric field. Hence, software–based on the magnetic field formulation tends to have better numerical convergence properties, whereas the electrical field formulation is preferred for nonlinear optics applications.

Equations 4.10 through 4.13 are relatively difficult to solve numerically. However, they can be significantly simplified for waveguides, which support modes that are nearly linearly polarized along either the x- or y-direction. For such waveguides two types of modes are guided: modes that are polarized either along the x-direction (x-polarized or quasi-transverse electric [TE] modes) and modes that are polarized along the y-direction (y-polarized or quasi-transverse magnetic [TM] modes). The main transverse components of the electromagnetic field for the x-polarized modes are E_x and H_y, while for the y-polarized ones they are E_y and H_x. The polarized approximation has proven particularly successful in the modeling of optical waveguides realized using planar technology and is therefore of great practical importance for semiconductor optoelectronic devices (Sujecki et al., 1998). The modal propagation constants and the corresponding field distributions for the x-polarized modes are obtained by solving Equation 4.10 or 4.13 while neglecting the minor field component:

$$\frac{\partial^2 E_x}{\partial y^2} + \frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial(\varepsilon E_x)}{\partial x}\right) + \left(\omega^2\mu\varepsilon - \beta^2\right)E_x = 0 \tag{4.22}$$

$$\frac{\partial^2 H_y}{\partial y^2} + \varepsilon\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial H_y}{\partial x}\right) + \left(\omega^2\mu\varepsilon - \beta^2\right)H_y = 0. \tag{4.23}$$

Similarly, the propagation constants and field distributions of y-polarized modes are obtained from Equations 4.11 and 4.13 by leaving only the terms containing the major field component:

$$\frac{\partial^2 E_y}{\partial x^2} + \frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial\left(\varepsilon E_y\right)}{\partial y}\right) + \left(\omega^2\mu\varepsilon - \beta^2\right)E_y = 0 \tag{4.24}$$

$$\frac{\partial^2 H_x}{\partial x^2} + \varepsilon\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial H_x}{\partial y}\right) + \left(\omega^2\mu\varepsilon - \beta^2\right)H_x = 0. \tag{4.25}$$

The solution of the partial differential equations for the major transverse field component is only much easier to implement and reduces the required computational effort. Further, improvement of computational efficiency can be achieved if ε varies slowly in the transverse plane. In such cases, the following scalar approximation can be used:

$$\left(\frac{\partial^2}{\partial x^2}+\frac{\partial^2}{\partial y^2}+\omega^2\mu\varepsilon\right)\varphi=\beta^2\varphi. \tag{4.26}$$

In principle, the scalar approximation is applicable only to waveguides with a small refractive index contrast. However, there are known instances of waveguides with a large refractive index contrasts for which plausible accuracy may be achieved under the scalar approximation (Riishede et al., 2003).

In summary, by deriving Equations 4.10 through 4.13, Equations 4.22 through 4.25 and 4.26, a hierarchy of approximations is established and this is applicable to optical waveguide analysis. Thus, depending upon which equation is solved by a particular solution method, the methods of calculating the modal propagation constants and the corresponding field distributions are divided into three categories. The vectorial methods attempt to solve the vectorial wave Equations 4.10 through 4.13. Methods that attempt to solve the Equations 4.22 through 4.25 are referred to as either semivectorial or using the polarized approximation. The scalar methods solve the scalar wave Equation 4.26. Compared to vectorial methods, scalar methods are more efficient but less accurate. The semivectorial methods (polarized approximation) provide a compromise between efficiency and accuracy when compared with the other two alternatives (Figure 4.2).

Finally, in order to illustrate the differences between the vectorial, semivectorial, and scalar methods, Table 4.1 presents normalized propagation constant values, which have been obtained for a rib waveguide structure with $H = 1$ μm, $W = 1.5$ μm, $n_f = 3.44$, $n_s = 3.4$, $n_c = 1.0$ (Figure 4.6). The outer slab thickness H is varied, between 0.1 and 0.9 μm. Operating wavelength is 1.15μm. The scalar and polarized results have been obtained using the finite difference method (FDM) (Sujecki et al., 1998), while the vectorial ones were taken from Hadley (2002) and Vassallo (1997b) for x- and y-polarized modes, respectively. All the results are believed to be calculated with an absolute error of 5×10^{-5}. The difference between the

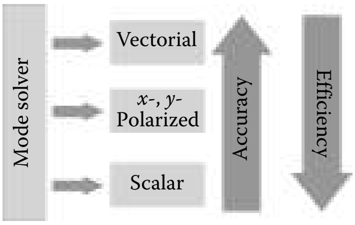


FIGURE 4.2 Diagram showing the hierarchy of approximations used in optical waveguide modeling.

TABLE 4.1 Normalized Propagation Constants B of Rib Dielectric Waveguide

h (μm)	Scalar Finite Difference	Polarized	Vectorial (Hadley, 2002)	Polarized	Vectorial (Vassallo, 1997b)
	Scalar	x-Polarized		y-Polarized	
0.1	0.3095	0.3018	0.3019	0.2674	0.2674
0.3	0.3176	0.3108	0.3110	0.2750	0.2751
0.5	0.3320	0.3267	0.3270	0.2888	0.2890
0.7	0.3542	0.3509	0.3512	0.3104	0.3107
0.9	0.3893	0.3884	0.3886	0.3453	0.3455

$H + h = 1$ μm; $W = 1.5$ μm; $n_f = 3.44$; $n_s = 3.4$; $n_c = 1.0$; $\lambda = 1.15$ μm (cf. Figure 4.6b).

polarized and vectorial analysis results does not exceed 3 on the fourth decimal place for either polarization. This illustrates the advantages of the polarized approximation in the case of rib waveguides. The results obtained by the scalar method, however, differ up to eight on the third decimal place for x-polarized modes and even in the first decimal place for the y-polarized modes. It should be noted that for $h = 0.9$ μm, the scalar approximation is accurate nearly on three decimal places for the x-polarized mode, despite a large refractive index contrast.

A survey of numerical methods used in calculations of propagation constants and field distributions of optical waveguides can be found in Sujecki (2014).

4.2 Modes of Optical Waveguides

In this section, we discuss modal characteristics of the selected longitudinally shift-invariant optical waveguides that are relevant to optoelectronic devices. We start with slab and circular waveguides, for which analytical solutions of Maxwell equations are available. Then, we consider selected index guiding optical waveguides for which Maxwell equations must be solved numerically. We then discuss the effective index method and finally provide a brief description of the concepts of radiation, leaky, and gain guided modes and also of waveguide array supermodes.

4.2.1 Slab Optical Waveguide

The refractive index distribution of a slab waveguide varies only along one selected axis of a rectangular coordinate system (cf. Figure 4.3). In any plane perpendicular to that axis, the refractive index distribution is homogenous, while the refractive index of the core n_f is larger than that of the substrate n_s and the cladding n_c and $n_s > n_c$. When a slab waveguide has a piecewise homogenous refractive index distribution (Figure 4.3), useful insight into the nature of the modes can be gained using the geometric optics. Figure 4.4 shows a cross section of a slab waveguide with an optical ray pattern, whereby γ is larger than the critical value: $\gamma = \sin^{-1}(n_s/n_f)$. The optical ray is thus totally reflected from the interfaces between the core and cladding and core and substrate. Since both interfaces are parallel, the subsequent incidence angles γ do

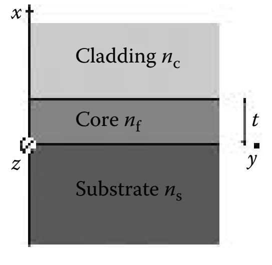


FIGURE 4.3 Three-layer slab waveguide in a rectangular coordinate system x, y, z.

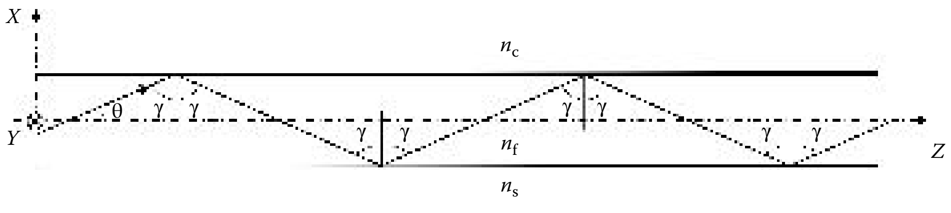


FIGURE 4.4 An optical ray pattern of a mode guided by a three-layer slab waveguide.

not change under subsequent reflections. Further, it can be observed that if the value of the incidence angle is larger than the critical angle, the optical ray is trapped within the core; and hence there is no power leakage. Furthermore, associating plane waves with optical rays leads to a realization that the waves generated through subsequent reflections interfere within the core. Thus, in order to avoid destructive interference the phase change for an optical wave upon the complete roundtrip between the core–cladding and core–substrate interfaces has to be equal to a multiple of 2π. This leads to a conclusion that the spectrum of the guided modes is discrete. The last intuitive observation that can be made with the help of ray optics is that if the plane wave associated with the ray is polarized along the y-axis, then it will not change the state of polarization upon subsequent ray reflections. This leads to a conclusion that the waveguide preserves linear polarization. All these observations are confirmed by the rigorous solution of the Maxwell equations (Sujecki, 2014).

In order to obtain the field distributions and propagation constants of the slab waveguide in the TE mode case, one needs to consider Equation 4.11. Since the derivatives along the y-direction are equal to zero, all coupling terms with E_x component can be neglected. For the three-layer waveguide (Figure 4.3), the field distribution of E_y component of the electromagnetic field therefore has the form

$$E_y = \begin{cases} C\exp\left(-p\left(x-t\right)\right) & \text{for} \quad x > t \\ A\cos\left(ht\right) + B\sin\left(ht\right) & \text{for} \quad 0 < x < t \\ D\exp\left(qx\right) & \text{for} \quad x < 0 \end{cases}, \tag{4.27}$$

where $p^2 = \beta^2 - n_c^2 k_0^2$, $q^2 = \beta^2 - n_s^2 k_0^2$, and $h^2 = n_f^2 k_0^2 - \beta^2$.

Substituting Equation 4.27 into Equation 4.10 and imposing the field continuity for E_y and H_z yields the dispersion equation for TE modes

$$ht = \operatorname{arctg}\frac{q}{h} + \operatorname{arctg}\frac{p}{h} + m\pi. \tag{4.28}$$

Following a similar procedure for TM modes we obtain

$$ht = \operatorname{arctg}\left(\frac{q}{h}\frac{n_f^2}{n_s^2}\right) + \operatorname{arctg}\left(\frac{p}{h}\frac{n_f^2}{n_c^2}\right) + m\pi, \tag{4.29}$$

whereby m is a parameter that sets the mode order.

Introducing the normalized frequency V, relative propagation constant B, and an asymmetry parameter A_E,

$$V = k_0 t\sqrt{n_f^2 - n_s^2}; \quad B = \frac{\beta^2 - n_s^2 k_0^2}{n_f^2 k_0^2 - n_s^2 k_0^2}; \quad A_E = \frac{n_s^2 - n_c^2}{n_f^2 - n_s^2}$$

allows recasting Equation 4.27 further into a more convenient normalized form

$$V\sqrt{1-B} - \operatorname{arctg}\sqrt{\frac{B + A_E}{1-B}} - \operatorname{arctg}\sqrt{\frac{B}{1-B}} - m\pi = 0 \tag{4.30}$$

for the TE modes and

$$V\sqrt{1-B} - \operatorname{arctg}\left(\frac{n_f^2}{n_c^2}\sqrt{\frac{B + A_E}{1-B}}\right) - \operatorname{arctg}\left(\frac{n_f^2}{n_s^2}\sqrt{\frac{B}{1-B}}\right) - m\pi = 0 \tag{4.31}$$

for the TM modes.

From Equations 4.28 through 4.31, one can obtain the values of the propagation constant for a guide mode of a slab waveguide. Once the propagation constant is calculated in the TE case, the field distribution can be obtained from Equation 4.27 while the constants A, B, C, D are calculated by solving

$$\begin{bmatrix} 1 & 0 & 0 & -1 \\ \cos(ht) & \sin(ht) & -1 & 0 \\ 0 & h & 0 & -q \\ -h\sin(ht) & h\cos(ht) & p & 0 \end{bmatrix} \begin{bmatrix} \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \\ \mathrm{D} \end{bmatrix} = 0. \tag{4.32}$$

For the TM case, the H_y component of the electromagnetic field is obtained using the formulae given by Equation 4.27, whereby the constants A, B, C, D are calculated by solving

$$\begin{bmatrix} 1 & 0 & 0 & -1 \\ \cos(ht) & \sin(ht) & -1 & 0 \\ 0 & \frac{h}{n_\mathrm{f}^2} & 0 & -\frac{q}{n_\mathrm{s}^2} \\ -\frac{h}{n_\mathrm{f}^2}\sin(ht) & \frac{h}{n_\mathrm{f}^2}\cos(ht) & \frac{p}{n_\mathrm{c}^2} & 0 \end{bmatrix} \begin{bmatrix} \mathrm{A} \\ \mathrm{B} \\ \mathrm{C} \\ \mathrm{D} \end{bmatrix} = 0. \tag{4.33}$$

In summary, the rigorous solution of the Maxwell equations shows that a slab waveguide supports two types of linearly polarized modes: TE and TM. The cutoff frequency of the first-order mode is obtained by setting $B = 0$ in Equation 4.30 for the TE case: $V = arctg\sqrt{A_\mathrm{E}} + m\pi$ and in Equation 4.31 for the TM case: $V = \mathrm{arctg}\left(\frac{n_\mathrm{f}^2}{n_\mathrm{c}^2}\sqrt{A_\mathrm{E}}\right) + \pi$.

An analytical solution can also be also obtained in the case of a multilayered slab waveguide, if the refractive index distribution is homogenous within each layer of an isotropic (Anemogiannis and Glytsis, 1992) or a uniaxial anisotropic medium (Majewski, 1984) and also for several cases of optical waveguides with a graded index profile (Tamir, 1975). Alternatively, for multilayer slab waveguides, numerical methods can be used (Anemogiannis et al., 1994; Ding and Chan, 1997; Anemogiannis et al., 1999; Sujecki, 2010).

4.2.2 Circular Optical Waveguide

Figure 4.5 shows the cross section of a step-index circular waveguide. For this waveguide structure, the core of the refractive index n_c and radius a is surrounded by an infinitely extending cladding with refractive index n_a, whereby $n_\mathrm{c} > n_\mathrm{a}$. For a step-index circular waveguide, Maxwell's equations can be solved analytically (Hondros and Debye, 1910).

In order to calculate the propagation constants of a circular optical waveguide, the Hondros–Debye equation has to be solved numerically

$$\left(X_\mathrm{m} + Y_\mathrm{m}\right)\left(n_\mathrm{c}^2 Y_\mathrm{m} + n_\mathrm{a}^2 X_\mathrm{m}\right) - \frac{m^2N^2}{\left(u^2B\right)^2} = 0, \tag{4.34}$$

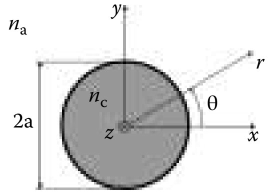


FIGURE 4.5 Step-index circular optical fiber.

where $X_m = \frac{K'_m(w)}{wK_m(w)}$, $Y_m = \frac{J'_m(u)}{uJ_m(u)}$, $B = \frac{N^2 - n_a^2}{n_c^2 - n_a^2}$ is the relative propagation constant, $N = \beta/k$ is the effective refractive index, k is the wave number, $u = \kappa a$ and $w = \gamma a$, $\kappa^2 = \omega^2 n_c^2 \varepsilon_0 \mu_0 - \beta^2$, $\gamma^2 = \beta^2 - \omega^2 n_a^2 \varepsilon_0 \mu_0$, and n_c and n_a are the refractive indices of the core and cladding, respectively. J_m and K_m denote Bessel functions of the first kind and modified Bessel functions of the second kind, while m is the function order. For $m = 0$, Equation 4.34 splits into two equations, the solutions of which give the propagation constants of the TE modes

$$\frac{J_1(u)}{uJ_0(u)} + \frac{K_1(w)}{wK_0(w)} = 0 \tag{4.35}$$

and TM modes

$$n_c^2 \frac{J_1(u)}{uJ_0(u)} + n_a^2 \frac{K_1(w)}{wK_0(w)} = 0. \tag{4.36}$$

Once the propagation constant of a mode is calculated, the longitudinal field component distribution can be obtained from

$$E_z(r, \theta, z) = A_E J_m(\kappa r)\, e^{j(-\beta z + m\theta)} \tag{4.37}$$

$$H_z(r, \theta, z) = A_H J_m(\kappa r)\, e^{j(-\beta z + m\theta)} \tag{4.38}$$

in the core and the cladding

$$E_z(r, \theta, z) = B_E K_m(\gamma r)\, e^{j(-\beta z + m\theta)} \tag{4.39}$$

$$H_z(r, \theta, z) = B_H K_m(\gamma r)\, e^{j(-\beta z + m\theta)}. \tag{4.40}$$

The transverse components are obtained from

$$E_r = -\frac{j}{\kappa^2}\left[\beta\kappa A_E J'_m(\kappa r) + j\omega\mu_0 \frac{m}{r} A_H J_m(\kappa r)\right] e^{j(-\beta z + m\theta)} \tag{4.41}$$

$$E_\theta = -\frac{j}{\kappa^2}\left[j\beta \frac{m}{r} A_E J_m(\kappa r) - \omega\mu_0 \kappa A_H J'_m(\kappa r)\right] e^{j(-\beta z + m\theta)} \tag{4.42}$$

$$H_r = -\frac{j}{\kappa^2}\left[\beta\kappa A_H J'_m(\kappa r) - j\omega\varepsilon_0 n_c^2 \frac{m}{r} A_E J_m(\kappa r)\right] e^{j(-\beta z + m\theta)} \tag{4.43}$$

$$H_\theta = -\frac{j}{\kappa^2}\left[j\beta \frac{m}{r} A_H J_m(\kappa r) + \omega\varepsilon_0 n_c^2 \kappa A_E J'_m(\kappa r)\right] e^{j(-\beta z + m\theta)} \tag{4.44}$$

in the core and

$$E_r = \frac{j}{\gamma^2}\left[\beta\gamma B_E K'_m(\gamma r) + j\omega\mu_0 \frac{m}{r} B_H K_m(\gamma r)\right] e^{j(-\beta z + m\theta)} \tag{4.45}$$

$$E_\theta = \frac{j}{\gamma^2}\left[j\beta \frac{m}{r} B_E K_m(\gamma r) - \omega\mu_0 \gamma B_H K'_m(\gamma r)\right] e^{j(-\beta z + m\theta)} \tag{4.46}$$

$$H_r = \frac{j}{\gamma^2}\left[\beta\gamma B_H K'_m(\gamma r) - j\omega\varepsilon_0 n_a^2 \frac{m}{r} B_E K_m(\gamma r)\right] e^{j(-\beta z + m\theta)} \tag{4.47}$$

$$H_\theta = \frac{j}{\gamma^2}\left[j\beta \frac{m}{r} B_H K_m(\gamma r) + \omega\varepsilon_0 n_a^2 \gamma B_E K'_m(\gamma r)\right] e^{j(-\beta z + m\theta)} \tag{4.48}$$

in the cladding, where the prime stands for the differentiation with respect to the function argument. The unknown constants A_E, A_H, B_E, and B_H are calculated by solving the homogenous set of algebraic equations:

$$\begin{bmatrix} J_m(u) & 0 & -K_m(w) & 0 \\ 0 & J_m(u) & 0 & -K_m(w) \\ j\frac{\beta m}{a\kappa^2}J_m(u) & +\frac{\omega\mu_0}{\kappa}J'_m(u) & j\beta\frac{m}{a\gamma^2}K_m(w) & +\frac{\omega\mu_0}{\gamma}K'_m(w) \\ \frac{\omega\varepsilon_0}{\kappa}n_c^2J'_m(u) & -j\frac{\beta m}{a\kappa^2}J_m(u) & +\frac{\omega\varepsilon_0}{\gamma}n_a^2K'_m(w) & -j\beta\frac{m}{a\gamma^2}K_m(w) \end{bmatrix}\begin{bmatrix} A_E \\ A_H \\ B_E \\ B_H \end{bmatrix} = 0. \tag{4.49}$$

Thus, a step-index circular waveguide guides three types of modes: hybrid modes, TE modes, and TM modes. The propagation constants of hybrid modes are obtained by solving the Hondros–Debye Equation 4.34, while propagation constants of TE and TM modes can be calculated by solving, respectively, Equations 4.35 and 4.36. The fundamental mode of the step-index circular waveguide is the HE_{11} mode. The single-mode region of the step-index circular optical waveguide is set by the cut-off normalized frequency of TE_{01} mode: $V = 2.40482555769577$, where $V = \frac{2\pi}{\lambda}a\sqrt{n_c^2 - n_a^2}$ and λ is the operating wavelength. The detailed description of the optical properties of step-index circular waveguides can be found in a number of textbooks (Okoshi, 1982; Snyder and Love, 1983; Majewski, 1991; Okamoto, 2006; Black and Gagnon, 2010). The analytical theory of a circular waveguide can be extended for multilayered structures (Yeh et al., 1978; Majewski, 1984, 1991).

If the refractive index contrast between the core and cladding is small, all the modes of the circular waveguide become linearly polarized. When $n_c \approx n_a$ Equation 4.34 reduces to

$$\frac{uJ_{l+1}(u)}{J_l(u)} = \frac{wK_{l+1}(w)}{K_l(w)}. \tag{4.50}$$

Equation 4.50 can also be obtained by solving the scalar wave Equation 4.26 for a step-index circular waveguide in the cylindrical coordinate system. The propagation constant of the fundamental mode in the scalar approximation is obtained from Equation 4.50 with $l = 0$. Table 4.2 shows a comparison for the normalized propagation constant $B = \frac{\beta^2 - n_a^2k^2}{n_c^2k^2 - n_a^2k^2}$ of the fundamental mode obtained solving the vectorial dispersion Equation 4.34 and the scalar dispersion Equation 4.50. These results are believed to be calculated with an absolute error of 5×10^{-7}. Table 3.13 shows the relative propagation constant values that have been obtained for the circular waveguide with $n_c = 1.47$, $\lambda = 1.55$ μm, $V = ak\sqrt{n_c^2 - n_a^2}$, whereas varying the cladding refractive index. These results show that the difference between the vectorial and scalar results grows with the increase in the refractive index contrast between the core and the cladding.

TABLE 4.2 Normalized Propagation Constant of the Fundamental Mode of a Single Step-Index Circular Optical Waveguide with $V = 2.4$, $n_c = 1.47$, $\lambda = 1.55$ μm

na	B: Vectorial (4.2.4)	B: Scalar (4.2.8)	Relative Difference
1.45	0.52691	0.53003	5.9e−003
1.4	0.51882	0.53003	2.2e−002
1.3	0.50126	0.53003	5.7e−002
1.2	0.48171	0.53003	1.0e−001
1.1	0.45996	0.53003	1.5e−001
1.0	0.43582	0.53003	2.3e−001

4.2.3 Index Guiding

Slab and circular waveguides are examples of structures that guide light using the phenomenon of total internal reflection. For such structures, the light is guided by the waveguide core, which has a larger refractive index than that of the surrounding medium. Hence, such waveguides are also referred to as index guiding waveguides. There are other types of index guiding waveguides, which are relevant to semiconductor devices; for instance: rectangular, rib and in-diffused waveguides, which are discussed in this section (cf. also Tamir, 1975; Ebeling, 1993; Maerz, 1995; Hunsperger, 2002). The theory of slab and circular waveguides is well established, and the guided mode designation is consistent in the literature. This is also the case for an elliptical waveguide (Dyott, 1996). Unfortunately, in the case of optical waveguides discussed in this section, a consistent mode designation does not exist (cf. for instance Goell, 1969; Gallick et al., 1992) this generates some confusion. Figure 4.6a shows the rectangular dielectric waveguide: the light beam is guided by the core, which has the refractive index n_{f} that is larger than that of the cladding $-n_{\mathrm{c}}$. For a square waveguide with side width w, the single mode condition is given approximately by $w \times (4\pi/\lambda) \times \sqrt{n_{\mathrm{f}}^2 - n_{\mathrm{c}}^2} < 2$ (Majewski and Sujecki, 1996b), where λ is the operating wavelength. For a rectangular waveguide, the single mode condition changes with the aspect ratio w/h. For large aspect ratios w/h, the single mode condition for rectangular waveguide can be approximated by the single mode condition of a slab waveguide (Marcatili, 1969). The detailed study of the modal characteristics of the rectangular waveguide modes can be found in Goell (1969) and Majewski and Sujecki (1996b,a).

Figure 4.6b shows the cross section of an optical rib waveguide. H denotes the outer slab thickness, h is the rib height, and w is the rib width. The areas between the vertical dashed lines and outside of the vertical dashed lines are referred to as the inner and outer slab, respectively. The refractive index of the core is larger than that of the substrate and cladding, i.e., $n_{\mathrm{f}} > \max(n_{\mathrm{s}}, n_{\mathrm{c}})$, which provides vertical index guiding. The lateral index guiding can be best explained with the help of the effective index method (Section 4.2.4). Since the thickness of the inner slab is larger than that of the outer one, the effective index of this layer is also larger than that of the outer one. Hence, it can be concluded that this structure also provides lateral confinement. An alternative way of explaining the waveguiding phenomenon in the case of rib waveguides is available in Marcatili (1974). The detailed studies of general properties of optical rib waveguides, including modal polarization, can be found in Soref et al. (1991), Chiang and Wong (1996), Majewski and Sujecki (1997), and Majewski and Sujecki (1998).

Another structure often used in integrated optics is an in-diffused waveguide (Figure 4.6c). It is fabricated by diffusion of dopants, which increase locally the refractive index in the substrate of refractive index n_{s}. The area with the locally increased value of the refractive index forms the core of the waveguide. The single mode properties of this waveguide strongly depend on the spatial distribution of the refractive index, which is a function of the distribution of the dopants. An experimental study of the properties of in-diffused optical waveguides can be found in Ctyroky et al. (1984).

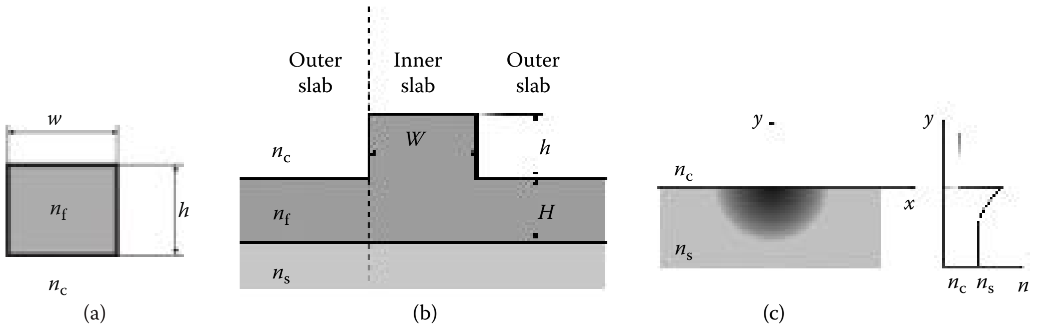


FIGURE 4.6 Rectangular waveguide (a), rib waveguide (b), and in-diffused waveguide (c).

4.2.4 Effective Index Method

It is possible to calculate a one-dimensional (1D) effective refractive index map for structures discussed in Section 4.2.4 so that the analytical techniques presented in Sections 4.2.1 and 4.2.2 can be applied to the calculation of the propagation constants of waveguides from Figure 4.6. Several techniques have been proposed in the literature for this purpose. The most popular effective index method (EIM) consists of approximating a more complex planar waveguide by a slab waveguide (Ramaswamy, 1974). This can be done for both semivectorial (4.1.7) and scalar (4.1.8) cases. EIM is based on the separation of variables, i.e., it relies on the assumption that a two-dimensional (2D) field distribution can be represented by a product of two 1D field distributions. As an illustrating example, we derive an approximate EIM solution in the case of an x-polarized fundamental mode of a rib waveguide. Setting the solution in the form $E_x(x, y, \omega) = X(x, \omega) \times Y(y, \omega)$ into Equation 4.22 gives

$$X\frac{\partial^2 Y}{\partial y^2} + Y\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial(\varepsilon X)}{\partial x}\right) + \omega^2\mu\varepsilon XY = \beta^2 XY. \tag{4.51}$$

Dividing both sides of Equation 4.51 by XY and moving the constants to the right-hand side (RHS) results in

$$\frac{1}{Y}\frac{\partial^2 Y}{\partial y^2} + \frac{1}{X}\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial(\varepsilon X)}{\partial x}\right) = -\left(\omega^2\mu\varepsilon - \beta^2\right), \tag{4.52}$$

which implies that a separation constant can be introduced to obtain two coupled 1D equations

$$\frac{\partial^2 Y}{\partial y^2} + \omega^2\mu\varepsilon Y = \beta_{\text{eff}}^2 Y \tag{4.53}$$

$$\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial(\varepsilon X)}{\partial x}\right) + \omega^2\mu\varepsilon_0 n_{\text{eff}}^2 X = \beta^2 X, \tag{4.54}$$

where the effective index of the slab waveguide fundamental mode $n_{\text{eff}} = \beta_{\text{eff}}/k_0$ while $k_0 = 2\pi/\lambda$. Once the dependence of n_{eff} on x is calculated using Equation 4.53, Equation 4.54 is used to obtain the propagation constant for the fundamental mode of the "effective" slab waveguide, which gives the approximate value of the propagation constant of the fundamental mode of the original 2D waveguide. The entire process is summarized in Figure 4.7.

Alternatively, one may consider a polar coordinate system and calculate the effective index along the azimuthal direction (Marcatili and Hardy, 1988). This requires a calculation of the propagation constant of a periodic waveguide (Figure 4.8). Once the effective index map is calculated, the approximate equivalent circular waveguide structure is obtained. The propagation constants of this equivalent circular waveguide can be calculated applying the analytical techniques discussed in Section 4.2.2.

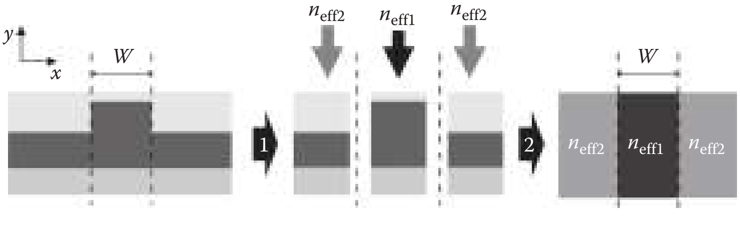


FIGURE 4.7 Schematic illustration of the effective index method in the case of a rib-loaded optical waveguide.

4.2.5 Radiation Modes, Leaky Waves, Gain Guiding, and Supermodes

If the angle of incidence for a ray shown in Figure 4.4 is smaller than the critical angle, the ray will undergo refraction at the boundary between the core and the cladding. In Figure 4.9, we show three possible scenarios for a slab waveguide with $n_f > n_s > n_c$. The case shown in Figure 4.9a corresponds to a guided mode that was already discussed in Section 4.2.1. The cases shown in Figures 4.9b and c correspond to radiation modes, i.e., modes that are not guided by the waveguide core. In fact, once the incidence angle is less than the critical angle, the ray may undergo refraction only at the interface between the core and substrate, which corresponds to a substrate radiation mode. If the ray undergoes refraction also at the cladding-core interface, the mode is referred to as free space radiation mode. Assuming that $\partial/\partial y = 0$ $\partial/\partial y = 0$, the solutions of the Maxwell equation for both modes can be obtained by solving Equations 4.24 and 4.22 for TE and TM modes, respectively (Marcuse, 1972, 1974; Ebeling, 1993). Figure 4.10 shows calculated distributions of the E_y electromagnetic field component

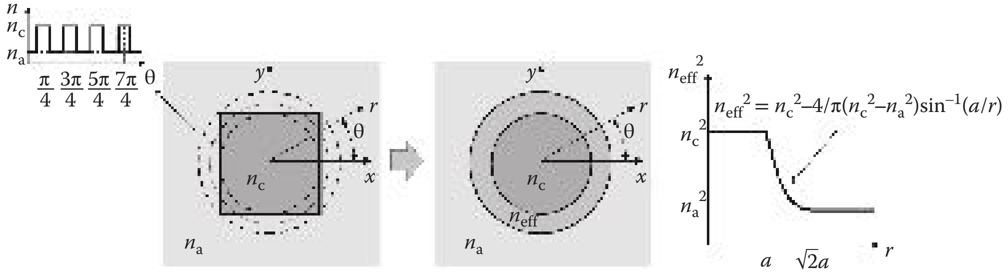

n
n_c
n_a
$\frac{\pi}{4}$ $\frac{3\pi}{4}$ $\frac{5\pi}{4}$ $\frac{7\pi}{4}$ θ
y
r
θ
x
n_c
n_a
y
r
θ
x
n_c
n_{eff}
n_a
n_{eff}^2
$n_{eff}^2 = n_c^2 - 4/\pi(n_c^2 - n_a^2)\sin^{-1}(a/r)$
n_c^2
n_a^2
a $\sqrt{2}a$ r

FIGURE 4.8 Schematic illustration of the azimuthal effective index method in the case of a square optical waveguide. (From Marcatili, E.A.J. and Hardy, A.A., *IEEE Journal of Quantum Electronics* 24, 766–774, 1988.)

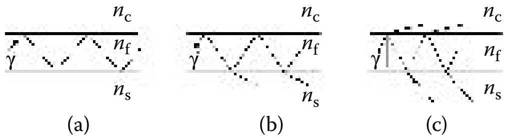


FIGURE 4.9 Optic ray paths in a slab waveguide for a guided mode (a), substrate radiation mode (b), and free space radiation mode (c).

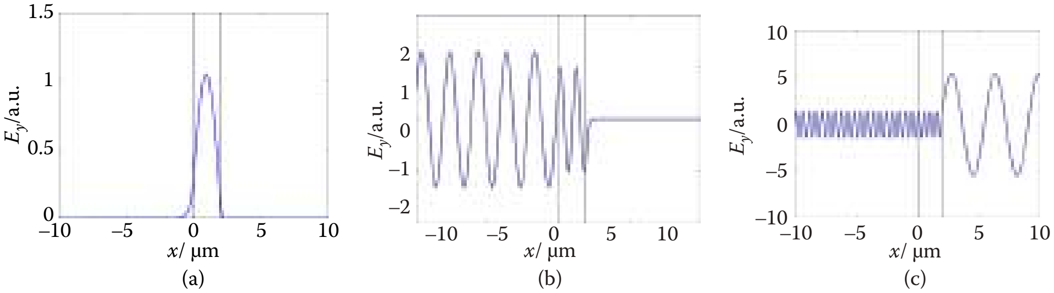


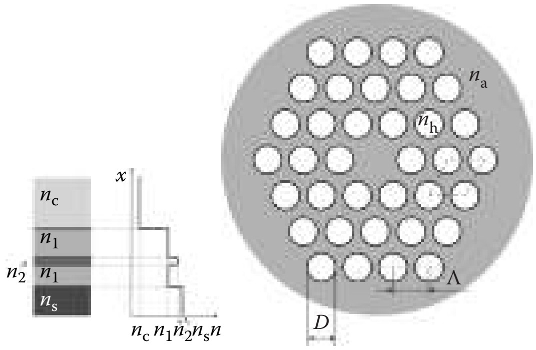
FIGURE 4.10 Field distributions of the E_y component of an electric field in a slab waveguide for a TE guided mode (a), substrate radiation mode (b), and free space radiation mode (c). The waveguide with is 2 μm, wavelength 1.55 μm, and $n_c = 1$, $n_f = 3.3$, and $n_s = 3.1$.

obtained by directly solving Maxwell equations for a guided mode and substrate and free space radiation modes.

Even if the refractive index distribution is purely real, Equations 4.28 through 4.31 admit solutions for the complex values of the propagation constant β. These modes are referred to as "leaky modes." A characteristic feature of leaky modes is an exponential growth of the field away from the core. A simple explanation of this field behavior is provided in Rozzi and Mongiardo (1997). Leaky modes are very important when studying the optical properties of waveguide modes below cut-off, bent waveguides and antireflection resonant optical waveguides (ARROWs) (Marcuse, 1974; Snyder and Love, 1983; Jablonski, 1994). Figure 4.11a shows an example cross section of an ARROW (Kubica, 2002). The top layer with refractive index n_1 guides the light and hence acts as the waveguide core. The layer with refractive index n_2 and the layer below it are designed to operate as a set of two antiresonant Fabry–Pérot resonators, which reduce the optical leakage from the core layer. ARROWs have found many applications, e.g., in laser diodes (Bhattacharya et al., 1996). Another waveguide that guides leaky waves is based on the concept of a photonic crystal (PC) and is relevant to vertical cavity surface emitting lasers (VCSELs) (Dems et al,, 2010). One example of such structure is a solid core photonic crystal waveguide (PCW), which is shown in Figure 4.11b. This waveguide consists of a set of air holes that surround the solid core. In this particular example, the air holes make a triangular lattice pattern. However, other arrangements of holes are also possible. In principle, the holes are arranged to prevent an optical wave from propagation within the air hole region at the waveguide operating wavelength. This is done by selecting the waveguide operating wavelength within the stop band of the PC (Joannopoulos et al., 1995). This results in an optical mode that is confined to the solid core (Bjarklev et al., 2003; Zolla, 2005; Poli et al., 2007). If the PC extended to infinity, such a waveguide would guide optical waves in the core region without any leakage. However, due to technological limitations, in practice only several PC layers surround the core, which results in the leakage of light.

Another wave-guiding mechanism that is relevant to optoelectronic devices is gain-guiding. Gain-guiding structures are exclusively used in semiconductor lasers. The current flowing through the active region of a laser diode creates an area of positive gain, which guides the light laterally (Petermann, 1991). The modes of gain-guided waveguides are known to have curved phase fronts (Petermann, 1991).

The final structure we discuss in this section is an optical waveguide array (Figure 4.12). Optical waveguide arrays are used in semiconductor lasers to scale up the output power of single emitter laser diodes. The modes of such structures are called supermodes to differentiate them from the modes of waveguides, which form the array. Figure 4.12 shows as an example of an array consisting of three rib waveguides. If each rib waveguide separately guides only one mode, then the entire array guides three supermodes. To illustrate the difference between the individual waveguide modes and the waveguide array supermodes, we used the EIM to calculate the field distributions $X(x)$, cf. Section 4.2.4, of all three supermodes (Figure 4.13) for an array consisting of three rib waveguides.



(a) (b)

FIGURE 4.11 Antireflection resonant optical waveguides (ARROW) (a) and photonic crystal waveguide (b).

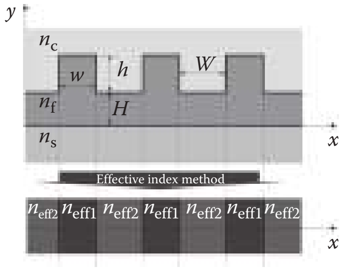

y
n_c
w
h
W
n_f
H
n_s
x
Effective index method
n_{eff2} n_{eff1} n_{eff2} n_{eff1} n_{eff2} n_{eff1} n_{eff2}
x

FIGURE 4.12 Optical waveguide array consisting of three rib waveguides and its effective index map.

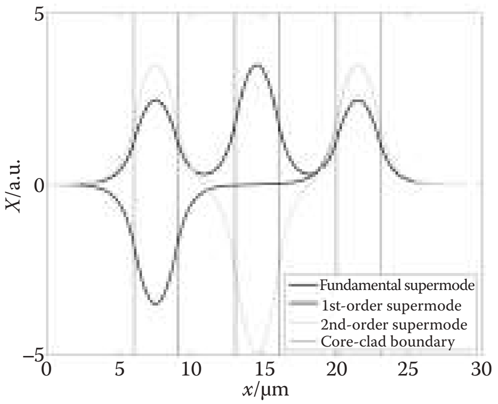


FIGURE 4.13 Lateral mode distributions $X(x)$ for an optical waveguide array consisting of three waveguides (Figure 4.12) calculated using the effective index method (EIM): $W = 4$ µm, $w = 3$ µm, $H = 4$ µm, $h = 2$ µm, $n_c = 1$, $n_f = 3.5$, $n_s = 1.45$, and $l = 1.55$ µm.

4.3 Beam Propagation Method

The operation of many optoelectronic devices relies on the phenomenon of light guiding along optical waveguides. For instance, optical waveguides are used in laser diodes, semiconductor laser amplifiers, electro-optic modulators, etc. The numerical techniques that were discussed in Section 4.2 are often insufficient for the study of light propagation in optoelectronic devices. This is because the waveguiding structures used in optoelectronic devices are not infinitely long and may not be straight. In principle, the electromagnetic field distribution within an optoelectronic device can be calculated using a general purpose Maxwell equations solver, e.g., finite difference time domain method (FDTDM). Such an approach, however, tends not to be practical because general purpose techniques calculate the field distribution in the entire domain. Hence, the resulting matrices storing all field sampling points within the computational domain are very large. For instance, 20 samples per wavelength for an optical waveguide that is 10,000 wavelengths long (a typical length of a high-power edge emitting laser diode) would result in 200,000 samples just for a 1D analysis. The beam propagation method (BPM) was developed to address these difficulties. Unlike the direct Maxwell equations solution methods, BPM allows calculating the spatial distribution of the output beam subject to the input beam shape (Figure 4.14) without the necessity of storing the field samples in the entire computational domain. In this section, we provide a short description of BPM.

In order to derive the basic equations relevant to BPM, we assume that the light propagation is taking place along the z-axis (Figure 4.1). In the rectangular coordinate system (Figure 4.3), the following set of

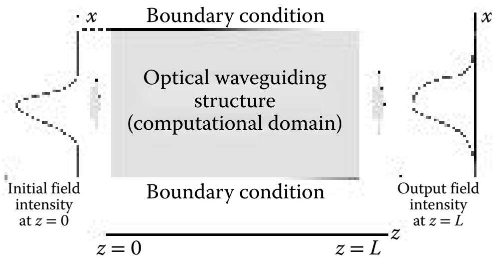

x
Boundary condition
Optical waveguiding structure (computational domain)
Boundary condition
Initial field intensity at z = 0
Output field intensity at z = L
x
z
z = 0
z = L

FIGURE 4.14 Schematic representation of the beam propagation method (BPM) application to an optical waveguiding structure in rectangular coordinate system.

three equations for the electric field vector components are obtained from Equation 4.3:

$$\frac{\partial^2 E_x}{\partial y^2}+\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial\left(\varepsilon E_x\right)}{\partial x}\right)+\omega^2\mu\varepsilon E_x+\frac{\partial^2 E_x}{\partial z^2}+\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial\left(\varepsilon E_y\right)}{\partial y}\right)-\frac{\partial^2 E_y}{\partial x\partial y}+\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}E_z\frac{\partial\varepsilon}{\partial z}\right)=0 \quad (4.55)$$

$$\frac{\partial^2 E_y}{\partial x^2}+\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial\left(\varepsilon E_y\right)}{\partial y}\right)+\omega^2\mu\varepsilon E_y+\frac{\partial^2 E_y}{\partial z^2}+\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial\left(\varepsilon E_x\right)}{\partial x}\right)-\frac{\partial^2 E_x}{\partial y\partial x}+\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}E_z\frac{\partial\varepsilon}{\partial z}\right)=0 \quad (4.56)$$

$$\frac{\partial^2 E_z}{\partial x^2}+\frac{\partial^2 E_z}{\partial y^2}+\omega^2\mu\varepsilon E_z+\frac{\partial}{\partial z}\left(\frac{1}{\varepsilon}\frac{\partial\left(\varepsilon E_z\right)}{\partial z}\right)+\frac{\partial}{\partial z}\left(\frac{1}{\varepsilon}E_x\frac{\partial\varepsilon}{\partial z}\right)+\frac{\partial}{\partial z}\left(\frac{1}{\varepsilon}E_y\frac{\partial\varepsilon}{\partial z}\right)=0. \quad (4.57)$$

Similarly for the magnetic field vector components:

$$\frac{\partial^2 H_x}{\partial x^2}+\varepsilon\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial H_x}{\partial y}\right)+\omega^2\mu\varepsilon H_x+\varepsilon\frac{\partial}{\partial z}\left(\frac{1}{\varepsilon}\frac{\partial H_x}{\partial z}\right)+\frac{1}{\varepsilon}\left(\frac{\partial\varepsilon}{\partial y}\frac{\partial H_y}{\partial x}+\frac{\partial\varepsilon}{\partial z}\frac{\partial H_z}{\partial x}\right)=0 \quad (4.58)$$

$$\varepsilon\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial H_y}{\partial x}\right)+\frac{\partial^2 H_y}{\partial y^2}+\omega^2\mu\varepsilon H_y+\varepsilon\frac{\partial}{\partial z}\left(\frac{1}{\varepsilon}\frac{\partial H_y}{\partial z}\right)+\frac{1}{\varepsilon}\left(\frac{\partial\varepsilon}{\partial x}\frac{\partial H_x}{\partial y}+\frac{\partial\varepsilon}{\partial z}\frac{\partial H_z}{\partial y}\right)=0 \quad (4.59)$$

$$\varepsilon\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial H_z}{\partial x}\right)+\varepsilon\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial H_z}{\partial y}\right)+\omega^2\mu\varepsilon H_z+\frac{\partial^2 H_z}{\partial z^2}+\frac{1}{\varepsilon}\left(\frac{\partial\varepsilon}{\partial x}\frac{\partial H_x}{\partial z}+\frac{\partial\varepsilon}{\partial y}\frac{\partial H_y}{\partial z}\right)=0. \quad (4.60)$$

We assume that the analyzed optical waveguiding structure varies slowly along the z-direction and hence $\partial\varepsilon/\partial z \approx 0$ (Vassallo, 1997a). As a consequence, the transverse components of the electromagnetic field decouple from the longitudinal ones and hence for the transverse field components, the set of Equations 4.55 through 4.57 can be approximated by the following equation:

$$\frac{\partial\vec{E}_{\mathrm{t}}}{\partial z^2}=-\left(A+\omega^2\mu\varepsilon\right)\vec{E}_{\mathrm{t}}, \quad (4.61)$$

and Equations 4.58 through 4.60 can be approximated by

$$\frac{\partial\vec{H}_{\mathrm{t}}}{\partial z^2}=-\left(B+\omega^2\mu\varepsilon\right)\vec{H}_{\mathrm{t}}, \quad (4.62)$$

where $\vec{E}_{\mathrm{t}}=\vec{i}_xE_x+\vec{i}_yE_y$and $\vec{H}_{\mathrm{t}}=\vec{i}_xH_x+\vec{i}_yH_y$. The elements of matrices A and B are provided in Tables 4.3 and 4.4.

TABLE 4.3 Elements a_{ji} of Matrix A

$a_{j,i}$	$i = 1$	$i = 2$
$j = 1$	$\frac{\partial^2}{\partial y^2} + \frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial(\varepsilon)}{\partial x}\right)$	$\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial(\varepsilon)}{\partial y}\right) - \frac{\partial^2}{\partial x \partial y}$
$j = 2$	$\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial(\varepsilon)}{\partial x}\right) - \frac{\partial^2}{\partial y \partial x}$	$\frac{\partial^2}{\partial x^2} + \frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial(\varepsilon)}{\partial y}\right)$

TABLE 4.4 Elements b_{ji} of Matrix B

$b_{j,i}$	$i = 1$	$i = 2$
$j = 1$	$\frac{\partial^2}{\partial x^2} + \varepsilon\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial}{\partial y}\right)$	$-\varepsilon\frac{\partial}{\partial y}\left(\frac{1}{\varepsilon}\frac{\partial}{\partial x}\right) + \frac{\partial^2}{\partial y \partial x}$
$j = 2$	$-\varepsilon\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial}{\partial y}\right) + \frac{\partial^2}{\partial x \partial y}$	$\frac{\partial^2}{\partial y^2} + \varepsilon\frac{\partial}{\partial x}\left(\frac{1}{\varepsilon}\frac{\partial}{\partial x}\right)$

Equations 4.61 and 4.62 form the basis of most BPM algorithms. Similarly, as in Section 4.1, Equations 4.61 and 4.62 can be further simplified for waveguiding structures, which support polarized modes. For the x-polarized modes, Equations 4.61 and 4.62 reduce to

$$\frac{\partial^2 E_x}{\partial z^2} = -\left(a_{11} + \omega^2\mu\varepsilon\right) E_x \tag{4.63}$$

and

$$\frac{\partial^2 H_y}{\partial z^2} = -\left(b_{22} + \omega^2\mu\varepsilon\right) H_y, \tag{4.64}$$

while for y-polarized modes we obtain

$$\frac{\partial^2 E_y}{\partial z^2} = -\left(a_{22} + \omega^2\mu\varepsilon\right) E_y \tag{4.65}$$

and

$$\frac{\partial^2 H_x}{\partial z^2} = -\left(b_{11} + \omega^2\mu\varepsilon\right) H_x. \tag{4.66}$$

The operators a_{ji} and b_{ji} are given in Tables 4.3 and 4.4.

Furthermore, if ε varies slowly with the transverse coordinates one can use the scalar approximation

$$\frac{\partial^2 \Phi}{\partial z^2} = -\left(s + \omega^2\mu\varepsilon\right) \Phi, \tag{4.67}$$

where Φ is the scalar potential and $s = \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial x^2}$.

Thus, we obtain the same hierarchy of approximations as in Section 4.1. Further improvement in efficiency can be achieved using the effective index approximation to calculate the effective index map and thus reduce the number of spatial dimensions considered. It can be observed that Equations 4.61 and 4.62 through Equation 4.67 have the same generic form, namely

$$\frac{\partial^2 F}{\partial z^2} = -\left(L + \omega^2\mu\varepsilon\right) F. \tag{4.68}$$

The first step in deriving a BPM algorithm for the solution of Equation 4.68 is the derivation of the one-way wave equation. This step is usually accomplished by formally factoring Equation 4.68

$$\left(\frac{\partial}{\partial z}-j\sqrt{L+\omega^2\mu\varepsilon}\right)\left(\frac{\partial}{\partial z}+j\sqrt{L+\omega^2\mu\varepsilon}\right)F=0.$$

If only one of the product components is considered, the so-called one-way wave equation is obtained

$$\frac{\partial \mathrm{F}}{\partial z}=-j\sqrt{L+\omega^2\mu\varepsilon}F. \tag{4.69}$$

After the introduction of the slowly varying (with z) envelope function Φ, defined by $F = \Phi(x, y, z) * \exp(-j\beta_r z)$ where β_r is a suitably selected reference propagation constant, Equation 4.69 transforms into

$$\frac{\partial \Phi}{\partial z}=j\beta_r\left(-\sqrt{1+\frac{L+\omega^2\mu\varepsilon-\beta_r^2}{\beta_r^2}}+1\right)\Phi. \tag{4.70}$$

The formal integration of Equation 4.70 yields

$$\Phi\left(z_0+\Delta z\right)=\exp\int\limits_{z=z_0}^{z=z_0+\Delta z}\left[j\beta_r\left(-\sqrt{1+\frac{L+\omega^2\mu\varepsilon-\beta_r^2}{\beta_r^2}}+1\right)\right]\Phi\left(z_0\right)dz. \tag{4.71}$$

If the computational domain (Figure 4.14) is divided into a concatenation of sections with a z-independent refractive index distribution, and within each section the rectangle numerical integration rule is applied, then Equation 4.71 reduces to

$$\Phi\left(z_0+\Delta z\right)=\exp\left[j\beta_r\left(-\sqrt{1+\frac{L+\omega^2\mu\varepsilon-\beta_r^2}{\beta_r^2}}+1\right)\right]\Delta z\Phi\left(z_0\right). \tag{4.72}$$

Equation 4.72 implies an algorithm, which calculates the value of the field distribution at the position $z_0 + \Delta z$ subject to the known field distribution at the previous position z_0. Thus, the entire calculation can be started from the initial field distribution at $z = 0$ (Figure 4.14) and continued until the output field distribution at $z = L$ is obtained. At each step, only the field distribution from the previous step needs to be stored. This makes a BPM algorithm much more manageable in terms of the required computer memory than general purpose Maxwell equations solvers.

The last step left is the conversion of Equation 4.72 into an efficient computer code. For this purpose, suitable approximations for the L, exponential, and square root operators need to be found. A number of algorithms were developed for this purpose. The most often used approximations for the L operator apply the FDM, the finite element method, or local waveguide mode expansion. In the case of the exponential and square root operators, the typical approach consists of expanding the operator using either Padé or Taylor series. For instance, an approximation of the square root operator using the Taylor series truncated after the first term results in a paraxial approximation, while applying Padé (1,1) expansion to the exponential operator, results in the Crank–Nicholson scheme. A comprehensive survey and a fairly detailed description of BPM operator approximation algorithms can be found in Yamauchi (2003) and Sujecki (2014). A popular choice is the selection of paraxial approximation for the square root operator, Padé (1,1) approximation for the exponential operator, and finite difference approximation for the L operator. Such selection yields the finite difference, Crank–Nicholson paraxial BPM algorithm (FDCNP-BPM). FDCNP-BPM is very effective if the wave propagation takes place within a narrow cone centered along the z-axis. If this condition is not fulfilled, then Padé approximations are used for the square root operator, resulting in wide angle (WA) BPM algorithms. In Section 4.4, we provide examples of BPM application.

4.4 Tapered and Bent Waveguides

The waveguides used in semiconductor lasers or semiconductor optical amplifiers may be intentionally tapered or bent. For instance, tapered lasers can offer high output power without compromising on the beam quality, while bent semiconductor optical amplifiers allow reducing the coupling of the back reflected waves into the amplifier. In this section, we discuss the optical properties of tapered and bent waveguides and their design methods.

4.4.1 Tapered Waveguides

A very useful intuitive insight into the operation of tapered optical waveguides can be gained through applying ray optics. Figure 4.15 shows an optical ray entering an optical taper with an initial incidence angle θ. The taper is symmetrical with respect to the z-axis and its angle is α. By means of geometric optics, it can readily be demonstrated that after the reflection from the boundary between the core and cladding the ray will travel downward and will impinge upon the lower boundary at an incidence angle $\theta-\alpha$ (Cada et al., 1988). The same applies to the subsequent reflections: at each subsequent reflection, the angle of incidence decreases by the value of the taper angle α. The obvious consequence of this fact is that even if the ray is traveling initially at an angle significantly exceeding the critical angle, then after several reflections it might start traveling at an angle that is less than the critical angle and thus start leaking power through refraction at the boundary between the core and cladding (Zheng, 1989). If this process is continued, the ray may eventually start traveling backward, thus leading to the conclusion that a part of the incident optical power is reflected back. All these predictions, arrived at intuitively using ray optics, have been confirmed with advanced numerical models and experiments. Further, when a taper is used to connect two waveguides, then it is obvious that the ray incidence angle is different at the output when compared with the input waveguide (Figure 4.15). This leads to a conclusion that a mode conversion may occur if both waveguides are multimode (cf. Li and Lit, 1986; Ravets et al., 2013).

Further useful insight might be obtained using the wave optics and the electromagnetic field theory. Such an analysis is particularly useful when attempting to establish guidelines for the design of tapers with negligibly small losses. The formulation of the wave equation in the cylindrical coordinate system leads to a conclusion that a lossless taper can be designed if the V parameter of the tapered waveguide is kept constant (Marcatili, 1985). This rule, however, applies without ambiguity only when tapering step index slab and circular waveguides. Also such prediction is subject to an assumption that an optical taper can be approximated by an infinitely extending wedge or cone. If a taper connects two waveguides (Figure 4.16), then the simulations show that a significant portion of the loss results from the mode mismatch at the input and output of the taper (Zheng, 1989). The conformal mapping shows that such a loss can be reduced by

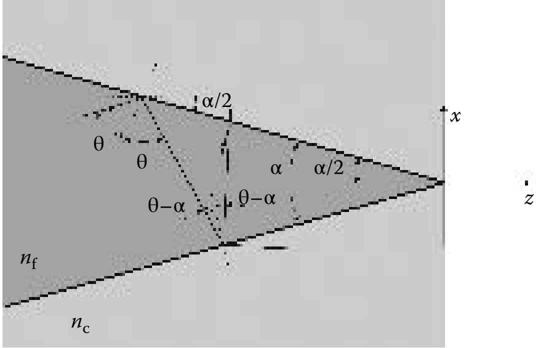


FIGURE 4.15 Optical ray path in an optical taper.

inserting a lens both at the taper input and output (Lee et al., 1997). Other advanced numerical simulations show that the loss can also be reduced by shaping the taper (Luyssaert et al., 2005).

An accurate analysis of tapered structures is complicated by the nonseparability of the wave equation in standard coordinate systems. Therefore, an application of numerical methods is necessary. BPM turned out to be particularly useful if the effect of back reflection is negligible.

Table 4.5 contains the parameters of four optical taper structures, which were studied extensively in the literature (Haes et al., 1996). Table 4.6 provides the calculated power content in the fundamental x-polarized mode at the end of the taper

$$P_c = \frac{\int E_y^*(x, z = 0)\, E_y(x, z = L)\, dx}{\int E_y^*(x, z = 0)\, E_y(x, z = 0)\, dx}. \tag{4.73}$$

Table 4.6 provides a comparison between the results calculated by FDCNP-BPM and finite difference Crank–Nicolson wide angle beam propagation method (FDCNWA-BPM), which uses Padé (1,1) approximation. The largest difference between both sets of results is one on the first decimal place. These results demonstrate the usefulness of the paraxial approximation in modeling and design of the slowly varying optical waveguiding structures.

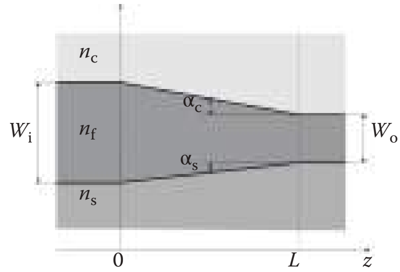


FIGURE 4.16 Optical taper connecting two optical slab waveguides.

TABLE 4.5 Slab Waveguide Taper Test Structures (Figure 4.16)

	W_i/μm	W_o/μm	L/μm	n_c	n_f	n_s	α_s
Taper 1	0.2	0.1	5.73	3.17	3.3	3.17	0
Taper 2	0.2	0.1	57.3	3.17	3.3	3.17	0
Taper 3	0.8	0.4	22.9	1.0	3.3	3.17	0
Taper 4	0.8	0.4	229	1.0	3.3	3.17	0

TABLE 4.6 Power Content in the Fundamental x-Polarized Mode at $z = L$ P_c Calculated from Equation 4.73 for Test Structures from Table 4.5 (Figure 4.16)

	Taper 1 (%)	Taper 2 (%)	Taper 3 (%)	Taper 4 (%)
FDCNP-BPM	91.18	97.29	82.24	96.63
FDCNWA-BPM	91.19	97.30	82.34	96.63

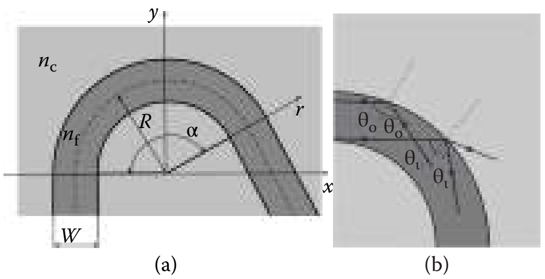
Source: Sujecki S., *International Journal of Electronics and Communications (AEU)*, 55, 185–190, 2001.

4.4.2 Bent Waveguides

Figure 4.17a shows a section of a circularly bent optical slab waveguide, which connects two straight waveguides. The bent radius *R* is measured with respect to the center of the waveguide. The angle α gives a measure of the azimuthal length of the bent. An optical beam experiences an optical loss, off axis offset and a polarization rotation while propagating through a bent waveguide section.

Some intuitive insight supporting the statement that an optical beam experiences power loss while propagating through a bend can be gained by observing that an optical mode of a straight waveguide after entering a bent waveguide section would have to preserve flat phase fronts in order to avoid losses. This, however, is not possible since the phase velocity within the bent waveguide section would have to grow with radius *r* if flat phase fronts were to be preserved. Thus, for a certain value of radius *r* the phase velocity would have to become larger than the light velocity of the medium thus implicating the necessity of phase front bending and the power loss through radiation from the bent (Marcatili, 1985). Alternatively, one can arrive at the same conclusion by tracking the paths of optical rays entering a bent waveguide (Figure 4.17b). An optical ray entering a bend near the outer boundary between the core and the cladding impinges upon the bend boundary at a large incidence angle and thus undergoes a total internal reflection. However, an optical ray, which enters the bend near the inner boundary between the core and cladding, has a small incidence angle and hence may also undergo refraction and hence lose power. These intuitive predictions have been confirmed by numerical simulations (cf. Hiremath et al., 2005).

The radiation into the outer space surrounding the bend is, however, not the only source of optical loss. An optical bend usually connects two straight waveguide sections. Hence, loss can be incurred by the mismatch between the field distribution of a straight waveguide mode and of the bent waveguide mode. This effect is related to the fact that the modes of an optical bend have different field distributions than those of the straight waveguide even if both straight and bent waveguides have the same transverse distribution of the refractive index (Hiremath et al., 2005). When compared with the modes of the straight waveguide, the modes of the bent waveguide have an off-axis offset. This effect can be most convincingly demonstrated using the conformal mapping to derive a refractive index distribution of an equivalent straight waveguide (Ladouceur and Labeye, 1995). In fact, conformal mapping provides another argument supporting the claim that an optical bend incurs a loss upon an optical beam. A logarithmic transformation, for instance, allows mapping of a section of a circular ring onto a straight stripe and can be used to derive the refractive index distribution of an equivalent straight waveguide (Heiblum and Harris, 1975). The resulting equivalent refractive index distribution grows away from the bent waveguide center. When the bent radius is large, the refractive index growth can be locally approximated by a linear function (Figure 4.18; Ladouceur and Labeye, 1995). Thus, at a certain value of the transverse coordinate, the value of the refractive index in the cladding exceeds the largest value of the refractive index in the core. This leads to a conclusion that all modes of such waveguide are leaky and hence inherently lost. Further, since the refractive index distribution is modified the fundamental mode intensity distribution of the bent waveguide is different from that



(a) (b)

FIGURE 4.17 Bent optical slab waveguide connecting two straight waveguides (a) and the optical ray paths in a bent waveguide (b).

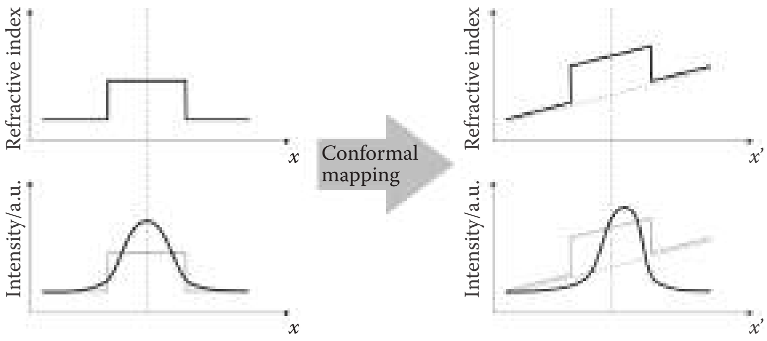


FIGURE 4.18 Schematic illustration of the mode offset in a bent waveguide.

of the straight one. The field is more concentrated in the outer part of the core and has an off-axis offset. Consequently, there is a mode mismatch between the straight and curved waveguide sections, which incurs a coupling loss. The effect of this phenomenon on the overall bent waveguide loss can be reduced by offsetting the straight waveguide (Kitoh et al., 1995). Another consequence of the fact that the modal field distributions in the straight and bent waveguides are different is the excitation of the first-order mode in the bent waveguide even if only the fundamental mode is incident from the straight waveguide. If this is the case, both the fundamental mode and the first-order mode propagate with different propagation constants along the bent waveguide and hence combine with different phases at the output. Thus, the resulting field pattern may not match the fundamental mode of the straight waveguide at the output, which results in a coupling loss. The loss resulting from the excitation of the first-order mode can be minimized either by appropriate selection of the bent radius and angle (Melloni et al., 2004) or by offsetting the straight waveguide (Hirono et al., 1998).

Finally, an optical beam, although propagating through a bent waveguide section, may undergo a change of its polarization state. This phenomenon is used to design polarization rotators (Deng et al., 2005). The calculation of the polarization rotation is not straightforward and requires either an application of the coupled mode theory or a vectorial BPM (Lui et al., 1998).

Finally, we note that a number of simple approximate techniques are available for the calculation of the bending radiation loss in various waveguide structures (Vassallo, 1991). However, a calculation of the full electromagnetic field distribution within an arbitrary bent waveguide requires an application of numerical methods. There are two ways in which the problem of calculating the electromagnetic field distribution within a bent waveguide can be formulated. One formulation of the problem consists of considering a complete ring and calculation of the complex resonant frequencies of the modes. In the other, the frequency is assumed to be a known parameter and the field distribution is calculated within the bent waveguide using for instance a BPM algorithm (Sujecki, 2014). Alternatively, for circularly bent slab waveguides an analytical solution for the propagation constants and field distributions of the guided modes is possible (Hiremath et al., 2005). The advantage of BPM is the possibility of inclusion in the model of all effects that accompany optical beam propagation through a bent waveguide. However, an optical mode solver provides a very useful insight into the bent waveguide characteristics.

References

Anemogiannis E, Glytsis EN (1992) Multilayer wave-guides—Efficient numerical-analysis of general structures. *Journal of Lightwave Technology* 10:1344–1351.

Anemogiannis E, Glytsis EN, Gaylord TK (1994) Efficient solution of eigenvalue equations of optical waveguiding structures. *Journal of Lightwave Technology* 12:2080–2084.

Anemogiannis E, Glytsis EN, Gaylord TK (1999) Determination of guided and leaky modes in lossless and lossy planar multilayer optical waveguides: Reflection pole method and wavevector density method. *Journal of Lightwave Technology* 17:929–941.

Bhattacharya A, Mawst LJ, Nesnidal MP, Lopez J, Botez D (1996) 0.4W CW diffraction limited beam Al free 0.98 μ/m wavelength three core ARROW-type diode lasers. *Electronics Letters* 32:657–658.

Bjarklev A, Bjarklev A, Broeng J (2003) *Photonic Crystal Fibres*. London: Kluwer Academic Publishers.

Black RJ, Gagnon L (2010) *Optical Waveguide Modes: Polarization, Coupling and Symmetry*. London: McGraw-Hill.

Cada M, Feng XA, Felsen LB (1988) Intrinsic modes in tapered optical wave-guides. *IEEE Journal of Quantum Electronics* 24:758–765.

Chiang KS, Wong WP (1996) Rib waveguides with degenerate polarised modes. *Electronics Letters* 32:1098–1099.

Ctyroky J, Hofman M, Janta J, Schrofel J (1984) 3-D analysis of $LiNbO_3$-Ti channel waveguides and directional-couplers. *IEEE Journal of Quantum Electronics* 20:400–409.

Dems M, Chung I-S, Peter N, Bischoff S, Panajotov K (2010) Numerical methods for modeling photonic-crystal VCSELs. *Optics Express* 18:16042–16054.

Deng HH, Yevick DO, Chaudhuri SK (2005) Bending characteristics of asymmetric SOI polarization rotators. *IEEE Photonics Technology Letters* 17:2113–2115.

Ding H, Chan KT (1997) Solving planar dielectric waveguide equations by counting the number of guided modes. *IEEE Photonics Technology Letters* 9:215–217.

Dyott RB (1996) *Elliptical Fiber Waveguides*. Boston: Artech House.

Ebeling KJ (1993) *Integrated Optoelectronics: Waveguide Optics, Photonics, Semiconductors*. Berlin: Springer-Verlag.

Gallick AT, Kerkhoven T, Ravaioli U (1992) Iterative solution of the eigenvalue problem for a dielectric waveguide. *IEEE Transactions on Microwave Theory and Techniques* 40:699–705.

Goell JE (1969) A circular harmonic computer analysis of rectangular dielectric waveguides. *Bell System Technical Journal* 48:2133–2160.

Hadley GR (2002) High accuracy finite difference equations for dielectric waveguide analysis II: Dielectric corners. *Journal of Lightwave Technology* 20:1219–1231.

Haes J et al. (1996) A comparison between different propagative schemes for simulation of tapered step index slab waveguides. *Journal of Lightwave Technology* 14:1557–1567.

Heiblum M, Harris JH (1975) Analysis of curved optical-waveguides by conformal transformation. *IEEE Journal of Quantum Electronics QE* 11:75–83.

Hiremath K, Hammer M, Stoffer R, Prkna L, Ctyroky J (2005) Analytic approach to dielectric optical bent slab waveguides. *Optical and Quantum Electronics* 37:37–61.

Hirono T, Kohtoku M, Yoshikuni Y, Lui WW, Yokoyama K (1998) Optimized offset to eliminate first-order mode excitation at the junction of straight and curved multimode waveguides. *IEEE Photonics Technology Letters* 10:982–984.

Hondros D, Debye P (1910) Electromagnetic waves in dielectrical wires. *Annalen der Physik* 32:465–476.

Hunsperger RG (2002) *Integrated Optics*. Berlin: Springer-Verlag.

Jablonski TF (1994) Complex modes in open lossless dielectric waveguides. *Journal of the Optical Society of America A: Optics, Image Science, and Vision* 11:1272–1282.

Joannopoulos JD, Meade RD, Winn JN (1995) *Photonic Crystals: Molding the Flow of Light*. Princeton, NJ: Princeton University Press.

Kitoh T, Takato N, Yasu M, Kawachi M (1995) Bending loss reduction in silica-based wave-guides by using lateral offsets. *Journal of Lightwave Technology* 13:555–562.

Kubica JM (2002) Analysis of ARROWs with thin cores. *Optical and Quantum Electronics* 34:737–745.

Ladouceur F, Labeye P (1995) A new general-approach to optical wave-guide path design. *Journal of Lightwave Technology* 13:481–492.

Lee CT, Wu ML, Hsu JM (1997) Beam propagation analysis for tapered waveguides: Taking account of the curved phase-front effect in paraxial approximation. *Journal of Lightwave Technology* 15:2183–2189.

Li YF, Lit JWY (1986) Mode-changes in step-index multimode fiber tapers. *Journal of the Optical Society of America A: Optics, Image Science, and Vision* 3:161–164.

Lui WW, Hirono T, Yokoyama K, Huang WP (1998) Polarization rotation in semiconductor bending waveguides: A coupled-mode theory formulation. *Journal of Lightwave Technology* 16:929–936.

Luyssaert B, Bienstman P, Vandersteegen P, Dumon P, Baets R (2005) Efficient nonadiabatic planar waveguide tapers. *Journal of Lightwave Technology* 23:2462–2468.

Maerz R (1995) *Integrated Optics: Design and Modelling*. London: Artech House.

Majewski A (1984) Numerical analysis of the multistep-index fiber having the uniaxial anisotropy. *Bulletin of the Polish Academy of Sciences, Technical Sciences* 32:709–714.

Majewski A (1991) *Theory and Design of Optical Fibres* (in Polish). Warsaw: Wydawnictwo Naukowo-Techniczne.

Majewski A, Sujecki S (1996a) Analiza wlasciwosci propagacyjnych swiatlowodów prostokatnych (in Polish). *Prace Naukowe Politechniki Warszawskiej* 110:5–35.

Majewski A, Sujecki S (1996b) Modes in rectangular fibres. *Optoelectronics Review* 4:45–50.

Majewski A, Sujecki S (1997) Rib lightguides and couplers. *Bull Polish Academy of Sciences: Technical Sciences* 45:145–150.

Majewski A, Sujecki S (1998) Polarisation characteristics of optical rib waveguides. *Optica Applicata* 28: 76–82.

Marcatili EAJ (1969) Dielectric rectangular waveguide and directional coupler for integrated optics. *Bell System Technical Journal* 48:2071–2102.

Marcatili EAJ (1974) Slab coupled waveguides. *Bell System Technical Journal* 53:645–674.

Marcatili EAJ (1985) Dielectric tapers with curved axes and no loss. *IEEE Journal of Quantum Electronics* 21:307–314.

Marcatili EAJ, Hardy AA (1988) The azimuthal effective-index method. *IEEE Journal of Quantum Electronics* 24:766–774.

Marcuse D (1972) *Light Transmission Optics*. New York, NY: Van Nostrand Reinhold.

Marcuse D (1974) *Theory of Dielectric Optical Waveguides*. San Diego, CA: Academic Press.

Melloni A, Cusmai G, Martinelli M (2004) Experimental confirmation of matched bends. *Optics Letters* 29:465–467.

Okamoto K (2006) *Fundamentals of Optical Waveguides*. London: Elsevier.

Okoshi T (1982) *Optical Fibres*. London: Academic Press.

Petermann K (1991) *Laser Diode Modulation and Noise*. London: Kluwer Academic Publishers.

Poli F, Cucinotta A, Selleri, S (2007) *Photonic Crystal Fibres*. Dordrecht: Springer-Verlag.

Ramaswamy W (1974) Strip-loaded film waveguides. *Bell System Technical Journal* 53:697–704.

Ravets S, Hoffman JE, Kordell PR, Wong-Campos JD, Rolston SL, Orozco LA (2013) Intermodal energy transfer in a tapered optical fiber: Optimizing transmission. *Journal of the Optical Society of America A: Optics, Image Science, and Vision* 30:2361–2371.

Riishede J, Mortensen NA, Laegsgaard J (2003) A “poor man’s approach” to modelling micro-structured optical fibres. *Journal of Optics A: Pure and Applied Optics* 5:534–538.

Rozzi T, Mongiardo M (1997) *Open Electromagnetic Waveguides*. London: IEE.

Snyder AW, Love, JD (1983) *Optical Waveguide Theory*. London: Chapman & Hall.

Soref RA, Schmidtchen J, Petermann K (1991) Large single mode rib waveguides in GeSi-Si and Si-on-SiO_2. *IEEE Journal of Quantum Electronics* 27:1971–1974.

Sujecki S (2001) New beam propagation algorithm for optical tapers. *International Journal of Electronics and Communications (AEU)* 55:185–190.

Sujecki S (2010) Arbitrary truncation order three-point finite difference method for optical waveguides with stepwise refractive index discontinuities. *Optics Letters* 35:4115–4117.

Sujecki S (2014) *Photonics Modelling and Design*. Boca Raton, FL: CRC Press.

Sujecki S, Benson TM, Sewell P, Kendall PC (1998) Novel vectorial analysis of optical waveguides. *Journal of Lightwave Technology* 16:1329–1335.

Tamir T (1975) *Integrated Optics*. Berlin: Springer-Verlag.

Vassallo C (1991) *Optical Waveguide Concepts*. Amsterdam: Elsevier.

Vassallo C (1997a) Difficulty with vectorial BPM. *Electronics Letters* 33:61–62.

Vassallo C (1997b) Mode solvers 1993–1995 optical mode solvers. *Optical and Quantum Electronics* 29:95–114.

Yamauchi J (2003) *Propagating Beam Analysis of Optical Waveguides*. Baldock, UK: Research Studies Press.

Yeh P, Yariv A, Marom E (1978) Theory of Bragg fiber. *Journal of the Optical Society of America* 68:1196–1201.

Zheng XH (1989) Understanding radiation from dielectric tapers. *Journal of the Optical Society of America A: Optics, Image Science, and Vision* 6:190–201.

Zolla F, Reneversez G, Nicolet A, Kuhlmev B, Guenneau S, Felbacq D. (2005) *Foundations of Photonic Crystal Fibres*. London: Imperial College Press.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

5

Heat Generation and Dissipation

5.1 Crystal Lattices 129
5.2 Crystal Vibrations and Phonons.. 131
5.3 Phonon Transport Equation 136
Interactions of Phonons among Themselves with Impurities and Boundaries • Phonon–Carrier Interaction
5.4 Boltzmann Equations for Carrier Transport 140
5.5 Macroscopic Balance Equations for Heat Transport in Semiconductors 142
Closure of the System of Macroscopic Evolution Equations
5.6 Energy-Transport Model and Its Numerical Discretization..... 146
Step 1 • Step 2
5.7 Thermal Properties of Graphene.. 152
Semiclassical Description • The Simulation Scheme • Numerical Results

Giovanni Mascali
and
Vittorio Romano

Thermal effects in solids are a manifestation of the vibrational motion of the molecules or atoms that make up the material. These phenomena can be described at several levels of accuracy. From a macroscopic point of view, a thermodynamical approach can be adopted, e.g., in the framework of linear irreversible thermodynamics (see e.g., [1]) or extended thermodynamics [2,3] (see e.g., [4–10,35]). More detailed models within solid-state physics introduce quasi-particles, called *phonons*, to take into account the energy transport in the crystal [11]. Phonons obey Bose–Einstein statistics and their evolution can be obtained, in a semiclassical context, by solving a system of Boltzmann equations.

In turn, thermal effects influence the electrical and optical properties of semiconductors. In particular, in micro- and nanoelectronics, the formation of hot spots and the high thermal power generation per unit area are major issues for an efficient design of semiconductor devices.

In this chapter, the basic concepts of phonon transport are introduced along with the derivation, from them, of the thermodynamical properties relevant to the electrical and optical features of a crystal lattice. Specific applications to graphene are presented.

5.1 Crystal Lattices

The structure of a solid is given by a periodic repetition of a set of atoms, called the *basis*, in the three-dimensional (3D) space [12]. In this way, a 3D crystal lattice is formed (see Figure 5.1) with a translational periodicity defined by three noncoplanar vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_3$, such that the crystal remains identical to itself

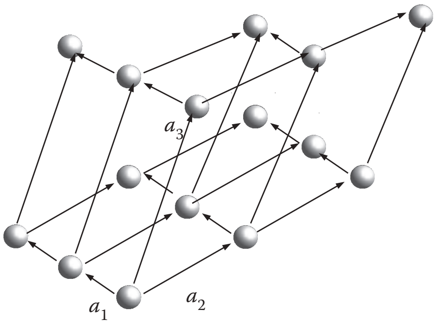


FIGURE 5.1 Example of crystal lattice.

if translated by a generic vector

$$\mathbf{T} = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3, \quad n_1, n_2, n_3 \in \mathbb{Z}.$$

The vectors $\boldsymbol{a}_1, \boldsymbol{a}_2, \boldsymbol{a}_2$ and the set

$$L = \{\mathbf{T} = n_1\boldsymbol{a}_1 + n_2\boldsymbol{a}_2 + n_3\boldsymbol{a}_3 \,:\, n_1, n_2, n_3 \in \mathbb{Z}\} \subset \mathbb{R}^3$$

are called *primitive vectors* and *Bravais lattice*, respectively. The *primitive cell* of L is a connected set $D \subset \mathbb{R}^3$ whose volume equals that of the parallelepiped spanned by the basis vectors, i.e., $vol(D) = \mathbf{a}_1 \cdot (\mathbf{a}_2 \times \mathbf{a}_3)^\dagger$, and fills the whole space if translated by the vectors belonging to L. In Figure 5.2, a primitive cell in a bidimensional case is depicted. In particular, the special primitive cell

$$D = \left\{ \mathbf{r} \in \mathbb{R}^3 \,:\, \mathbf{r} = \sum_{n=1}^{3} \alpha_n \boldsymbol{a}_n,\ \alpha_n \in \left[-\frac{1}{2}, \frac{1}{2}\right] \right\}$$

is called the *Wigner–Seitz cell*. The lengths of the primitive vectors are the *lattice constants*. For materials having the diamond structure, like those made of carbon atoms, the primitive cell is cubic and the three lattice constants are equal.

The *reciprocal lattice* L^* of L is defined by

$$L^* = \left\{\mathbf{G} = n_1\boldsymbol{a}_1^* + n_2\boldsymbol{a}_2^* + n_3\boldsymbol{a}_3^* \,:\, n_1, n_2, n_3 \in \mathbb{Z}\right\},$$

where the *primitive vectors* $\boldsymbol{a}_1^*, \boldsymbol{a}_2^*, \boldsymbol{a}_3^* \in \mathbb{R}^3$ constitute the dual basis given by

$$\boldsymbol{a}_1^* = 2\pi\frac{\boldsymbol{a}_2 \times \boldsymbol{a}_3}{\boldsymbol{a}_1 \cdot \boldsymbol{a}_2 \times \boldsymbol{a}_3}; \quad \boldsymbol{a}_2^* = 2\pi\frac{\boldsymbol{a}_3 \times \boldsymbol{a}_1}{\boldsymbol{a}_1 \cdot \boldsymbol{a}_2 \times \boldsymbol{a}_3}; \quad \boldsymbol{a}_3^* = 2\pi\frac{\boldsymbol{a}_1 \times \boldsymbol{a}_2}{\boldsymbol{a}_1 \cdot \boldsymbol{a}_2 \times \boldsymbol{a}_3}.$$

They satisfy the relations $\boldsymbol{a}_m \cdot \boldsymbol{a}_n^* = 2\pi\delta_{mn}$ with $m, n = 1, 2, 3$ and, therefore, $e^{i\boldsymbol{a}_m \cdot \boldsymbol{a}_n^*} = 1$.

[†] The symbol × denotes the vector product.

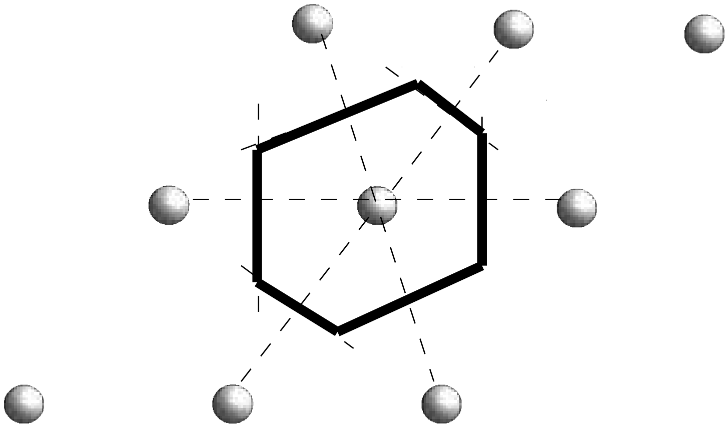


FIGURE 5.2 Primitive cell for a two-dimensional lattice.

According to the previous definitions, the volume of the reciprocal primitive cell is $\boldsymbol{a}_1^* \cdot (\boldsymbol{a}_2^* \times \boldsymbol{a}_3^*)$, while the Wigner–Seitz cell of the reciprocal lattice is

$$\mathcal{B} = \left\{ \mathbf{k} \in \mathbb{R}^3 \, : \, \mathbf{k} = \sum_{n=1}^{3} \beta_n \boldsymbol{a}_n^*, \; \beta_n \in \left[-\frac{1}{2}, \frac{1}{2} \right] \right\} \tag{5.1}$$

and is called the first *Brillouin zone*.

Strictly speaking, the previous picture is valid only for an infinite medium. If the sample has a finite size, boundary effects should be also considered. However, for the scope of this chapter, it is reasonable to assume as valid the bulk (that is of the infinite case) properties of the crystal.

5.2 Crystal Vibrations and Phonons

The most basic manifestation of the crystal temperature is the vibration of the atoms around their equilibrium positions represented by the points of the ideal Bravais lattice. In solids, under the adiabatic approximation, thermal motion of atoms around their equilibrium positions can be described as an ensemble of normal modes of vibration of the crystal lattice.

The interactions among the atoms depend on the specific chemical bond, e.g., ionic or covalent. As the first approximation, they can be considered elastic-type interactions. In order to grasp the main features without entering into the mathematical details, let us consider the simplified case of a chain of N equal particles of mass m connected by equal massless springs with elastic constant k (see Figure 5.3, left). The equilibrium separation is the lattice constant a. The particles can vibrate along the direction of the chain producing *longitudinal waves* or along the perpendicular directions producing *transversal waves*. For simplicity, we assume that the particles can vibrate only along the longitudinal direction. Let x_j be the displacement of the jth particle from its equilibrium position.

In order to deal with an infinite chain, cyclic boundary conditions are imposed

$$x_{j+N}(t) = x_j(t), \quad t > 0,$$

with t time and $N \in \mathbb{N}$ the period. Moreover, only the interactions of each atom with the nearest neighbors are retained. Under such hypotheses, one gets the following equations of motion:

$$m\ddot{x}_j = k(x_{j+1} + x_{j-1} - 2x_j), \quad j = 1, 2, \cdots, N. \tag{5.2}$$

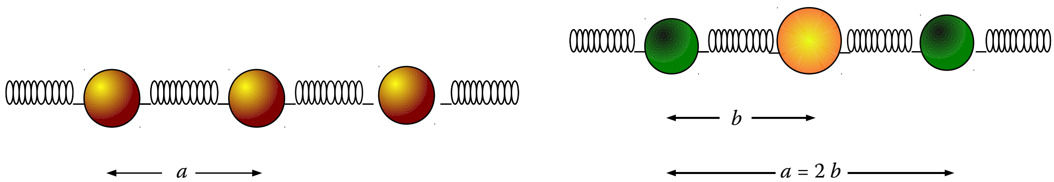


FIGURE 5.3 Schematic representation of a monoatomic (left) and a biatomic (right) linear chain.

We look for solutions of the type $x_j = \xi e^{i(qja-\omega(q)t)}$, where ξ is the amplitude, i is the imaginary unit, and q is the one-dimensional [1D] wave vector. After substituting in Equation 5.2, we have nontrivial solutions if and only if the angular frequency $\omega(q)$ satisfies the *dispersion relation*

$$\omega(q) = 2\sqrt{k/m}\left|\sin\left(\frac{qa}{2}\right)\right|.$$

By imposing the boundary conditions, we have

$$e^{iqNa} = 1 \Rightarrow q = \frac{2\pi}{Na}n = \frac{2\pi}{L}n, \quad n \in \mathbb{Z},$$

where $L = Na$ is the length of the portion of chain made of the N atoms.

The distance between two consecutive wave vectors is $\frac{2\pi}{L}$ and therefore the density of states, that is, the number of wave vectors per unit length, is $d(q) = \frac{L}{2\pi}$. Since under the transformation $q \mapsto q + l\frac{2\pi}{a}, l \in \mathbb{Z}$, the solution does not change, it is enough to limit q to the interval $[-\pi/a, \pi/a]$, which is the first Brillouin zone for the case considered. There are exactly N distinct modes. Of course, the solutions are periodic with the first Brillouin zone the interval of periodicity.

When the wavelength is much longer than the interatomic distance, $|qa| \ll 1$, the frequency is approximately given by the linear relation

$$\omega(q) = \sqrt{\frac{k}{m}}\,a|q|.$$

Note that in this limit, the waves are not dispersive and the group velocity (equal to the phase velocity) is $\sqrt{\frac{k}{m}}\,a$, which represents the sound velocity of the crystal in the continuum limit.

Let us now consider a linear chain composed of two types of different particles, with alternate masses m_1 and m_2 (see Figure 5.3, right). It is a simple model of a 1D crystal with two atoms in each unit cell. Let us assume that the lattice distance is $a = 2b$, b being the equilibrium distance between two adjacent atoms. The equations of motions are now

$$\begin{cases} m_1\ddot{x}_{2j+1} = k(x_{2j+2} + x_{2j} - 2x_{2j+1}), \\ m_2\ddot{x}_{2j} \;\;= k(x_{2j+1} + x_{2j-1} - 2x_{2j}). \end{cases}$$

We look for solutions of the form

$$x_{2j+1} = \xi\, e^{i((2j+1)bq-\omega(q)t)},$$

$$x_{2j} = \eta\, e^{i(2jbq-\omega(q)t)}.$$

The even and odd indices represent particles with mass m_1 and m_2, respectively. Proceeding as in the previous case, we obtain the dispersion relation

$$\omega^2(q) = k\left(\frac{1}{m_1} + \frac{1}{m_2}\right)\left[1 \mp \sqrt{1 - \frac{4m_1m_2}{(m_1+m_2)^2}\sin^2(bq)}\right].$$

In the limit $|qb| \ll 1$, it becomes

$$\omega^2(q) = \begin{cases} \dfrac{2k}{m_1+m_2}b^2q^2, & \text{acoustic branch,} \\ 2k\left(\dfrac{1}{m_1} + \dfrac{1}{m_2}\right), & \text{optical branch.} \end{cases}$$

Note that near $q = 0$, the acoustic branch is almost linear (*Debye approximation*), while the optical branch is almost flat (*Einstein approximation*). Therefore, near $q = 0$, the optical branch has a negligible group velocity and does not transport energy.

It is also instructive to analyze the ratio of the displacements

$$\frac{\xi}{\eta} = \frac{2k\cos(qb)}{2k - m_1\omega^2} \sim \begin{cases} 1 & \text{for acoustic modes} \\ -\dfrac{m_2}{m_1} & \text{for optical modes} \end{cases} \quad \text{as } q \mapsto 0.$$

Therefore, the acoustic modes produce oscillations along the same direction, while the optical modes produce oscillations in opposite directions.

In the 3D situation we have different directions of polarization with longitudinal and traversal oscillations, for both optical and acoustic modes. Each unit (primitive) cell generally consists of more than one atom, which is, therefore, characterized by four numbers, the first three, $\mathbf{n} = (n_1, n_2, n_3)$, with $n_i \in \mathbb{Z}$, $i = 1, 2, 3$, identify the cell to which it belongs, and the fourth, $s \in \mathbb{N}$, represents its number in the cell. Let $\mathbf{u}_s(\mathbf{n})$ be the displacements of the atoms in their vibrations. The index s runs from 1 to the number of atoms in each primitive cell. The Lagrangian of a crystal lattice, considered as a mechanical system of particles vibrating around their equilibrium positions, by retaining in the potential only up to the quadratic terms in the displacements (*harmonic approximation*), is given by

$$\mathcal{L} = \frac{1}{2}\sum_{\mathbf{n},s} m_s|\dot{\mathbf{u}}_s(\mathbf{n})|^2 - \frac{1}{2}\sum_{\mathbf{n},\mathbf{n}',s,s'} < \Lambda^{ss'}(\mathbf{n}-\mathbf{n}')\mathbf{u}_s(\mathbf{n}), \mathbf{u}_{s'}(\mathbf{n}') >, \tag{5.3}$$

where the dot means time derivative, $<,>$ denotes the standard scalar product, m_s is the mass of the atoms, and the matrices $\Lambda^{ss'}$ depend only on the differences $\mathbf{n} - \mathbf{n}'$, since interaction forces depend only on the relative position of the atoms. The $\Lambda^{ss'}$ satisfies the symmetry relation

$$\Lambda^{ss'}(\mathbf{n}) = \left(\Lambda^{s's}(-\mathbf{n})\right)^T, \tag{5.4}$$

along with some further relations, which derive from the fact that a parallel displacement or a rotation of the lattice as a whole gives rise to no forces (see [13]). In Equation 5.4, the superscript T indicates transposition.

The Lagrange equations of motion read

$$m_s\ddot{\mathbf{u}}_s = -\sum_{\mathbf{n}',s'} \Lambda^{ss'}(\mathbf{n}-\mathbf{n}')\mathbf{u}_{s'}(\mathbf{n}'). \tag{5.5}$$

Let us look for solutions in the form of plane waves

$$\mathbf{u}_s(\mathbf{n}) = \mathbf{e}_s(\mathbf{q})\exp[i(\mathbf{q}\cdot\mathbf{r}_\mathbf{n} - \omega t)], \tag{5.6}$$

with $\mathbf{r}_\mathbf{n}$ being the radius vector of any particular vertex of the cell $\mathbf{n}$, which can be used to define its position, ω being the angular frequency, $\mathbf{q}$ being the wave vector, and $\mathbf{e}_s$ being the polarization vector, which is the same for equivalent atoms in different cells. Substituting Equation 5.6 into Equation 5.5, after some algebra one gets

$$\sum_{s'} \tilde{\Lambda}^{ss'}(\mathbf{q})\,\mathbf{e}_{s'} - \omega^2 m_s \mathbf{e}_s = 0, \tag{5.7}$$

where

$$\tilde{\Lambda}^{ss'}(\mathbf{q}) := \sum_{\mathbf{n}} \Lambda^{ss'}(\mathbf{n}) \exp(-i\mathbf{q}\cdot\mathbf{r}_\mathbf{n}). \tag{5.8}$$

Equation 5.7 has nontrivial solutions if and only if the determinant of the $\nu \times \nu$ block matrix, whose generic block of indices s and s' is given by

$$\tilde{\Lambda}^{ss'}(\mathbf{q}) - \omega^2 m_{s'} \delta_{ss'} I, \tag{5.9}$$

vanishes, I being the 3×3 identity matrix. If ν is the number of atoms per cell, the order of the determinant is 3ν, therefore Equation 5.9 is an algebraic equation of degree 3ν in the unknown ω^2. Each solution determines ω as a function of the wave vector $\mathbf{q}$, which means that there are 3ν branches of such function

$$\omega = \omega_p(\mathbf{q}), \quad p = 1, \ldots, 3\nu,$$

which, as said, is called dispersion relation, p labeling the various branches. From Equations 5.8 and 5.4, it follows that

$$\tilde{\Lambda}^{ss'}(\mathbf{q}) = \tilde{\Lambda}^{s's}(-\mathbf{q}) = [\tilde{\Lambda}^{s's}(\mathbf{q})]^*, \tag{5.10}$$

where the star indicates complex conjugation. As a consequence, the $\tilde{\Lambda}^{ss'}(\mathbf{q})$ forms a Hermitian matrix, where the eigenvectors corresponding to different eigenvalues are orthogonal, i.e.,

$$\sum_{s=1}^{\nu} m_s \mathbf{u}_s^{(p)} \cdot \mathbf{u}_s^{(p')*} = 0, \quad \text{for} \quad p \neq p'. \tag{5.11}$$

Due to the symmetry of the mechanical equations of motion under time reversal, the dispersion relation has to be even

$$\omega(\mathbf{q}) = \omega(-\mathbf{q}).$$

Moreover, the wave vector $\mathbf{q}$ appears in Equation 5.6 only in the exponential factor $\exp(i\mathbf{q}\cdot\mathbf{r}_\mathbf{n})$, which does not change with the substitution

$$\mathbf{q} \to \mathbf{q} + \mathbf{G}, \qquad \mathbf{G} = n_1 \boldsymbol{a}_1^* + n_2 \boldsymbol{a}_2^* + n_3 \boldsymbol{a}_3^*, \; n_1, n_2, n_3 \in \mathbb{Z}, \tag{5.12}$$

where $\mathbf{G}$ is an arbitrary vector of the reciprocal lattice. Therefore, the wave vector is physically indeterminate, meaning that values of $\mathbf{q}$ differing by $\mathbf{G}$ are physically equivalent. This implies that the function $\omega(\mathbf{q})$ is periodic in the reciprocal lattice

$$\omega(\mathbf{q} + \mathbf{G}) = \omega(\mathbf{q}), \tag{5.13}$$

and can be restricted to a single cell. From a geometrical point of view, the dispersion relation $\omega(\mathbf{q})$ is represented by a finite four-dimensional hypersurface, with 3ν sheets, corresponding to the various branches, which may intersect.

As seen for the simple 1D chain, it is possible to ensure also in the 3D case that some of the branches of the vibrational spectrum, for wavelengths large with respect to the physical dimension of a unit cell, correspond to sound waves in the crystal and for them the dispersion relation is a first-order homogeneous function of the components of the wave vector, for $\mathbf{q}$ small enough. These waves are called acoustic waves.

In lattices with more than one atom per cell, there are $3(\nu - 1)$ further types of waves, whose frequency does not vanish at $\mathbf{q} = 0$, but tends to a finite constant. Such vibrations are called optical vibrations, and for them the atoms in each cell are in relative motion [13].

For example, in silicon and gallium arsenide there are two atoms per cell and, therefore, six types of elastic waves: three acoustic modes and three optical modes. Of the three acoustic modes, one is longitudinal (LA), that is the atoms are displaced in the direction of the wave propagation, and the other two are transverse (TA), in which the atoms are displaced in a transverse direction. The same is true for the optical phonons (LO and TO). The transversal modes are therefore doubly degenerate. In Figure 5.4, we represent the dispersion relations in Si along the main crystallographic directions.

What the text mentioned earlier is strictly valid only in the harmonic approximation, in which the various monochromatic waves freely propagate through the crystal without interacting. Higher-order terms are taken into account as various processes of decay and scattering among these waves. Their description is rather complex and a simplified approach based on a relaxation time approximation is usually adopted in the device simulations.

The picture we have presented thus far is based on classical mechanics. However, the applications in micro- and nanoelectronics need a quantum approach to the lattice vibrations. Since the $\widetilde{\Lambda}^{ss'}(\mathbf{q})$ is a Hermitian matrix, it is possible to write the classical equations of motion in canonical coordinates, decoupling them in independent harmonic oscillators [12,14]. As a consequence, the analogous quantum Hamiltonian is the sum of the Hamiltonians of $3\nu N$ independent quantum oscillators [12,14], with N being the total number of cells, each of them having quantized energy levels given by

$$\epsilon_n = \hbar\omega\left(n + \frac{1}{2}\right), \qquad n = 0, 1, 2, \ldots,$$

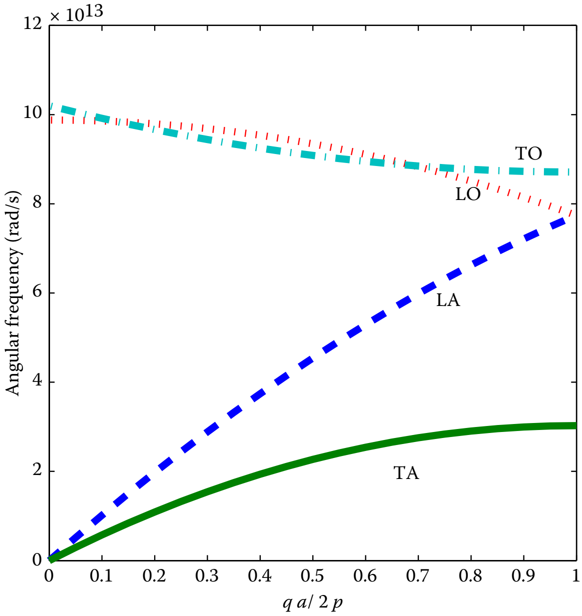


FIGURE 5.4 Phonon dispersion relations in silicon along the <100> crystallographic directions.

From the particle point of view, each quantum state can be considered as formed by n fictitious particles called *phonons* having energy $\hbar\omega$ in analogy with photons. Here $\hbar$ is the reduced Planck constant. The processes, which change the state from the energy level ϵ_n to ϵ_{n+1} (respectively, ϵ_{n-1}) can be interpreted as the *creation* (respectively, *destruction*) of a phonon.

The quasimomentum is associated with phonons:

$$\mathbf{p} = \hbar\mathbf{q}.$$

It is almost analogous to the ordinary momentum, with the important difference that it is defined up to an arbitrary constant vector $\hbar\mathbf{G}$, with $\mathbf{G}$ belonging to the reciprocal lattice. The phonon velocity is given by the group velocity of the corresponding classical waves, $\mathbf{v} = \frac{\partial\omega}{\partial\mathbf{q}}$, which can be rewritten in the form

$$\mathbf{v} = \frac{\partial\epsilon(\mathbf{p})}{\partial\mathbf{p}}.$$

In the approximation successive to the harmonic one, various elastic and inelastic phonon collision processes occur, which tends to drive the system toward equilibrium (see below for more details). In these processes, the energy and momentum conservation laws must be satisfied, the latter up to an additive vector of the form $\hbar\mathbf{G}$

$$\sum\mathbf{p} = \sum\mathbf{p}' + \hbar\mathbf{G},$$

where $\mathbf{p}$ and $\mathbf{p}'$ are the momenta of the phonons before and after the collision.

An arbitrary number of identical phonons can be created simultaneously in the lattice, which means that the phonon gas obeys the Bose statistics. Since the total number of phonons is determined by the equilibrium conditions, the phonon chemical potential is zero [13]. Therefore, the equilibrium occupation number $\bar{g}(\mathbf{p})$ of a quantum state with momentum $\mathbf{p}$ and energy $\epsilon(\mathbf{p})$ is determined by the Planck distribution function

$$\bar{g}(\mathbf{p}) = 1/(\exp(\epsilon(\mathbf{p})/T_{\mathrm{L}}) - 1),$$

where T_{L} is the lattice temperature.

5.3 Phonon Transport Equation

If the typical phonon wavelength is much shorter than the phonon mean free path, phonons can be treated as semiclassical particles that obey, as said, the Bose statistics. In this case, the state of the phonon gas can be described through the distribution function $g^p(\mathbf{q})$, in which time evolution is governed by the Boltzmann–Peierls (BP) equation

$$\frac{\partial g^p}{\partial t} + \mathbf{v}_p \cdot \nabla_{\mathbf{x}} g^p = \sum_\eta C^p_\eta(g^p) + C^p_{(\mathrm{oth})}(g^p), \tag{5.14}$$

the index p labeling the several branches. The right-hand side is the collision operator. The term $\sum_\eta C^p_\eta(g^p)$ represents the phonon–phonon collisions, and collisions of phonons with impurities, boundaries, and defects, the index η labeling the various types of scatterings. The other contribution in the collision operator is due to the interaction of phonons with other particles, e.g., electrons and photons. In addition to the phonon–phonon collisions, in the next sections we describe only the details of the phonon–electron collisions because they are of major interest in this chapter.

5.3.1 Interactions of Phonons among Themselves with Impurities and Boundaries

More specifically, phonon interaction processes, which do not involve charge carriers or photons, can be divided into intrinsic ones, which arise from the anharmonicity of the interatomic forces, and extrinsic ones, due to phonon scatterings at the boundaries of the crystal and at various types of crystal defects and imperfections [11]. The anharmonic scatterings can be *normal* processes (N-processes), in which the phonon total momentum after a collision is conserved, and *umklapp* processes (U-processes) in which the total momentum changes by a reciprocal lattice vector after a collision. On the other hand, all extrinsic processes do not conserve the total momentum after a collision and together with the umklapp processes are usually called resistive processes [15].

The anharmonic interactions involve an optical phonon, which decays into a transversal acoustic phonon and a longitudinal phonon belonging either to the acoustic or to the optical branch [16]. The expression of the corresponding collision operator in the BP equation for the optical phonons, first given by Klemens [17], reads

$$C^{p_O}_{p_O\leftrightarrow TA+p'} = \frac{\gamma^2 y^{p_O}}{3\pi^2 \rho v_s^2}\int_B \epsilon_{p_O}\omega'_{p'}\,\omega_{TA}(\mathbf{q}'')\delta(\omega_{p_O}-\omega'_{p'}-\omega_{TA}(\mathbf{q}''))$$

$$\times\left[\left(g^{p_O}(\mathbf{q})+1\right)g^{p'}(\mathbf{q}')g^{TA}(\mathbf{q}'')-g^{p_O}(\mathbf{q})\left(g^{p'}(\mathbf{q}')+1\right)\left(g^{TA}(\mathbf{q}'')+1\right)\right]\mathrm{d}\mathbf{q}', \qquad (5.15)$$

where $y^p = \frac{1}{(2\pi)^3}$ is the phonon density of states, $p_O = LO, TO$, $p' = LO, LA$, $\mathbf{q}'' = \mathbf{q}-\mathbf{q}'+\mathbf{G}$ (stemming from the generalized momentum conservation), γ is the Grüneisen parameter, v_s is the sound velocity, ρ is the material density, and δ is the Dirac delta function [16]. Analogous are the expressions for the corresponding collision operators appearing in the BP equation relative to the acoustic phonons. Relaxation time approximations for the operator Equation 5.15 can be found in [18].

For the interactions of the acoustic phonons with boundaries and impurities, the relaxation time approximation is commonly used for the operators, which reads [18]

$$C^{p_A}_{\eta} = -\frac{g^{p_A}-g^{p_A(0)}}{\tau^{\eta}_{p}},$$

where $\eta = B, I$, with B and I, respectively, the scattering with boundaries and impurities. The relaxation times are given by

$$[\tau^B_{p_A}]^{-1} = \frac{v^{p_A}}{d}, \quad p_A = LA, TA,$$

with d being the effective diameter of the sample

$$[\tau^I_{p_A}]^{-1} = \frac{V\Gamma}{4\pi\hbar^4 (v_s^{p_A})^3}\epsilon^4_{p_A}, \quad p_A = LA, TA,$$

where, as impurities, only the isolated defects of mass different from those of the host have been considered. In this case, V is the volume per atom and Γ is the mass-fluctuation phonon scattering parameter for a single element made up of several naturally occurring isotopes and is given by

$$\Gamma = \sum_i c_i\left[\frac{m_i-\bar{m}}{\bar{m}}\right]^2, \quad \bar{m} = \sum_i c_i m_i,$$

with m_i the atomic mass of the ith isotope and c_i the fractional atomic natural abundance.

5.3.2 Phonon–Carrier Interaction

Within the semiclassical approximation for electrons, we indicate by f^A the distribution function of the charge carriers occupying the valley labeled by A. The valleys are the neighborhoods of states centered in the lowest minima of the lowest conduction bands and in the highest maxima of the highest valence bands [12] in the carrier wave vector space. In fact, the electrons are occupying those states which mainly contribute to the charge transport in semiconductors. The contribution of the electrons in the valence bands can be conveniently treated in terms of the presence of particles called holes (missing electrons), which have electric charge, crystal momentum, and energy opposite to those of the missing electrons in the valence band considered [7,8,19]. The electron/hole state is characterized by the index A of the band they occupy, by their crystal momentum, $\hbar\mathbf{k}$, with $\mathbf{k}$ wave vector, and spin state, which can be up or down. An electron/hole population for each neighborhood of the lowest/highest minima/maxima of the lowest/highest conduction/valence bands has to be considered.

In order to describe phonon interactions with carriers, we consider the particular case of silicon and we limit ourselves to the electrons in the conduction valleys. More general cases involving compound semiconductors and interactions with holes can be treated similarly [8,20,24].

As regards the phonon dispersion relations in silicon, the following isotropic quadratic approximations can be used (see Figure 5.4):

$$\epsilon_p = \epsilon_0^p + \hbar v_s^p\,|\mathbf{q}| + \hbar c^p\,|\mathbf{q}|^2\,,\ |\mathbf{q}| \in \left[0, \frac{2\pi}{a}\right],\ p = LA,\ TA,\ LO,\ TO, \tag{5.16}$$

where the first Brillouin zone is approximated by a sphere of radius $\frac{2\pi}{a}$, with a being the silicon lattice constant; A, O, L, and T stand for acoustic, optical, longitudinal, and transversal, respectively; and the coefficients in Equation 5.16 can be found in [21].

For silicon, electrons in conduction bands, which mainly contribute to the charge transport, are those in the valleys centered in the six-band minima, near the equivalent X symmetry points, of the lowest conduction band; see Figure 5.5 for an illustration of the first Brillouin zone. In the Kane ellipsoidal approximation [19], in each valley the dependence of the electron energy $\mathcal{E}$ (here and whenever possible we omit the valley index, for simplicity) on the wave vector $\mathbf{k}$ is given by

$$\mathcal{E} = \frac{\hbar^2}{2}\gamma(\mathcal{E})\sum_{i=1}^{3}\frac{(k_i-\kappa_i)^2}{m_i}, \qquad \gamma(\mathcal{E}) := \left(1+\alpha\mathcal{E}\right)^{-1},$$

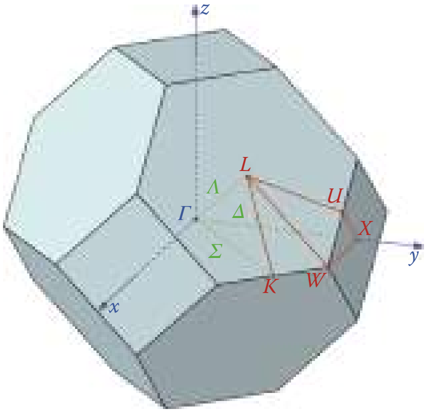


FIGURE 5.5 First Brillouin zone of silicon.

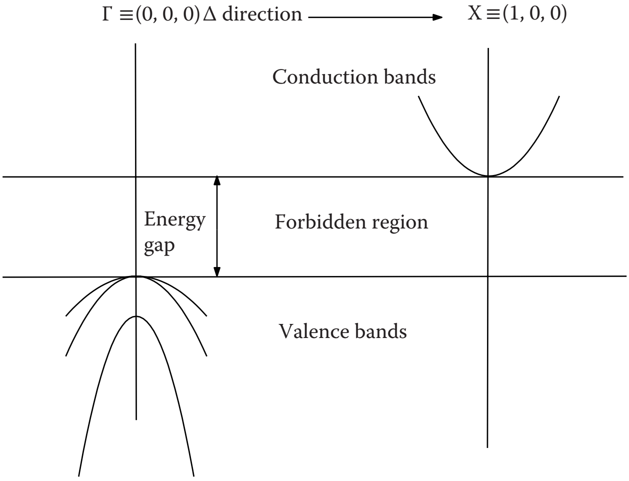


FIGURE 5.6 Schematic representation of the energy bands in silicon.

where the energy is referred to the minimum of the valley, κ_i, $i = 1, 2, 3$ are the coordinates of the minimum, α is the nonparabolicity parameter, and m_i^{-1}, $i = 1, 2, 3$ are the eigenvalues of the inverse effective valley mass tensor. Both the nonparabolicity parameter and the effective masses are temperature dependent [21]. The first Brillouin zone is usually extended to all $\mathbb{R}^3$. Figure 5.6 shows a schematic representation of the energy bands in silicon.

Since the electrons in the valleys, which correspond to minima lying on the same principal axis, have the same mass tensor [19], they can be considered a unique population. Therefore, one can consider three electron populations, depending on which axis the minimum of the valley, where the electrons live, belongs to. We label the populations by $e = x, y, z$.

The general form of the collision operators relative to the interactions of phonons with electrons, contributing to the term $C^p_{(\mathrm{oth})}$ in Equation 5.14, is [4,22]

$$C^p_{\iota, ee'}(g^p, f^e, f^{e'}) = \int_{\mathbb{R}^3}\int_{\mathbb{R}^3} \left[w^{ee'+}_{\iota,p}(\mathbf{k}, \mathbf{k}', \mathbf{q})\kappa_1(g^p, f^e, f^{e'})\right] d\mathbf{k}'\, d\mathbf{k},$$

with

$$\kappa_1(g^p, f^e, f^{e'}) := \left(g^p + 1\right) f^{e'}(\mathbf{k}') - g^p f^e(\mathbf{k}),$$

$$w^{ee'\pm}_{\iota,p}(\mathbf{k}, \mathbf{k}', \mathbf{q}) := s^{ee'}_{\iota,p}(\mathbf{q})\,\delta\big(\mathcal{E}_{e'}(\mathbf{k}') - \mathcal{E}_e(\mathbf{k}) \mp \epsilon_p\big)\,\delta\big(\mathbf{k}' - \mathbf{k} \mp \mathbf{q} + \mathbf{G}\big).$$

The phonon–electron scatterings can be intravalley ($\iota = sv$, $e = e'$, $p = LA, TA$),[†] which involve only acoustic phonons and conserve the total energy and momentum, or intervalley [19]. The latter, which pertain both to the acoustic and the optical phonons, in their turn can be distinguished in g-type ones ($\iota = dv$, $e = e'$, $p = LA, TA, LO$), which take electrons between equivalent valleys or valleys that have the transversal mass in the same direction, and f-type ones ($\iota = dv$, $e \neq e'$, $p = TA, LA/LO, TO$), taking electrons to nonequivalent valleys. Both of them are umklapp processes, which do not conserve the total momentum and involve a reciprocal lattice vector **G** [21]. The phonon wave vectors interested in the intervalley transitions remain very close to the vector joining the minima of the initial and final valleys and therefore they are usually taken as constant and, after reduction to the first Brillouin zone, equal to $\frac{2\pi}{a}(1, 0.15, 0.15)$ for f-type scatterings and to $\frac{2\pi}{a}(0.3, 0, 0)$ for g-type scatterings. The scattering functions $s^{ee'}_{\iota,p}$ are given by [19]

[†] *sv* and *dv*, respectively, stand for same valley and different valleys.

$$
\begin{aligned}
s^{ee}_{sv,LO} &= s^{ee}_{sv,TO} = 0, \quad s^{ee'}_{sv,p} = 0, \; e = x, y, z, \; e \neq e', \; p = LA, TA, LO, TO,\\
s^{ee}_{sv,p} &= \frac{\hbar\,|\mathbf{q}|^2 D^2_{a,p}}{8\,\pi^2 \rho\, \epsilon_p} I^2(|\mathbf{q}|), \; p = LA, TA, \; e = x, y, z,\\
s^{ee}_{dv,p} &= \frac{\hbar(\Delta^{ee}_p)^2}{8\,\pi^2 \rho\, \bar{\epsilon}^{ee}_p}, \; p = LA, TA, LO, \; e = x, y, z,\\
s^{ee'}_{dv,p} &= \frac{\hbar(\Delta^{ee'}_p)^2}{4\,\pi^2 \rho\, \bar{\epsilon}^{ee'}_p}, \; p = TA, LA/LO, TO, \qquad e \neq e', \; e, e' = x, y, z,
\end{aligned}
$$

where $D_{a,p}$, $p = LA, TA$ are the acoustic deformation potentials, $I(|\mathbf{q}|)$ is the overlap integral, ρ is the material density, $\Delta^{ee'}_p$, $e, e' = x, y, z$, $p = LA, TA, LO, TO$ are the intervalley deformation potentials, and $\bar{\epsilon}^{ee'}_p$, $e, e' = x, y, z$, $p = LA, TA, LO, TO$ are the phonon energies involved in the f-type and g-type transitions. The values of the various parameters and the expression of the overlap integral appearing in the above-mentioned scattering functions can be found in [21].

5.4 Boltzmann Equations for Carrier Transport

Before macroscopic balance equations for heat transport in semiconductors can be written, it is necessary to write down the evolution equations for the carrier distribution functions, which read

$$
\begin{aligned}
&\frac{\partial f^A}{\partial t} + \mathbf{v}_A \cdot \nabla_{\mathbf{x}} f^A + \frac{q_A}{\hbar}\,\mathbf{E} \cdot \nabla_{\mathbf{k}} f^A = C^A_{im}(f^A) + \sum_{\iota, A', p} C^{AA'}_{\iota,p}(f^A, f^{A'}, g^p) + \sum_{A'} \mathcal{I}[f^A, f^{A'}],\\
&\Delta(\varepsilon_{\mathrm{d}}\, \phi) = q\big(n(\mathbf{x}, t) - p(\mathbf{x}, t) + N_{\mathrm{a}}(\mathbf{x}) - N_{\mathrm{d}}(\mathbf{x})\big), \qquad (5.17)
\end{aligned}
$$

where q_A is the charge of the carriers in the Ath valley and q its absolute value, $\mathbf{v}_A = \frac{1}{\hbar}\nabla_{\mathbf{k}}\mathcal{E}_A$ is the carrier group velocity, ι labels the type of scattering between carriers and phonons, ε_{d} is the material permittivity, ϕ and **E** are the electric potential and field, N_{a} and N_{d} are the acceptor and donor concentrations, n is the total electron density and p is the total hole density. The Boltzmann Equations 5.17 are coupled among them through the Poisson Equation 5.18 and some of the collision operators that appear at the right-hand side of Equation 5.17.

The operator C^A_{im} [19] takes into account scatterings between carriers and impurities, which are elastic and intravalley. Its form is

$$
C^A_{im}(f^A) = \int_{\mathbb{R}^3} \left[w^A_{im}(\mathbf{k}', \mathbf{k}) f^A(\mathbf{k}') - w^A_{im}(\mathbf{k}, \mathbf{k}') f^A(\mathbf{k})\right] d\mathbf{k}',
$$

the impurity scattering transition rate being given by

$$
w^A_{im}(\mathbf{k}, \mathbf{k}') \;=\; \mathcal{K}^{(im)}_A \frac{1}{\left[|\mathbf{k} - \mathbf{k}'|^2 + \lambda^2_A\right]^2}\, \delta(\mathcal{E}_A(\mathbf{k}') - \mathcal{E}_A(\mathbf{k})),
$$

with $\lambda_{_A} = \sqrt{\frac{(N_{\mathrm{a/d}})q^2}{\varepsilon_{\mathrm{d}} k_B T_{\mathrm{L}}}}$ is the inverse Debye length and $\mathcal{K}_A^{(im)} = \frac{Z^2_{\mathrm{a/d}} N_{\mathrm{a/d}}\, q^4}{4\pi\hbar\varepsilon_{\mathrm{d}}^2}$, where Z is the impurity charge number and k_B is the Boltzmann constant. In the calculation of $\lambda_{_A}$ and $\mathcal{K}_A^{(im)}$, N_{a}, and Z_{d}, respectively, N_{d} and Z_{d} have to be taken according to whether interactions with donors or impurities are considered.

The collision operators $C_{\iota,p}^{AA'}(f^A, f^{A'}, g^p)$ describe interactions between carriers and phonons and [22]

$$C_{\iota p}^{A\,A'}(f^A, f^{A'}, g^p) = \int_{\mathcal{B}} \int_{\mathbb{R}^3} \big[w_{\iota p}^{AA'+}(\mathbf{k}, \mathbf{k}', \mathbf{q}) \kappa^1_{AA'}(g_p, f^A, f^{A'})$$

$$+ w_{\iota p}^{AA'-}(\mathbf{k}, \mathbf{k}', \mathbf{q}) \kappa^2_{AA'}(g_p, f^A, f^{A'}) \big] \, d\mathbf{k}' \, d\mathbf{q},$$

where

$$\kappa^2_{AA'}(g^p, f^A, f^{A'}) := g^p f^{A'}(\mathbf{k}') - \left(g^p + 1 \right) f^A(\mathbf{k}).$$

Earlier we reported the expressions of the κ that are valid in the case when both the initial and the final valleys belong to conduction bands and the expressions in the interband case are analogous [23]. The scattering functions $s_{\iota,p}^{AA'}$ are characteristic of the type of interaction of carriers, for example with acoustic and nonpolar optical phonons for elemental semiconductors like Si (see the previous section) and Ge, and also with polar optical phonons for compound semiconductors, like GaAs and SiC. For their expressions, we refer the interested reader to [19,20,24].

In the generation–recombination collision operator $\sum_{A'} \mathcal{I}[f^A, f^{A'}]$, if A is a conduction valley the sum is extended to the valence valleys and vice versa. It describes the processes in which electron–hole pairs are created by exciting an electron from the valence to the conduction band and vice versa, electrons and holes from the conduction and the valence band recombine and are annihilated. These processes are of several types; here we consider only the most important ones.

In the direct band-to-band Auger mechanism, three carriers are involved. In generation, an electron–hole pair is created at the expense of the energy of a highly energetic electron or hole. The opposite process transfers the excess energy to an electron or hole.

The Shockley–Read–Hall (SRH) generation/recombination occurs by phonon absorption/emission. This process is trap-assisted utilizing defects with energy levels within the semiconductor band-gap.

The Auger and SRH operators, in the relaxation time approximation and in the simpler case of a single conduction (n) and valence (h) valley, read [25]

$$\mathcal{I}[f^A, f^{\bar{A}}] = -\left(\Gamma_n n + \Gamma_h p\right)\left(\widetilde{n} f^A - {n_i}^2 \mathcal{M}_A\right) - \frac{\widetilde{n} f^A - {n_i}^2 \mathcal{M}_A}{\tau_h(n + n_i) + \tau_n(p + n_i)}, \tag{5.18}$$

where $\bar{A} = h$ if $A = e$ and vice versa, $\widetilde{n} = p$ or n according to whether the Ath valley is populated by electrons or holes, the $\mathcal{M}_A$'s are the Maxwellians normalized to unit, Γ_n and Γ_h are the Auger electron and hole constant coefficients, τ_n and τ_h are the carrier lifetimes, and n_i is the intrinsic concentration [19].

In direct band-gap semiconductors, the generation/recombination processes can also involve photons. In fact, in recombinations there can be spontaneous emissions of photons, which are of fundamental importance in light-emitting optoelectronic devices [26]. Similarly, in generations, the transition energy can be provided by photons, which is the key mechanism in photodetectors and other electroabsorption devices. The operator corresponding to photon-assisted generation/recombination processes reads [26,27]

$$\int w_{\uparrow}(\mathbf{k}', \mathbf{k}) f_h(\mathbf{k}')(1 - f_n(\mathbf{k})) - w_{\downarrow}(\mathbf{k}, \mathbf{k}') f_n(\mathbf{k})(1 - f_h(\mathbf{k}')) d\mathbf{k}'$$

for the electrons in the conduction band, f_n being their distribution function, while f_h is the distribution function for the electrons in the valence band. In the above formula

$$w_{\uparrow}(\mathbf{k}',\mathbf{k}) = \frac{\hbar\pi q^2}{6 m_e \omega_{\mathrm{f}} \sqrt{\varepsilon_{\mathrm{d}}}\left(\sqrt{\varepsilon_{\mathrm{d}}} + \omega_{\mathrm{f}} \frac{d\sqrt{\varepsilon_{\mathrm{d}}}}{d\omega_{\mathrm{f}}}\right)} \delta(\mathbf{k}' - \mathbf{k})\delta(\mathcal{E}_n(\mathbf{k}) - \mathcal{E}_h(\mathbf{k}) - \hbar\omega_{\mathrm{f}})\, \frac{g_{\mathrm{f}}}{V_{\mathrm{f}}},$$

$$w_{\downarrow}(\mathbf{k}',\mathbf{k}) = \frac{1 + g_{\mathrm{f}}}{g_{\mathrm{f}}} w_{\uparrow}(\mathbf{k}',\mathbf{k}),$$

where $\tilde{E}$ is a material-dependent parameter by having the dimensions of an energy [27], m_e is the free electron mass, ω_{f} is the photon frequency, g_{f} is the photon distribution function, and V_{f} is the photon mode volume. An analogous expression holds for holes.

5.5 Macroscopic Balance Equations for Heat Transport in Semiconductors

From a macroscopic point of view, thermal effects are described, within the standard linear irreversible thermodynamics, by the heat conduction equation with a Fourier-like law for the heat flux

$$\rho c_{\mathrm{V}} \frac{\partial}{\partial t} T - \mathrm{div}\,[K(T)\nabla T] = H, \tag{5.19}$$

where T is the common local equilibrium temperature of electrons, holes, and lattice, ρ is the density of the material, c_{V} is the specific heat at constant volume, $K(T)$ is the thermal conductivity, and H is the thermal production term. The validity of the Fourier law is rather questionable in the case of solids at low temperature in the ballistic case [2,3] but is valid if the resistive processes dominate. The dependence of K on T is usually deduced from the experimental data and can depend on the characteristic scales involved in the problem. For bulk materials, accurate analytical formulas are known [28,29]. For nanoscale devices, the surface scattering plays a major role and significant departures from the bulk case are observed [28].

The same considerations hold for the production H, which contains several effects, e.g., the Joule and Thomson ones in addition to radiative recombination and optical generation. The correct modeling of H involves several phenomenological parameters. A review of the heuristic approaches can be found in [1] where an in-depth discussion of the drift-diffusion models with varying (but equal for electrons, holes, and lattice) temperature is performed. A systematic approach to get constitutive relations directly from the phonon transport equations will be adopted in the next section of this chapter. However, in order to provide the reader with a physical intuition about what will be deduced, we first review the standard energy model based on the drift-diffusion approximation.

The standard drift-diffusion model augmented with the heat conduction equation for the common temperature T, the so-called *energy model*, is given by the following set of evolution equations:

$$\frac{\partial n}{\partial t} - \frac{1}{q}\nabla \cdot \mathbf{J}_n = -R + G, \quad \mathbf{J}_n = \mu_n k_B T \nabla n - q\, n\mu_n \nabla\phi + D_n \nabla T, \tag{5.20}$$

$$\frac{\partial p}{\partial t} + \frac{1}{q}\nabla \cdot \mathbf{J}_p = -R + G, \quad \mathbf{J}_p = -\mu_p k_B T \nabla p - q\, p\mu_p \nabla\phi - D_p \nabla T, \tag{5.21}$$

$$\rho c_{\mathrm{V}} \frac{\partial T}{\partial t} - \nabla \cdot (K(T)\nabla T) = H \tag{5.22}$$

$$\nabla\left(\varepsilon_{\mathrm{d}} \nabla\phi\right) = q(n - p - N_D + N_A). \tag{5.23}$$

In the expressions of the electron and hole currents, $\mathbf{J}_n$ and $\mathbf{J}_p$, μ_n and μ_p are the mobilities that depend on the electric field and other parameters [29], while the last terms express the *Seebeck effect* (a temperature gradient induces an electric current), D_n and D_p being thermal coefficients. They are also written as $D_n = n\mu_n P_n$ and $D_p = p\mu_p P_p$ with P_n and P_p playing the role of thermoelectric powers.

The recombination and generation terms, R and G (see [29] for a complete review), can be directly deduced from Equations 5.18 and 5.19 by integration over the $\mathbf{k}$ space.

Regarding the production H, in [1], the following expression has been proposed, in the steady-state case:

$$H = \underbrace{\frac{|\mathbf{J}_n|^2}{q\mu_n n} + \frac{|\mathbf{J}_p|^2}{q\mu_p p}}_{\text{Joule effect}} + T\left(\frac{k_B}{q}\log\frac{n}{n_i} + P_n\right)\nabla\cdot\mathbf{J}_n - T\left(\frac{k_B}{q}\log\frac{p}{n_i} + P_p\right)\nabla\cdot\mathbf{J}_p$$

$$+ \underbrace{T\left(\mathbf{J}_n\cdot\nabla P_n + \mathbf{J}_p\cdot\nabla P_p\right)}_{\text{Thomson effect}} + H_{\text{rad}}. \tag{5.24}$$

The two intermediate terms can be interpreted by observing that in the steady state $\nabla\cdot\mathbf{J}_n = -\nabla\cdot\mathbf{J}_p = q(R-G)$. Thereby, the intermediate terms are the recombination-generation heat generation. The last term H_{rad} is the radiative contribution.

Equations 5.20 through 5.22 and in particular expression Equation 5.24 contain several coefficients to be prescribed, e.g., the thermoelectric powers. In the rest of the section, we proceed with a systematic way to tackle the question of determining macroscopic balance equations.

These macroscopic models can be constructed starting from the Boltzmann–Peierls system for phonons and carriers by taking suitable moments of the distribution functions. In particular, the following functions of the phonon and carrier wave vectors $\{\psi_p(\mathbf{q})\} := \{\epsilon_p, \epsilon_p\mathbf{v}^p\}$ and $\{\psi_A(\mathbf{k})\} := \{1, \mathbf{v}^A, \mathcal{E}_A, \mathcal{E}_A\mathbf{v}^A\}$, can be considered, to which the macroscopic state variables below correspond:

$$\begin{pmatrix} W^p \\ \mathbf{Q}^p \end{pmatrix} = y^p \int_{\mathcal{B}} \begin{pmatrix} \epsilon_p \\ \epsilon_p\mathbf{v}^p \end{pmatrix} g^p d\mathbf{q}, \tag{5.25}$$

$$\begin{pmatrix} n_A \\ W^A \end{pmatrix} = y^A \int_{\mathbb{R}^3} \begin{pmatrix} 1 \\ \frac{\mathcal{E}_A}{n_A} \end{pmatrix} f^A d\mathbf{k}, \qquad \begin{pmatrix} \mathbf{V}^A \\ \mathbf{S}^A \end{pmatrix} = \frac{y^A}{n_A} \int_{\mathbb{R}^3} \begin{pmatrix} \mathbf{v}^A \\ \mathcal{E}_A\mathbf{v}^A \end{pmatrix} f^A d\mathbf{k}, \tag{5.26}$$

which, respectively, are the phonon energy and energy flux densities, and the carrier density, average energy, velocity, and energy flux per carrier in the valley A, $y^A = \frac{2}{(2\pi)^3}$ being the carrier density of states. The index p runs over all the phonon branches and A over the carrier valleys. The phonon momentum densities are given by

$$\mathbf{P}^p = y^p \int_{\mathcal{B}} \hbar\mathbf{q}\, g^p \mathrm{d}\mathbf{q},$$

where p runs over the phonon branches. Here, the minimal number of moments necessary for describing the thermal energy transport have been chosen, but this number, if required by the physical problem under study, can be easily extended to cover, e.g., an arbitrary number of scalar and vector moments both for phonons and carriers, by taking into account higher energy powers [5,15,30]. These further moments are needed to deal with situations very far from equilibrium.

The evolution equations for the state variables 5.25 and 5.26 can be obtained directly from the Boltzmann equations by integration:

$$\frac{\partial}{\partial t}\begin{pmatrix} W^p \\ Q_i^p \end{pmatrix} + \sum_j \frac{\partial}{\partial x_j}\begin{pmatrix} Q_j^p \\ T_{ij}^p \end{pmatrix} = \begin{pmatrix} C_{W^p} \\ C_{Q_i^p} \end{pmatrix}, \qquad i = 1, 2, 3. \tag{5.27}$$

$$\frac{\partial}{\partial t}\begin{pmatrix} n_A \\ n_A W^A \end{pmatrix} + \sum_j \frac{\partial}{\partial x_j}\left[n_A \begin{pmatrix} V_j^A \\ S_j^A \end{pmatrix}\right] - n_A\, q_A \begin{pmatrix} 0 \\ \sum_j E_j\, V_j^A \end{pmatrix} = n_A \begin{pmatrix} C_{n_A} \\ C_{W_A} \end{pmatrix}, \tag{5.28}$$

$$\frac{\partial}{\partial t}\left[n_A \begin{pmatrix} V_i^A \\ S_i^A \end{pmatrix}\right] + \sum_j \frac{\partial}{\partial x_j}\left[n_A \begin{pmatrix} F_{ij}^{A(0)} \\ F_{ij}^{A(1)} \end{pmatrix}\right]$$

$$- n_A\, q_A \sum_j E_j \begin{pmatrix} G_{ij}^{A(0)} \\ G_{ij}^{A(1)} \end{pmatrix} = n_A \begin{pmatrix} C_{V_i^A} \\ C_{S_i^A} \end{pmatrix}, \qquad i = 1, 2, 3. \tag{5.29}$$

The equations for the phonons are macroscopic equations for heat transport and together with those for the carriers, they constitute a generalization of the system Equations 5.20 through 5.22, where other moments of the distribution functions are considered besides the densities and the energies.

In the Equations 5.27 through 5.29, there appear extra fluxes and production terms. Those relative to phonons are defined by

$$T_{ij}^p = y^p \int \epsilon_p \mathbf{v}_i^p\, \mathbf{v}_j^p\, g^p(\mathbf{q})d\mathbf{q}, \qquad \text{flux of the energy flux,}$$

$$\begin{pmatrix} C_{W^p} \\ C_{Q_i^p} \end{pmatrix} = y^p \int \begin{pmatrix} \epsilon_p \\ \epsilon_p \mathbf{v}_i^p \end{pmatrix}\left[C_{(\mathrm{oth})}^p + \sum_\eta C_\eta^p \right] d\mathbf{q}, \quad \begin{pmatrix} \text{energy production} \\ \text{energy flux production} \end{pmatrix}.$$

While for the carriers, we have

$$\begin{pmatrix} F_{ij}^{A(0)} \\ F_{ij}^{A(1)} \end{pmatrix} = \frac{y^A}{n_A}\int \begin{pmatrix} 1 \\ \mathcal{E}_A \end{pmatrix} v_i^A v_j^A f^A d\mathbf{k}, \quad \begin{pmatrix} \text{velocity flux,} \\ \text{flux of the energy flux} \end{pmatrix}$$

$$\begin{pmatrix} G_{ij}^{A(0)} \\ G_{ij}^{A(1)} \end{pmatrix} = \frac{y^A}{\hbar\, n_A}\int \begin{pmatrix} \dfrac{\partial v_i^A}{\partial k_j} \\ \dfrac{\partial \mathcal{E}_A v_i^A}{\partial k_j} \end{pmatrix} f^A d\mathbf{k},$$

$$n_A\, C_{M_{\psi_A}} = y^A \int \psi_A(\mathbf{k})\left\{ C_{im}(f^A) + \sum_{\iota, A', p} C_{\iota,p}^{AA'}(f^A, f^{A'}, g^p) + \sum_{A'} \mathcal{I}[f^A, f^{A'}] \right\} d\mathbf{k}, \quad M_{\psi_A}\text{-production,}$$

with $\{M_{\psi_A}\} := \{n_A, \mathbf{V}_A, W_A, \mathbf{S}_A\}$. Therefore, in the evolution equations, the number of the unknowns is greater than that of the equations, which means that constitutive equations are needed for the extra variables $T_{ij}^p, F_{ij}^{A(0)}, G_{ij}^{A(0)}, F_{ij}^{A(1)}, G_{ij}^{A(1)}, C_{W^p}, C_{\mathbf{Q}^p}, C_{n_A}, C_{W^A}, C_{\mathbf{V}^A}, C_{\mathbf{S}^A}$.

5.5.1 Closure of the System of Macroscopic Evolution Equations

A systematic way to find the closure relations is founded on a universal physical principle: the *maximum entropy principle* (MEP) [3,31,32]. MEP states that, if a certain number of moments is known, then the least-biased distribution functions, which can be used for evaluating the unknown moments, are those maximizing the total entropy functional under the constraint that they reproduce the known moments. In the case under consideration, neglecting the mutual interactions among the subsystems, the total entropy is the sum of entropy of the several particle populations

$$\mathcal{S} = -k_B \left\{ \sum_A y^A \int_{\mathbb{R}^3} \left(f^A \ln f^A - f^A \right) d\mathbf{k} + \sum_p y^p \int_{\mathcal{B}} \left[g^p \ln g^p \right.\right.$$

$$\left.\left. - \left(1 + g^p\right) \ln \left(1 + g^p\right) \right] d\mathbf{q} \right\}.$$

The first terms are the contribution of electrons and holes according to the Fermi statistics in the non degenerate case, the latter ones are the contribution of the phonons according to Bose statistics [33].

The solution of this maximization problem, linearized with respect to the vector variables [20,24], is given by

$$g^p_{ME} = \frac{1}{\exp\left(\epsilon_p \Lambda_{W^p}\right) - 1} - \frac{\epsilon_p \exp\left(\epsilon_p \Lambda_{W^p}\right)}{\left(\exp\left(\epsilon_p \Lambda_{W^p}\right) - 1\right)^2} \mathbf{v}^p \cdot \Lambda_{\mathbf{Q}^p},$$

$$f^A_{ME} = \exp\left(-\Lambda^A - \Lambda_{W^A} \mathcal{E}_A\right)\left(1 - \mathbf{v}^A \cdot \left(\Lambda_{\mathbf{V}^A} + \mathcal{E}_A \Lambda_{\mathbf{S}^A}\right)\right),$$

where the Λ's are Lagrange multipliers related to the field variables by means of the constraint relations Equations 5.25 and 5.26.

For example, in the previously considered case of silicon, the scalar constraints can be rewritten as follows:

$$W^p = 4\pi y^p d_1^p(\Lambda_{W^p}), \qquad \begin{pmatrix} n_e \\ n_e W^e \end{pmatrix} = \frac{\sqrt{2 m_e}\, m_e y^e J_0}{\hbar^3} e^{-\Lambda^e} \begin{pmatrix} d_0(\Lambda_{W^e}) \\ d_1(\Lambda_{W^e}) \end{pmatrix},$$

where

$$d_1^p(x) := \int_0^{\frac{2\pi}{a}} t^2 \frac{\epsilon_p(t)}{\exp\left(\epsilon_p(t)\, x\right) - 1}\, dt, \quad d_n(x) := \int_0^{\infty} t^n \exp(-xt) \sqrt{\frac{t}{\gamma^5(t)}} \left(\gamma(t) - t \frac{d\gamma(t)}{dt} \right) dt,$$

$$J_0 := \int_{S^2} \psi^{\frac{3}{2}} d\Omega, \quad \psi^{-1} := \sum_{i=1}^{3} \frac{m_e k_i^2}{m_i |\mathbf{k}|^2},$$

S^2 being the unit sphere surface. Inverting, one has

$$\Lambda_{W^p} = h_p^{-1}(W^p) \quad \Lambda^e = -\log \left(\frac{\hbar^3 n_e}{\sqrt{2 m_e}\, m_e y^e J_0 d_0} \right), \quad \Lambda_{W^e} = h_e^{-1}(W^e), \tag{5.30}$$

h_p^{-1} and h_e^{-1} being the inverse functions of $h_p(x) := 4\pi y^p d_1^p(x)$ and $h_e(x) := \frac{d_1(x)}{d_0(x)}$, respectively.

Remark. It is important to notice that according to extended thermodynamics [3], the following relation between the energy Lagrange multiplier of a population and the corresponding local temperature holds: $k_B T = \dfrac{1}{\Lambda_W} = \dfrac{1}{h^{-1}(W)}$.

Similarly, inverting the vector constraints, one obtains

$$\Lambda_{Q_i^p} = \frac{3}{4\pi y^p}\left(\frac{\partial}{\partial \Lambda_{W^p}} p_p^1(\Lambda_{W^p})\right)^{-1} Q_i^p, \; p_p^1(x) := \int_0^{\frac{2\pi}{a}} \frac{\epsilon_p(t)\, t^2 \left|\mathbf{v}^p(t)\right|^2}{\exp\left(\epsilon_p(t)x\right) - 1} dt, \tag{5.31}$$

$$\Lambda_{V_i^e} = \frac{b_{11}(W^e)}{J_{1,i}^e} V_i^e + \frac{b_{12}(W^e)}{J_{1,i}^e} S_i^e, \; \Lambda_{S_i^e} = \frac{b_{12}(W^e)}{J_{1,i}^e} V_i^e + \frac{b_{22}(W^e)}{J_{1,i}^e} S_i^e. \tag{5.32}$$

Here $J_{1,i}^e := \int_{S^2} \frac{m_e^2}{(m_i^e)^2} \psi^{\frac{5}{2}} n_i^2 \, d\Omega$, and the b_{kl} are the elements of the matrix **B**, which is the inverse of the symmetric matrix **A** of elements

$$a_{kl} = -\frac{p^{k+l-2}}{m_e J_0 d_0}, \; k, l = 1, 2, \; \text{with } p^n = p^n(x) := \int_0^{\infty} \frac{2t^{n+\frac{3}{2}} \gamma^{\frac{1}{2}}(t)}{\gamma(t) - t\,\frac{\gamma(t)}{dt}} \, e^{-xt} dt.$$

Substituting Equations 5.30 through 5.32 into the MEP distributions, these latter will depend on (x, t) only through the state variables. In turn, inserting the MEP distributions into the integrals defining the extra variables, the needed closure relations can be eventually obtained.

The procedure requires the combersome computation of integrals over the wave vector space of phonons and carriers; however, it is similar for any semiconductor material and it has been done for Si in [4,7–9, 34–39], for compound semiconductors in [20,24], and for graphene in [23].

5.6 Energy-Transport Model and Its Numerical Discretization

By using MEP, the resulting equations form a hyperbolic system of balance laws [6]. In order to have a formulation which has a more evident physical interpretation and is directly comparable with system Equations 5.20 through 5.22, from the hyperbolic model, by performing a suitable scaling [10], it is possible to get a system of coupled nonlinear parabolic equations, known in the literature as the *energy-transport model* [40]. The energy-transport model can be cast in a drift-diffusion form and, therefore, after suitable modifications, the well-known Scharfetter–Gummel scheme can be employed for the numerical discretization. For a discretization of the original hyperbolic equations see [41].

Let us start by noting that, in the simpler case with one conduction and one valence band, the productions $C_{\mathbf{Q}^p}$ $C_{\mathbf{V}^A}$, $C_{\mathbf{S}^A}$, with p running over the phonon branches and $A = n, h$, appearing in Equations 5.27 through 5.29, have the general form

$$n_A\, C_{\mathbf{V}^A} = \sum_B \left(c_{\mathrm{VV}}^B n_B \mathbf{V}^B + c_{\mathrm{VS}}^B n_B \mathbf{S}^B\right) + \sum_p c_{\mathrm{VQ}}^p \mathbf{Q}^p,$$

$$n_A\, C_{\mathbf{S}^A} = \sum_B \left(c_{\mathrm{SV}}^B n_B \mathbf{V}^B + c_{\mathrm{SS}}^B n_B \mathbf{S}^B\right) + \sum_p c_{\mathrm{SQ}}^p \mathbf{Q}^p,$$

$$C_{\mathbf{Q}^p} = \sum_B \left(c_{\mathrm{QV}}^B n_B \mathbf{V}^B + c_{\mathrm{QS}}^B n_B \mathbf{S}^B\right) + \sum_{p'} c_{\mathrm{QQ}}^{pp'} \mathbf{Q}^{p'},$$

where the c's depend on electron, hole, and phonon temperatures, through their energies. For simplicity, we assume that in the previous expressions the electron–hole mixed terms are negligible as well as the carrier–phonon terms, so that they become

$$n_A\, C_{\mathbf{V}^A} = c^A_{\mathrm{VV}} n_A \mathbf{V}^A + c^A_{\mathrm{VS}} n_A \mathbf{S}^A,$$

$$n_A\, C_{\mathbf{S}^A} = c^A_{\mathrm{SV}} n_A \mathbf{V}^A + c^A_{\mathrm{SS}} n_A \mathbf{S}^A,$$

$$C_{\mathbf{Q}^p} = \sum_{p'} c^{pp'}_{\mathrm{QQ}} \mathbf{Q}^{p'}.$$

Moreover, we assume that all the phonons have the same local temperature T_{L}. Therefore, summing Equations 5.27_1 over p, we have

$$\frac{\partial W}{\partial t} + \nabla_{\mathbf{x}} \cdot \mathbf{Q} = \sum_p C_{W^p} =: H, \tag{5.33}$$

with $W := \sum_p W^p$ and $\mathbf{Q} := \sum_p \mathbf{Q}^p$. Noting that

$$\frac{\partial W}{\partial t} = \frac{dW}{dT_{\mathrm{L}}} \frac{\partial T_{\mathrm{L}}}{\partial t} = \rho c_{\mathrm{V}} \frac{\partial T_{\mathrm{L}}}{\partial t},$$

from Equations 5.28 and 5.33 one can write the following energy-transport system:

$$\frac{\partial n}{\partial t} + \operatorname{div}\left(n\,\mathbf{V}_n\right) = -R + G, \tag{5.34}$$

$$\frac{\partial p}{\partial t} + \operatorname{div}\left(p\,\mathbf{V}_h\right) = -R + G, \tag{5.35}$$

$$\frac{\partial \left(nW_n\right)}{\partial t} + \operatorname{div}\left(n\,\mathbf{S}_n\right) + nq\mathbf{V}_n \cdot \nabla\phi = nC_{W_n}, \tag{5.36}$$

$$\frac{\partial \left(pW_h\right)}{\partial t} + \operatorname{div}\left(p\,\mathbf{S}_h\right) - pq\mathbf{V}_h \cdot \nabla\phi = pC_{W_h}, \tag{5.37}$$

$$\rho c_V \frac{\partial T_{\mathrm{L}}}{\partial t} + \operatorname{div}(\mathbf{Q}) = H, \tag{5.38}$$

$$\mathbf{E} = -\nabla\phi, \quad \Delta(\varepsilon_{\mathrm{d}}\phi) = -q(N_{\mathrm{d}} - N_{\mathrm{a}} - n + p), \tag{5.39}$$

where W_n, W_h, $\mathbf{V}_n$, $\mathbf{V}_h$, $\mathbf{S}_n$, $\mathbf{S}_h$ are the electron and hole average energies, velocities, and energy fluxes, c_{V} is the lattice specific heat, and C_{W_n}, C_{W_h} are the electron and hole energy production terms, which in the relaxation time approximation are given by

$$C_{W_A} = -\frac{W_A - W_A^0}{\tau_{W_A}}, \quad A = n, h, \tag{5.40}$$

with $W_A^0 = W_A(1/k_B T_{\mathrm{L}}) \approx 3/2 k_B T_{\mathrm{L}}$ and τ_{W_A}, $A = n, h$, the energy relaxation times.

In Equations 5.36 and 5.37, the terms $-nq\mathbf{V}_n \cdot \nabla\phi, pq\mathbf{V}_h \cdot \nabla\phi$ represent the energy gain due to the Joule heating, while the terms nC_{W_n}, pC_{W_h} represent the energy loss and gain due to the recombination and generation of electrons and holes, and the energy exchange with the lattice.

It remains to be seen how closure relations for $\mathbf{V}_n$, $\mathbf{V}_h$, $\mathbf{S}_n$, $\mathbf{S}_h$, and $\mathbf{Q}$ can be obtained. For this reason, one can exploit the above-mentioned scaling (the interested reader is referred to [42,43] for more details), which is substantially equivalent to applying the Maxwell procedure to Equations 5.27_2 and 5.29. At the first order, this procedure consists of introducing the equilibrium values, $\mathbf{V}_n = \mathbf{V}_h = \mathbf{S}_n = \mathbf{S}_h = \mathbf{Q} = 0$, in the left-hand sides of Equations 5.27 and 5.29. So doing, one obtains a linear system, which, once solved, gives

$$\mathbf{V}_n = D^n_{11}(W_n, T_\mathrm{L})\,\nabla \log n + D^n_{12}(W_n, T_\mathrm{L})\,\nabla W_n + D^n_{13}(W_n, T_\mathrm{L})\,\nabla\phi, \tag{5.41}$$

$$\mathbf{S}_n = D^n_{21}(W_n, T_\mathrm{L})\,\nabla \log n + D^n_{22}(W_n, T_\mathrm{L})\nabla\, W_n + D^n_{23}(W_n, T_\mathrm{L})\,\nabla\phi, \tag{5.42}$$

$$\mathbf{Q} = -K(T_\mathrm{L})\nabla T_\mathrm{L}, \tag{5.43}$$

with

$$D^n_{11}(W_n, T_\mathrm{L}) = D_n^{-1}\left[c^n_{\mathrm{VS}}F^{(1)}_n - \frac{c^n_{\mathrm{SS}}}{\lambda_{W_n}}\right],\quad D^n_{12}(W_n, T_\mathrm{L}) = D_n^{-1}\left[c^n_{\mathrm{VS}}F^{(1)\prime}_n - \left(\frac{c^n_{\mathrm{SS}}}{\lambda'_{W_n}}\right)\right], \tag{5.44}$$

$$D^n_{13}(W_n, T_\mathrm{L}) = D_n^{-1}\left[c^n_{\mathrm{SS}}q - c^n_{\mathrm{VS}}\,q\,G^{(1)}_n\right], \tag{5.45}$$

$$D^n_{21}(W_n, T_\mathrm{L}) = -D_n^{-1}\left[c^n_{\mathrm{VV}}F^{(1)}_n - \frac{c^n_{\mathrm{SV}}}{\lambda_{W_n}}\right],\quad D^n_{22}(W_n, T_\mathrm{L}) = -D_n^{-1}\left[c^n_{\mathrm{VV}}F^{(1)\prime}_n - \left(\frac{c^n_{\mathrm{SV}}}{\lambda'_{W_n}}\right)\right], \tag{5.46}$$

$$D^n_{23}(W_n, T_\mathrm{L}) = -D_n^{-1}\left[c^n_{\mathrm{SV}}q - c^n_{\mathrm{VV}}\,q\,G^{(1)}_n\right],\quad D_n(W, T_\mathrm{L}) = c^n_{\mathrm{VS}}c^n_{\mathrm{SV}} - c^n_{\mathrm{SS}}c^n_{\mathrm{VV}}. \tag{5.47}$$

Analogous expressions are valid for $\mathbf{V}_h$ and $\mathbf{S}_h$. The coefficients c^A_{ij} depend on W_A and T_L, while the prime indicates a derivative with respect to W. The fluxes $F^{(1)}_A$ and $G^{(1)}_A$ depend on W_A. The complete expressions of c^A_{ij}, $F^{(1)}_A$, and $G^{(1)}_A$ are reported in [43,44], and an explicit expression of the lattice conductivity can be found in [5]. In this way, a generalization of system Equations 5.20 through 5.22 has been obtained where no free parameter is present.

At the source and drain contacts of devices, the Robin boundary condition

$$-K(T_\mathrm{L})\frac{\partial T_\mathrm{L}}{\partial n} = R^{-1}_{\mathrm{th}}(T_\mathrm{L} - T_{\mathrm{env}}) \tag{5.48}$$

is assumed, where R_{th} is the thermal resistivity of the contact, T_{env} is the environment temperature, and $\partial/\partial n$ denotes the derivative along the outer normal at the contacts. R_{th} depends on the material of the contact.

On the lateral boundaries and oxide/silicon interfaces of devices no-flux condition for the temperature is used and Dirichlet condition at the bulk contacts is assumed. The electron energy at the source, drain, and bulk contacts are taken as equal to that corresponding to the lattice temperature (Ohmic contacts). The other boundary conditions are described in [44].

The crystal lattice temperature T_L changes much more slowly than other variables. For instance, the typical time of the lattice temperature for reaching the steady state in the simulations is of the order of thousands of picoseconds, while for the other fields it is of the order of picoseconds. This double-scale behavior can be exploited by adopting the following multirate integration scheme:

- Step 1. First the balance equations for electrons and holes can be integrated, with the crystal lattice energy and the electric field frozen at the time step $k-1$. This gives the electron and hole densities and energies at the time step k, which can be written schematically as

$$\frac{\partial \mathcal{U}^k}{\partial t} + F(\mathcal{U}^k, \phi^{k-1}, T^{k-1}_\mathrm{L}) = 0, \tag{5.49}$$

with $\mathcal{U} = (n, p, W_n, W_h)$. Here, $k = 1, \ldots, N$ is the index of the integration interval $\Delta t_k = [t_{k-1}, t_k]$, with $t_k = t_{k-1} + \Delta t$, with Δt being the time size of the synchronization window.

- Step 2. The lattice energy balance equation is integrated, with n, p, W_n, and W_h given by step 1

$$\rho c_{\mathrm{V}} \frac{\partial T_{\mathrm{L}}^k}{\partial t} - \mathrm{div}\left[K(T_{\mathrm{L}}^k)\nabla T_{\mathrm{L}}^k\right] = H(\mathcal{U}^k, T_{\mathrm{L}}^k), \tag{5.50}$$

along with the Poisson equation with $n = n^k$ and $p = p^k$.

For steps 1 and 2, different time steps for the numerical integration over the interval Δt_k are used. Typically, the time step for integration of Equation 5.50 can be taken as 100 times larger than the time step for Equation 5.49.

5.6.1 Step 1

The numerical scheme is based on an exponential fitting like that employed in the Scharfetter–Gummel scheme for the drift-diffusion model of semiconductors [29]. The basic idea is to split the particle and energy density currents as the difference of two terms. Each of them is written by introducing suitable mean mobilities in order to get expressions of the currents similar to those arising in other energy-transport models known in literature [45]. The equations are spatially discretized on a regular grid. The details of the numerical scheme can be found in [44]. Here, a brief account is given.

For the sake of simplicity, the numerical method is presented only for the electron part, setting the generation–recombination term equal to zero and omitting the index n in the electron variables. The inclusion of holes and the coupling with electrons can be performed straightforwardly.

First, as said, the current density $\mathbf{J} = n\mathbf{V}$ and the energy-flux density $\mathbf{Z} = n\,\mathbf{S}$ are rewritten as

$$\mathbf{J} = \mathbf{J}^{(1)} - \mathbf{J}^{(2)}, \quad \mathbf{Z} \; = \; \mathbf{Z}^{(1)} - \mathbf{Z}^{(2)}, \tag{5.51}$$

and then each term is put into a drift-diffusion form

$$\mathbf{J}^{(1)} = \frac{c_{\mathrm{SS}}}{D}\left[\nabla(n\lambda_W^{-1}) - qn\nabla\phi\right], \quad \mathbf{J}^{(2)} = \frac{c_{\mathrm{VS}}}{D}\left[\nabla\left(nF^{(1)}\right) - qnF^{(1)}\lambda_W\nabla\phi\right], \tag{5.52}$$

$$\mathbf{Z}^{(1)} = \frac{c_{\mathrm{VV}}}{D}\left[\nabla\left(nF^{(1)}\right) - qnF^{(1)}\lambda_W\nabla\phi\right], \quad \mathbf{Z}^{(2)} = \frac{c_{\mathrm{VS}}}{D}\left[\nabla(n\lambda_W^{-1}) - qn\nabla\phi\right]. \tag{5.53}$$

In the two-dimensional (2D) case, the grid points (x_i, y_j) are introduced, with $x_{i+1} - x_i = h =$ constant and $y_{j+1} - y_j = k =$ constant, and the middle points $(x_i, y_{j\pm 1/2}) = (x_i, y_j \pm k/2)$ and $(x_{i\pm 1/2}, y_j) = (x_i \pm h/2, y_j)$. A uniform time step Δt is used and $u_{i,j}^l = u(x_i, y_j, l\,\Delta t)$ is set.

By indicating with J_x and J_y the x and y components of the current density $\mathbf{J}$ and by Z_x and Z_y the x and y components of $\mathbf{Z}$, the balance Equations 5.34 and 5.36 are discretized as

$$\frac{n_i^{l+1} - n_i^l}{\Delta t} + \frac{(J_x)_{i+1/2,j} - (J_x)_{i-1/2,j}}{h} + \frac{(J_y)_{i,j+1/2} - (J_y)_{i,j-1/2}}{k} + O(h^2, k^2, \Delta t) = 0, \tag{5.54}$$

$$\begin{aligned} &\frac{(n\,W)_i^{l+1} - (n\,W)_i^l}{\Delta t} + \frac{(Z_x)_{i+1/2,j} - (Z_x)_{i-1/2,j}}{h} + \frac{(Z_y)_{i,j+1/2} - (Z_y)_{i,j-1/2}}{k} + \\ &-q\frac{(J_x)_{i+1/2,j} + (J_x)_{i-1/2,j}}{2}\,\frac{\phi_{i+1,j} - \phi_{i-1,j}}{2h} - q\frac{(J_y)_{i,j+1/2} + (J_y)_{i,j-1/2}}{2}\,\frac{\phi_{i,j+1} - \phi_{i,j-1}}{2k} \\ &\qquad + n_{i,j}\frac{W_{i,j} - W_0}{(\tau_W)_{i,j}} + O(h^2, k^2, \Delta t) = 0. \end{aligned} \tag{5.55}$$

The variables without temporal index must be evaluated at time level l.

In order to calculate the components of the currents in the middle points, let us consider the sets

$$I_{i+1/2,j} = [x_i, x_{i+1}] \times [y_{j-1/2}, y_{j+1/2}], \quad I_{i,j+1/2} = [x_{i-1/2}, x_{i+1/2}] \times [y_j, y_{j+1}]$$

and expand $J_x^{(r)}$, $r = 1, 2$ in Taylor's series in $I_{i+1/2,j}$

$$J_x^{(r)}(x,y) = (J_x^{(r)})_{i+1/2,j} + (x - x_{i+1/2})\left(\frac{\partial J_x^{(r)}}{\partial x}\right)_{i+1/2,j} + (y - y_j)\left(\frac{\partial J_x^{(r)}}{\partial y}\right)_{i+1/2,j} + o(\Delta x, \Delta y)$$

and $J_y^{(r)}$, $r = 1, 2$ in Taylor's series in $I_{i,j+1/2}$

$$J_y^{(r)}(x,y) = (J_y^{(r)})_{i,j+1/2} + (x - x_i)\left(\frac{\partial J_y^{(r)}}{\partial x}\right)_{i,j+1/2} + (y - y_{j+1/2})\left(\frac{\partial J_y^{(r)}}{\partial y}\right)_{i,j+1/2} + o(\Delta x, \Delta y).$$

First, $U_T = \frac{1}{q\lambda_W}$ is introduced, which plays the role of a thermal potential [44] and indicates by $\overline{U}_T$ its piecewise constant approximation, which is given by $\overline{U}_T = \dfrac{\lambda_W^{-1}(W_{i,j}) + \lambda_W^{-1}(W_{i+1,j})}{2q}$ in the cell $I_{i+1/2,j}$ and by $\overline{U}_T = \dfrac{\lambda_W^{-1}(W_{i,j+1}) + \lambda_W^{-1}(W_{i,j})}{2q}$ in the cell $I_{i,j+1/2}$. Then the *local* mobilities

$$g_{11} = -\frac{\bar{c}_{SS}}{\overline{D}} n\lambda_W^{-1}, \quad g_{12} = -\frac{\bar{c}_{VS}}{\overline{D}} nF^{(1)}, \tag{5.56}$$

$$g_{21} = -\frac{\bar{c}_{VV}}{\overline{D}} nF^{(1)}, \quad g_{22} = -\frac{\bar{c}_{VS}}{\overline{D}} n\lambda_W^{-1} \tag{5.57}$$

are introduced, where $\bar{c}_{pq}$ is a piecewise constant approximation of c_{pq} p, q = V, S, given by $\bar{c}_{pq} = c_{pq}\left(\frac{W_{i,j}+W_{i+1,j}}{2}\right)$ in the cell $I_{i+1/2,j}$ and by $\bar{c}_{pq} = c_{pq}\left(\frac{W_{i,j}+W_{i,j+1}}{2}\right)$ in the cell $I_{i,j+1/2}$, and, similarly to [46], the *local* Slotboom variables

$$s_{kr} = \exp\left(-\phi/\overline{U}_T\right) g_{kr} \quad k, r = 1, 2$$

that satisfy

$$\nabla s_{1r} \simeq -\exp\left(-\phi/\overline{U}_T\right) \mathbf{J}^{(r)}, \quad \nabla s_{2r} \simeq -\exp\left(-\phi/\overline{U}_T\right) \mathbf{Z}^{(r)} \quad r = 1, 2. \tag{5.58}$$

From the x component of Equation 5.58$_1$, we have

$$\frac{\partial s_{1r}(x, y_j)}{\partial x} \simeq -\exp\left(-\phi/\overline{U}_T\right) J_x^{(r)}(x, y_j)$$

$$= -\exp\left(-\phi/\overline{U}_T\right)\left\{(J_x^{(r)})_{i+1/2,j} + (x - x_{i+1/2})\left(\frac{\partial J_x^{(r)}}{\partial x}\right)_{i+1/2,j} + o(\Delta x, \Delta y)\right\}, \tag{5.59}$$

which, after integration over $[x_i, x_{i+1}]$, where a piecewise linear approximation is used for ϕ and some algebra, gives

$$(J_x^{(r)})_{i+1/2,j} = -z_{i+1/2,j} \coth z_{i+1/2,j} \frac{(g_{1r})_{i+1,j} - (g_{1r})_{i,j}}{h} + z_{i+1/2,j} \frac{(g_{1r})_{i+1,j} + (g_{1r})_{i,j}}{h}, \quad r = 1, 2, \tag{5.60}$$

where $z_{i+1/2,j} = \frac{\phi_{i+1,j} - \phi_{i,j}}{2\bar{U}_T}$.

Likewise, by evaluating the y component of Equation 5.58_2 and integrating over $[y_j, y_{j+1}]$, we find

$$(J_y^{(r)})_{i,j+1/2} = -z_{i,j+1/2} \coth z_{i,j+1/2} \frac{(g_{1r})_{i,j+1} - (g_{1r})_{i,j}}{k} + z_{i,j+1/2} \frac{(g_{1r})_{i,j+1} + (g_{1r})_{i,j}}{k}, \quad r = 1, 2, \tag{5.61}$$

where $z_{i,j+1/2} = \frac{\phi_{i,j+1} - \phi_{i,j}}{2\bar{U}_T}$. Using the same procedure, the following discrete expression for the components of the energy flux is obtained:

$$(Z_x^{(r)})_{i+1/2,j} = -z_{i+1/2,j} \coth z_{i+1/2,j} \frac{(g_{2r})_{i+1,j} - (g_{2r})_{i,j}}{h} + z_{i+1/2,j} \frac{(g_{2r})_{i+1,j} + (g_{2r})_{i,j}}{h}, \tag{5.62}$$

$$(Z_y^{(r)})_{i,j+1/2} = -z_{i,j+1/2} \coth z_{i,j+1/2} \frac{(g_{2r})_{i,j+1} - (g_{2r})_{i,j}}{k} + z_{i,j+1/2} \frac{(g_{2r})_{i,j+1} + (g_{2r})_{i,j}}{k}, \quad r = 1, 2. \tag{5.63}$$

The error in formulas Equations 5.60 through 5.63 is $o(h, k)$.

The Poisson equation is solved by replacing it with

$$\phi_t - \operatorname{div}\left(\varepsilon_d \nabla \phi\right) = q(N_D - N_A - n). \tag{5.64}$$

The solution of Equation 5.64 as $t \mapsto +\infty$ is the same as that of the original Poisson equation, at least in the smooth case.

If a time step $\Delta\hat{t}$ is introduced and $\phi_{ij}^r = \phi(x_i, y_j, r\Delta\hat{t})$ is set, Equation 5.64 can be discretized in an explicit way as

$$\phi_{ij}^{r+1} = \phi_{ij}^r + \varepsilon_d \Delta\hat{t} \left[\frac{1}{h^2}\left(\phi_{i+1,j} - 2\phi_{i,j} + \phi_{i-1,j}\right) + \frac{1}{k^2}\left(\phi_{i,j+1} - 2\phi_{i,j} + \phi_{i,j-1}\right) + q(C_{i,j} - n_{i,j})\right], \tag{5.65}$$

where $C_{ij} = N_d(x_i, y_j) - N_a(x_i, y_j)$, with the notable advantage of taking easily into account the different types of boundary conditions. The price to pay is that at each time step, one needs to reach the stationary state of Equation 5.64 by using a time step satisfying the Courant–Friedrichs–Lewy (CFL) condition, common for parabolic equations,

$$\Delta\hat{t} \le \frac{1}{2} \frac{1}{\frac{1}{h^2} + \frac{1}{k^2}}.$$

However, the computational effort is comparable with that required by direct methods.

5.6.2 Step 2

A coordinate splitting technique [47] is used for the solution of the lattice energy equation for the variable $u = k_B T_L$, with time step Δt_T. The splitting technique allows an efficient employment of stable implicit time schemes. The procedure consists of two steps with two suboperators (with corresponding boundary conditions)

$$\rho c_V \frac{u^{n+1/2} - u^n}{\Delta t_T} = \frac{\partial}{\partial x}\left[K(T_L^n)\frac{\partial u^{n+1/2}}{\partial x}\right] + \frac{k_B}{2}H^{n+1/2} \tag{5.66}$$

and

$$\rho c_V \frac{u^{n+1} - u^{n+1/2}}{\Delta t_T} = \frac{\partial}{\partial y}\left[K(T_L^n)\frac{\partial u^{n+1}}{\partial y}\right] + \frac{k_B}{2}H^{n+1}. \tag{5.67}$$

Here, $u^n = u(x, y, n\Delta t_T)$. The scheme is absolutely stable and approximates the equation of the lattice energy with first-order accuracy in time. For the approximation of the spatial derivatives, the standard stencil with three points can been chosen. For instance, the approximation of Equation 5.67 is

$$\rho c_V u_{i,j}^{n+1} = \rho c_V u_{i,j}^{n+1/2} + \frac{\Delta t_T}{k^2}\left[\frac{\tilde{K}_{i,j} + \tilde{K}_{i,j+1}}{2}(u_{i,j+1}^{n+1} - u_{i,j}^{n+1}) - \frac{\tilde{K}_{i,j} + \tilde{K}_{i,j-1}}{2}(u_{i,j}^{n+1} - u_{i,j-1}^{n+1})\right] + \frac{k_B}{2}H^{n+1},$$

where $\tilde{K}_{i,j} = K(T_{\mathrm{L}i,j})$. Of course, such a discretization is valid in the interior points of the mesh.

The Robin boundary condition Equation 5.48 is approximated as

$$-K\left(\frac{u_{i,1}^n + u_{i,0}^n}{2k_B}\right)\frac{u_{i,1}^{n+1} - u_{i,0}^{n+1}}{k} = R_{\mathrm{th}}^{-1}(u_{i,0}^{n+1} - k_B T_{\mathrm{env}}). \tag{5.68}$$

Here, it has been assumed that at the portion of boundary where the Robin condition holds, one has $j = 0$ and the closest interior points have $j = 1$. The resulting linear system can be solved efficiently with the tridiagonal matrix factorization procedure.

5.7 Thermal Properties of Graphene

Recent years have witnessed a great interest in 2D materials for their promising applications. The most widely investigated one is graphene, but lately the single-layer transition metal dichalcogenides (TMDCs), such as molybdenum disulfite and tungsten diselenite, and black phosphorus have received certain attention [48]. In this section, we analyze the thermal and electrical properties of suspended graphene monolayers. What follows is based on Ref. [49].

5.7.1 Semiclassical Description

Graphene is considered a potential new semiconductor material for future applications in nanoelectronic [50–55] and optoelectronic devices [48] because it has very good mechanical properties, it is an excellent heat and electricity conductor, and it also has noteworthy optical properties. It is a 2D allotrope of carbon that consists of carbon atoms tightly packed into a honeycomb hexagonal lattice (see Figure 5.7), due to their sp^2 hybridization. The atoms in graphene form two interpenetrating triangular Bravais sublattices and each unit cell contain two atoms belonging to different sublattices [51]. The first Brillouin zone is hexagonal like the Bravais lattice. There are six σ bonds per unit cell whose orbitals yield six bands, three of which (called σ) are below the Fermi energy and three of which are above (the σ^* bands).

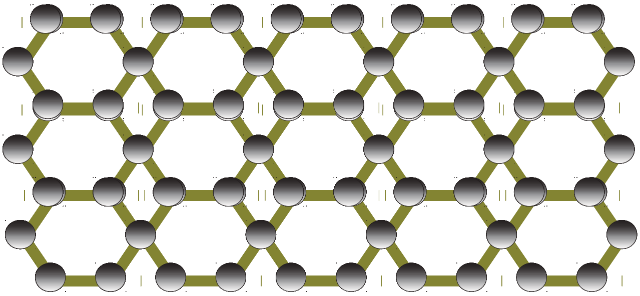


FIGURE 5.7 Honeycomb crystal lattice of graphene.

Moreover, each primitive cell contributes two $2p_z$-orbitals that participate in π bonding. The p_z electrons can be treated independently from the other valence electrons, since the overlap between the corresponding orbitals and the s or p_x and p_y orbitals is strictly zero by symmetry. Therefore, within the tight binding approximation, the energy dispersion relations of the π bands can be calculated by solving the eigenvalue problem for the approximate[†] effective tight-binding Hamiltonian

$$\mathcal{H}(\mathbf{k}) = t_O \begin{pmatrix} 0 & \gamma^*_{\mathbf{k}} \\ \gamma_{\mathbf{k}} & 0 \end{pmatrix},$$

with t_O the nearest neighbor (nn) C–C tight binding overlap energy and

$$\gamma_{\mathbf{k}} := 1 + e^{i\mathbf{k}\cdot\mathbf{a}_1} + e^{i\mathbf{k}\cdot\mathbf{a}_2}$$

the sum of the nn phase factors, $\mathbf{a}_1$ and $\mathbf{a}_2$ being the primitive vectors.

One finds

$$\mathcal{E}^{\pm} = \pm t_O |\gamma_{\mathbf{k}}|,$$

where the + refers to conduction (π^*) band and the − refers to the valence (π) band. The points where the bands touch each other are called Dirac points, at these points $\mathcal{E}^{\pm} = 0$. They coincide with the vertices of the first Brillouin zone and are divided into two nonequivalent sets of three points. These two sets are named K− and K′−points.

In the proximity of the Dirac points, for $|\mathbf{k}|a << 1$ (for simplicity, here and in the following expression wave vectors are measured with respect to the K(K′) point), a being the distance between nn carbon atoms, the energy dispersion relation can be approximated by (see Figure 5.8)

$$\mathcal{E}^{\pm} = \pm\hbar v_F |\mathbf{k}|, \tag{5.69}$$

with $v_F = \dfrac{3|t_O|a}{2\hbar}$ the Fermi velocity. Electrons that mostly contribute to the charge transport in intrinsic graphene are indeed those in the two valleys around the Dirac points. Therefore, four populations of electrons will be taken into account, which will be labeled ℓ =K, K′ and $s = -1, 1$ (or equivalently $s = \pi, \pi^*$), respectively. On account of Equation 5.69, their energy is given by

$$\mathcal{E}_{\ell,s} = s\hbar v_F |\mathbf{k} - \mathbf{k}_{\ell}|,$$

[†] It does not take into account the next nearest neighbor hopping correction.

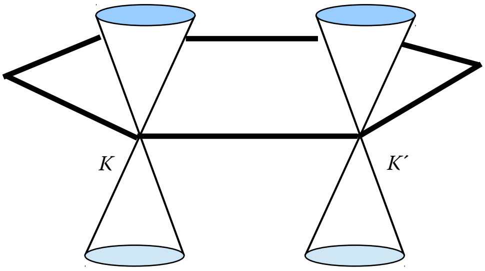


FIGURE 5.8 Energy bands in graphene around the Dirac points K and K′.

with $\mathbf{k}_\ell$ the position of the Dirac point ℓ. Because of the linearity of these relations, electrons behave as massless Dirac fermions. This linear dispersion, with zero band-gap, gives rise to the appeal of graphene for nanophotonic applications. Graphene has a high sensitivity to optical signals through various mechanisms of light–matter interaction. For this reason, it is applied in photodetectors and modulators at near-infrared, visible, and ultraviolet spectrum range, due to its broadband absorption of light through interband transitions [33]. Moreover, since graphene supports plasmons, it is also employed in plasmonics [33].

In a semiclassical kinetic setting, the charge transport in graphene is described by four Boltzmann equations, one for electrons in the valence band (π) and one for electrons in the conduction band (π*), that in turn can belong to the K or K′ valley,

$$\frac{\partial f_{\ell,s}(t,\mathbf{x},\mathbf{k})}{\partial t} + \mathbf{v}_{\ell,s} \cdot \nabla_{\mathbf{x}} f_{\ell,s}(t,\mathbf{x},\mathbf{k}) - \frac{q}{\hbar}\mathbf{E} \cdot \nabla_{\mathbf{k}} f_{\ell,s}(t,\mathbf{x},\mathbf{k}) = \left.\frac{df_{\ell,s}}{dt}(t,\mathbf{x},\mathbf{k})\right|_{e-ph}, \tag{5.70}$$

where $f_{\ell,s}(t,\mathbf{x},\mathbf{k})$ represents the distribution function of charge carriers, in the valence or conduction band ($s = -1$ or $s = 1$) and valley ℓ (K or K′), at position $\mathbf{x}$, time t, and with wave vector $\mathbf{k}$. We denote by $\nabla_{\mathbf{x}}$ and $\nabla_{\mathbf{k}}$ the gradients with respect to the position and the wave vector, respectively. The relationship between the group velocity $\mathbf{v}_{\ell,s}$ and the band energy $\mathcal{E}_{\ell,s}$ can be expressed as

$$\mathbf{v}_{\ell,s} = \frac{1}{\hbar}\nabla_{\mathbf{k}}\,\mathcal{E}_{\ell,s}\,.$$

The electric field $\mathbf{E}$ is assumed as external here. On the right-hand sides of Equation 5.70 are the collision terms, representing the interactions of electrons with acoustic, Γ-*LO* and Γ-*TO*, respectively, and K-phonons.

Scattering with acoustic phonons is intravalley and intraband and a linear dispersion relation

$$\omega_{ac} = v_s|\mathbf{q}|$$

is assumed for them, with ω_{ac} the acoustic phonon frequency and v_s the sound speed in graphene.

The Γ-*LO* and Γ-*TO* phonons are in-plane (with respect to the plane of the lattice) modes, which give rise to electron intravalley transitions, which mostly involve phonons with wave vectors near to the center Γ of the Brillouin zone. The wave vectors $\mathbf{q}$ of these phonons are near to zero and for their dispersion relations the Einstein approximation can be used, according to which $\hbar\omega \approx \text{const}$, with ω the phonon frequency. Such transitions can be both intraband and interband.

The K-phonons are not a real phonon branch. Their name is due to the fact that their wave vectors are close to the K- or K′ point [53]. They belong to an optical branch and induce intervalley scatterings. An

Einstein approximation with a mean phonon energy is used for them. Interaction of electrons with out-of-plane phonon modes, the so-called *Z* phonons, or with in-plane phonons having wave vectors far from the Γ or *K* point is negligible [56], even though they can play some role in the thermal effects. For an accurate description of phonon dispersion relations and thermal conductivity in graphene, we refer the interested reader to [57,58].

The general form of the collision term can be written as [51,53]

$$\left.\frac{df_{\ell,s}}{dt}(t,\mathbf{x},\mathbf{k})\right|_{e-ph} = \sum_{\ell',s'}\left[\int_{\mathcal{B}} S_{\ell',s',\ell,s}(\mathbf{k}',\mathbf{k})f_{\ell',s'}(t,\mathbf{x},\mathbf{k}')\left(1-f_{\ell,s}(t,\mathbf{x},\mathbf{k})\right)d\mathbf{k}' - \int_{\mathcal{B}} S_{\ell,s,\ell',s'}(\mathbf{k},\mathbf{k}')f_{\ell,s}(t,\mathbf{x},\mathbf{k})\left(1-f_{\ell',s'}(t,\mathbf{x},\mathbf{k}')\right)d\mathbf{k}'\right],$$

where the total transition rate $S_{\ell',s',\ell,s}(\mathbf{k}',\mathbf{k})$ is given by the sum of the contributions of the several types of scatterings described earlier:

$$\begin{aligned} S_{\ell',s',\ell,s}(\mathbf{k}',\mathbf{k}) &= \sum_{p}\left|G^{(p)}_{\ell',s',\ell,s}(\mathbf{k}',\mathbf{k})\right|^2\Big[\left(g_p^- + 1\right)\delta\Big(\mathcal{E}_{\ell,s}(\mathbf{k}) - \mathcal{E}_{\ell',s'}(\mathbf{k}') + \hbar\,\omega_p\Big) \\ &\quad + g_p^+\,\delta\left(\mathcal{E}_{\ell,s}(\mathbf{k}) - \mathcal{E}_{\ell',s'}(\mathbf{k}') - \hbar\,\omega_p\right)\Big]. \end{aligned} \tag{5.71}$$

The index *p* labels the *p*th phonon population. The $\left|G^{(p)}_{\ell',s',\ell,s}(\mathbf{k}',\mathbf{k})\right|^2$'s are the electron–phonon coupling matrix elements, which describe the interaction mechanism of an electron with a *p*-phonon, from the state of wave vector $\mathbf{k}'$ belonging to the valley ℓ' and band s' to the state $\mathbf{k}$ belonging to the valley ℓ and band *s*. The symbol δ denotes the Dirac distribution, ω_p is the *p*th phonon frequency, $g_p(\mathbf{q})$ is the phonon distribution for the *p*-type phonons. In Equation 5.71, $g_p^{\pm} = g_p\left(\mathbf{q}^{\pm}\right)$, where $\mathbf{q}^{\pm} = \pm\left(\mathbf{k}' - \mathbf{k}\right)$, stemming from the momentum conservation. For acoustic phonons, we consider the elastic approximation according to which the transition rate reads

$$\frac{1}{(2\,\pi)^2}\,\frac{\pi\,D_{ac}^2\,k_B\,T}{2\hbar\,\sigma_m\,v_s^2}\left(1+\cos\vartheta_{\mathbf{k},\mathbf{k}'}\right)\delta\left(\mathcal{E}(\mathbf{k}') - \mathcal{E}(\mathbf{k})\right), \tag{5.72}$$

where D_{ac} is the acoustic phonon-coupling constant, σ_m is the graphene areal density, and $\vartheta_{\mathbf{k},\mathbf{k}'}$ is the convex angle between $\mathbf{k}$ and $\mathbf{k}'$.

The electron–phonon coupling matrix elements for the intraband scatterings of the electrons in the conduction band with the *LO*, the *TO*, and the K-phonons are [53]

$$\left|G^{(LO)}(\mathbf{k}',\mathbf{k})\right|^2 = \frac{1}{(2\,\pi)^2}\,\frac{\pi\,D_O^2}{\sigma_m\,\omega_O}\left(1 - \cos(\vartheta_{\mathbf{k},\mathbf{k}'-\mathbf{k}} + \vartheta_{\mathbf{k}',\mathbf{k}'-\mathbf{k}})\right) \tag{5.73}$$

$$\left|G^{(TO)}(\mathbf{k}',\mathbf{k})\right|^2 = \frac{1}{(2\,\pi)^2}\,\frac{\pi\,D_O^2}{\sigma_m\,\omega_O}\left(1 + \cos(\vartheta_{\mathbf{k},\mathbf{k}'-\mathbf{k}} + \vartheta_{\mathbf{k}',\mathbf{k}'-\mathbf{k}})\right) \tag{5.74}$$

$$\left|G^{(K)}(\mathbf{k}',\mathbf{k})\right|^2 = \frac{1}{(2\,\pi)^2}\,\frac{2\pi\,D_K^2}{\sigma_m\,\omega_K}\left(1 - \cos\vartheta_{\mathbf{k},\mathbf{k}'}\right), \tag{5.75}$$

where D_O is the optical phonon coupling constant, ω_O is the optical phonon frequency, D_K is the K-phonon coupling constant and ω_K is the K-phonon frequency. The angles $\vartheta_{\mathbf{k},\mathbf{k}'-\mathbf{k}}$ and $\vartheta_{\mathbf{k}',\mathbf{k}'-\mathbf{k}}$ denote the convex angles between $\mathbf{k}$ and $\mathbf{k}' - \mathbf{k}$ and between $\mathbf{k}'$ and $\mathbf{k}' - \mathbf{k}$, respectively.

TABLE 5.1 Physical Parameters for the Scattering Rates

σ_m	7.6×10^{-8} g/cm^2
v_F	10^6 m/s
v_s	2×10^4 m/s
D_{ac}	6.8 eV
$\hbar\,\omega_O$	164.6 meV
D_O	10^9 eV/cm
$\hbar\,\omega_K$	124 meV
D_K	3.5×10^8 eV/cm

In the literature, there are several values for the coupling constants entering the collision terms. For example, for the acoustic deformation potential one can find values ranging from 2.6 to 29 eV. A similar degree of uncertainty is found for the optical and K-phonon coupling constants as well. In the following numerical simulations of monolayer graphene, the parameters proposed in [56,59] have been adopted. They are reported in Table 5.1. By applying a gate voltage, the Fermi energy can be varied. In the sequel, we consider only the case of high values of the Fermi energy, which is equivalent for conventional semiconductors to an n-type doping. Under such a condition, the dynamics of the electrons belonging to the valence band can be neglected, and this is the reason why only the intraband coupling matrix elements for the electrons of the conduction band were reported earlier. In order to simplify the notation the indices s and ℓ will be omitted in the following.

From the semiclassical transport equations, using the procedure shown in the previous sections (see e.g., [24,50,60–63]), one can also formulate macroscopic models that are more suited for computer-aided design (CAD) purposes because they avoid the numerical solutions of the Boltzmann equations, even if they introduce some approximation for the closure relations that are needed.

The evolution of the phonon distributions is governed by the following Boltzmann equations:

$$\frac{\partial g_\eta}{\partial t} = C_\eta, \quad \eta = LO, TO, K, \tag{5.76}$$

$$\frac{\partial g_{\mathrm{ac}}}{\partial t} + \mathbf{c}_{\mathrm{ac}} \cdot \nabla_{\mathbf{x}} g_{\mathrm{ac}} = C_{\mathrm{ac}}, \tag{5.77}$$

with $\mathbf{c}_{\mathrm{ac}} = \nabla_{\mathbf{q}}\omega_{\mathrm{ac}} = v_s \frac{\mathbf{q}}{|\mathbf{q}|}$ the acoustic group velocity (on account of the Einstein approximation, the group velocity of the other phonons is negligible). The collision terms C_η and C_{ac} describe the interaction of the phonons with the electrons and the other phonons.

The phonon–phonon collision term is very complicated and represents a formidable task from a numerical point of view. For this reason, in the applications, the Bhatnagar–Gross–Krook (BGK) approximation is usually employed. It gives a simple way to describe the tendency of each species of phonons to an equilibrium distribution. We adopt the same approach as in [4] modeling the phonon–phonon interaction by requiring that each type of phonons relaxes to a local equilibrium given by a Bose–Einstein distribution evaluated, for all the phonon species, at a common local equilibrium temperature T_{LE}, determined by the procedure we specify below. Therefore, we split the phonon collision terms as follows:

$$C_\eta = C_{\eta-e} - \frac{g_\eta - g_\eta^{\mathrm{LE}}}{\tau_\eta}, \tag{5.78}$$

$$C_{\mathrm{ac}} = C_{\mathrm{ac}-e} - \frac{g_{\mathrm{ac}} - g_{\mathrm{ac}}^{\mathrm{LE}}}{\tau_{\mathrm{ac}}}, \tag{5.79}$$

where $C_{\eta-e}$ and $C_{\text{ac}-e}$ describe the phonon–electron collisions, τ_η and τ_{ac} are the phonon relaxation times, and g_η^{LE} and $g_{\text{ac}}^{\text{LE}}$ are the local equilibrium phonon distributions given by

$$g_\eta^{\text{LE}} = \left[e^{\hbar\omega_\eta/k_B T_{\text{LE}}} - 1\right]^{-1}, \tag{5.80}$$

$$g_{\text{ac}}^{\text{LE}} = \left[e^{\hbar v_s q/k_B T_{\text{LE}}} - 1\right]^{-1}. \tag{5.81}$$

Let us introduce the average phonon energies

$$W_{LO+TO} = \frac{2}{(2\pi)^2}\int_{\mathcal{B}} \hbar\omega_{LO/TO}\, g_{LO/TO}\, d\mathbf{q}, \quad W_K = \frac{1}{(2\pi)^2}\int_{\mathcal{B}} \hbar\omega_K\, g_K\, d\mathbf{q}, \tag{5.82}$$

$$W_{\text{ac}} = \frac{3}{(2\pi)^2}\int_{\mathbb{R}^2} \hbar\omega_{\text{ac}}\, g_{\text{ac}}\, d\mathbf{q}, \tag{5.83}$$

where the LO and the TO phonons, having the same energy, have been considered, in the first approximation, as a unique population ($g_{LO} \approx g_{TO}$) with a density of states equal to $\frac{2}{(2\pi)^2}$, while the density of states of the acoustic phonons is equal to $\frac{3}{(2\pi)^2}$ on account of their three polarization states.

By multiplying, for each species, the phonon Boltzmann equation with the phonon energy and by integrating with respect to the wave vector $\mathbf{q}$, one gets the following macroscopic balance equations for the average phonon energies

$$\frac{\partial W_{LO+TO}}{\partial t} = C_{W_{LO+TO}}, \quad \frac{\partial W_K}{\partial t} = C_{W_K}, \tag{5.84}$$

$$\frac{\partial W_{\text{ac}}}{\partial t} = -\nabla_{\mathbf{x}} \cdot \mathbf{Q}_{\text{ac}} + C_{W_{\text{ac}}}, \tag{5.85}$$

where $\mathbf{Q}_{\text{ac}}$ is the phonon energy-flux. While

$$C_{W_{LO+TO}} = C_{W_{LO+TO}}^{p-e} - \frac{W_{LO+TO} - W_{LO+TO}^{\text{LE}}}{\tau_{LO+TO}}, \quad C_{W_K} = C_{W_K}^{p-e} - \frac{W_K - W_K^{\text{LE}}}{\tau_K} \tag{5.86}$$

$$C_{W_{\text{ac}}} = C_{W_{\text{ac}}}^{p-e} - \frac{W_{\text{ac}} - W_{\text{ac}}^{\text{LE}}}{\tau_{\text{ac}}}, \tag{5.87}$$

are the energy-production terms with

$$W_{LO+TO}^{\text{LE}} = \frac{2}{(2\pi)^2}\int_{\mathcal{B}} \hbar\omega_{LO/TO} \left[e^{\hbar\omega_\eta/k_B T_{\text{LE}}} - 1\right]^{-1} d\mathbf{q}, \tag{5.88}$$

$$W_K^{\text{LE}} = \frac{1}{(2\pi)^2}\int_{\mathcal{B}} \hbar\omega_K \left[e^{\hbar\omega_\eta/k_B T_{\text{LE}}} - 1\right]^{-1} d\mathbf{q}, \tag{5.89}$$

$$W_{\text{ac}}^{\text{LE}} = \frac{3}{(2\pi)^2}\int_{\mathcal{B}} \hbar\omega_{\text{ac}} \left[e^{\hbar v_s q/k_B T_{\text{LE}}} - 1\right]^{-1} d\mathbf{q}, \tag{5.90}$$

the average energies in local equilibrium, while $C^{p-e}_{W_{LO+TO}}$, $C^{p-e}_{W_K}$, and $C^{p-e}_{W_{\mathrm{ac}}}$ are the contributions arising from the phonon–electron interactions. Due to energy conservation, these satisfy the relationships

$$C^{e-p}_{W_{LO+TO}} + C^{p-e}_{W_{LO+TO}} = 0, \quad C^{e-p}_{W_K} + C^{p-e}_{W_K} = 0, \tag{5.91}$$

$$C^{e-p}_{W_{\mathrm{ac}}} + C^{p-e}_{W_{\mathrm{ac}}} = 0, \tag{5.92}$$

where $C^{e-p}_{W_{LO+TO}}$ and $C^{e-p}_{W_{\mathrm{ac}}}$ are the energy production terms for electrons.

Since the electron flow does not explicitly enter the phonon–phonon scatterings, the conservation of the total phonon energy in absence of an external source implies

$$\frac{W_{LO+TO} - W^{\mathrm{LE}}_{LO+TO}}{\tau_{LO+TO}} + \frac{W_K - W^{\mathrm{LE}}_K}{\tau_K} + \frac{W_{\mathrm{ac}} - W^{LE}_{\mathrm{ac}}}{\tau_{\mathrm{ac}}} = 0. \tag{5.93}$$

Now we can define the local equilibrium temperature as follows (see [4]).

Definition 5.7.1 The temperature T_{LE} is the common temperature that we must assign to each species in order to have

$$W_{LO+TO} + W_K + W_{\mathrm{ac}} = W^{\mathrm{LE}}_{LO+TO} + W^{\mathrm{LE}}_K + W^{\mathrm{LE}}_{\mathrm{ac}}. \tag{5.94}$$

In other words, T_{LE} is the common temperature each phonon species should to be in thermodynamic equilibrium in order to preserve the total energy.

By taking into account the expressions 5.88 through 5.90, we find that the nonlinear equation T_{LE} must satisfy

$$\begin{aligned} h(T_{\mathrm{LE}}) :=& \frac{2\hbar\omega_{LO/TO} A}{(2\pi)^2}\left[e^{\hbar\omega_{LO/TO}/k_B T_{\mathrm{LE}}} - 1\right]^{-1} + \frac{\hbar\omega_K A}{(2\pi)^2}\left[e^{\hbar\omega_K/k_B T_{\mathrm{LE}}} - 1\right]^{-1} \\ &+ \left(k_B T_{\mathrm{LE}}\right)^3 \frac{3\zeta(3)}{\pi\hbar^2 v_s^2} = W_{LO+TO} + W_K + W_{\mathrm{ac}}, \end{aligned} \tag{5.95}$$

where $\zeta(\cdot)$ is the zeta function and $A = \frac{8\sqrt{3}\pi^2}{9a^2}$ is the area of the first Brillouin zone, with $a = 0.142$ nm. In Equation 5.95, W_{LO+TO}, W_K, and W_{ac} are the current values of the average phonon energies.

We observe that

$$\begin{aligned} h'(T_{\mathrm{LE}}) =& \frac{2(\hbar\omega_{LO/TO})^2 A}{(2\pi)^2(k_B T_{\mathrm{LE}})^2} e^{\hbar\omega_{LO/TO}/k_B T_{\mathrm{LE}}}\left[e^{\hbar\omega_{LO/TO}/k_B T_{\mathrm{LE}}} - 1\right]^{-2} \\ &+ \frac{(\hbar\omega_K)^2 A}{(2\pi)^2(k_B T_{\mathrm{LE}})^2} e^{\hbar\omega_K/k_B T_{\mathrm{LE}}}\left[e^{\hbar\omega_K/k_B T_{\mathrm{LE}}} - 1\right]^{-2} + \left(k_B T_{\mathrm{LE}}\right)^2 \frac{9\zeta(3)}{\pi\hbar^2 v_s^2} > 0, \quad \forall\, T_{\mathrm{LE}} > 0, \\ &\lim_{T_{\mathrm{LE}}\mapsto 0^+} h(T_{\mathrm{LE}}) = 0, \quad \lim_{T_{\mathrm{LE}}\mapsto +\infty} h(T_{\mathrm{LE}}) = +\infty, \end{aligned}$$

and therefore Equation 5.95 admits a unique solution for any assigned positive W_{LO+TO}, W_K, and W_{ac}.

By taking into account the definition of T_{LE}, a simple way to satisfy Equation 5.93 is to take

$$\tau_{LO+TO} = \tau_K = \tau_{\mathrm{ac}} = \tau. \tag{5.96}$$

In the sequel, the phonon relaxation times will be taken as equal. Moreover, we assume what follows.

Assumption 5.7.1 The phonon distributions are considered as sums of even and odd parts, and the former are supposed to be locally of Bose–Einstein type.

As a consequence, from Equations 5.88 to 5.90 with T_{LE} substituted by the respective phonon temperatures, one finds that the temperatures of the phonon species are related to the respective average energies according to

$$k_B T_{LO/TO} = \frac{\hbar\omega_{LO/TO}}{\ln\left(1 + \dfrac{2A\hbar\omega_{LO/TO}}{(2\pi)^2 W_{LO+TO}}\right)}, \quad k_B T_K = \frac{\hbar\omega_K}{\ln\left(1 + \dfrac{A\hbar\omega_K}{(2\pi)^2 W_K}\right)}, \tag{5.97}$$

$$k_B T_{\mathrm{ac}} = W_{\mathrm{ac}}^{1/3}\left(\frac{3\zeta(3)}{\pi\hbar^2 v_s^2}\right)^{-1/3}. \tag{5.98}$$

5.7.2 The Simulation Scheme

For a homogeneous monolayer graphene sheet under a constant electric field **E**, the only significant components of the evolution equations are those parallel to the field and there is no dependence on the spatial variables. By choosing a reference frame in the plane of the graphene sheet with the x-axis parallel to **E**, the complete model consists of the following equations:

$$\begin{aligned}\frac{\partial f(t,\mathbf{k})}{\partial t} - \frac{q}{\hbar}E\frac{\partial f(t,\mathbf{k})}{\partial k_x} &= \int S(\mathbf{k}',\mathbf{k}) f(t,\mathbf{k}')\left(1 - f(t,\mathbf{k})\right) d\mathbf{k}', \\ &\quad - \int S(\mathbf{k},\mathbf{k}') f(t,\mathbf{k})\left(1 - f(t,\mathbf{k}')\right) d\mathbf{k}',\end{aligned} \tag{5.99}$$

$$\frac{dW_{LO+TO}(t)}{dt} = C^{p-e}_{W_{LO+TO}} - \frac{W_{LO+TO} - W^{\mathrm{LE}}_{LO+TO}}{\tau}, \quad \frac{dW_K(t)}{dt} = C^{p-e}_{W_K} - \frac{W_K - W^{\mathrm{LE}}_K}{\tau}, \tag{5.100}$$

$$\frac{dW_{\mathrm{ac}}(t)}{dt} = C^{p-e}_{W_{\mathrm{ac}}} - \frac{W_{\mathrm{ac}} - W^{\mathrm{LE}}_{\mathrm{ac}}}{\tau}. \tag{5.101}$$

An equation similar to Equation 5.99 holds for the K' valley. As the initial condition for the electrons we take the Fermi–Dirac distribution

$$f(0,\mathbf{k}) = \frac{1}{1 + \exp\left(\dfrac{\mathcal{E}(\mathbf{k}) - \mathcal{E}_F}{k_B T_0}\right)},$$

where T_0 is the room temperature (300 K) and $\mathcal{E}_F$ is the Fermi energy, which is related to the initial charge density by

$$\rho(0) = \frac{4}{(2\pi)^2}\int f(0,\mathbf{k})\, d\mathbf{k}. \tag{5.102}$$

In Equation 5.102, the factor 4 arises because we are considering both the two states of spin and the degeneracy (equal to 2) of the valley. Note that in the unipolar case ρ remains constant, $\rho(t) = \rho(0)$, as a consequence of the charge conservation.

Regarding the phonons, we assume that initially all the phonons are at the room temperature T_0. Therefore, the initial conditions for the phonon average energies are given by Equations 5.88 through 5.90 with $T_{\mathrm{LE}} = T_0$.

Let us introduce a uniform time-step Δt and denote by t_n the nth time level. For each interval $[t_n, t_n + \Delta t]$ we solve Equation 5.99 by a Direct Simulation Monte Carlo scheme (DSMC) by freezing the phonon temperatures at the values they have at $t = t_n$. A crucial question is to properly take into account the degeneracy effects in the scattering terms. Usually, the approach proposed in [64] is adopted. However, by so doing, occupation numbers greater than one can be obtained [65] with the manifest violation of the Pauli exclusion principle. In order to overcome such a difficulty we employ the new DSMC scheme presented in [65], which guarantees physically correct occupation numbers. For the sake of shortness we skip the details and refer the interested reader to the paper quoted earlier.

The remaining Equations 5.100 and 5.101 are discretized by an explicit Euler method with $T_{\rm LE}$ evaluated at the previous time step by solving the nonlinear relation Equation 5.95.

In order to complete the numerical scheme, we have to evaluate the production terms $C^{p-e}_{W_{LO+TO}}$, $C^{p-e}_{W_K}$, and $C^{p-e}_{W_{\rm ac}}$. They represent the rate of variation of phonon energy per unit time and are proportional to the difference between the number of emission and absorption processes per unit time due to the electron–phonon scatterings. Taking advantage of the intermediate results coming from the DSMC part, in each time window $[t_{n-1}, t_n]$ we count, for each phonon species, the number of emission scatterings C^+_η and absorption scatterings C^-_η, $\eta = LO + TO, K$. If N_e is the number of particles used in the MC method, each simulation particle has a statistical density weight given by ρ/N_e. Therefore, we can estimate the phonon energy production term as

$$C^{p-e}_{W_\eta} = \frac{\rho}{N_e\,\Delta t}\hbar\omega_\eta \left(C^+_\eta - C^-_\eta\right), \tag{5.103}$$

similar to the procedure adopted for the simulation of other semiconductors [28,66].

Since for the acoustic transition rate we have assumed the elastic approximation, it follows that $C^{p-e}_{W_{\rm ac}} = 0$.

5.7.3 Numerical Results

Crystal heating due to an electron flow is considered rather important in graphene. It can influence the properties related to the electrical characteristics but can also create hot spots with rather high temperatures and the possibility of damaging the material. The main aim of the realized simulations is to evaluate the heating rate of a monolayer graphene under a constant electric field.

For the DSMC part, $N_e = 10^4$ particles have been used. The time step is set as $\Delta t = 2.5$ fs and a constant phonon relaxation time has been adopted, $\tau = 5$ ps, which is a value already used in the literature, also in consideration of the fact that only the phonons that more strongly interact with electrons, having as said an almost constant frequency, are significantly brought out of local equilibrium. Here we are neglecting the presence of defects, for a study of their influence on the phonon mean free path, see [67]. In [53], a temperature-dependent phonon relaxation time has been adopted but it seems to be valid only at low electric fields. More accurate expressions are still lacking. Several Fermi energies have been considered in order to investigate the dependence of the temperature increasing not only on the applied field but also on the electron density. We have also performed a comparison with the case when all phonons are kept in equilibrium at the room temperature in order to analyze the influence of the crystal heating on the characteristic curves.

First, we analyze the case with $\mathcal{E}_F = 0.3$ eV by considering several applied fields. In Figure 5.9, the evolution of the temperatures of each type of phonons along with $T_{\rm LE}$ is plotted. The latter can be identified as the temperature of the crystal lattice and can therefore be directly related to the measurements. The most energetic phonons are the optical ones, while the least energetic phonons are the acoustic ones. One observes that in the first 5 ps the rise in temperature increases, as expected, with the electric field. Roughly, it seems that the maximum $T_{\rm LE}$ after 5 ps depends on the electric field in a linear way. The temperature variation when $E = 20$ kV/cm is about 100 K, a relevant effect that should require a particular attention

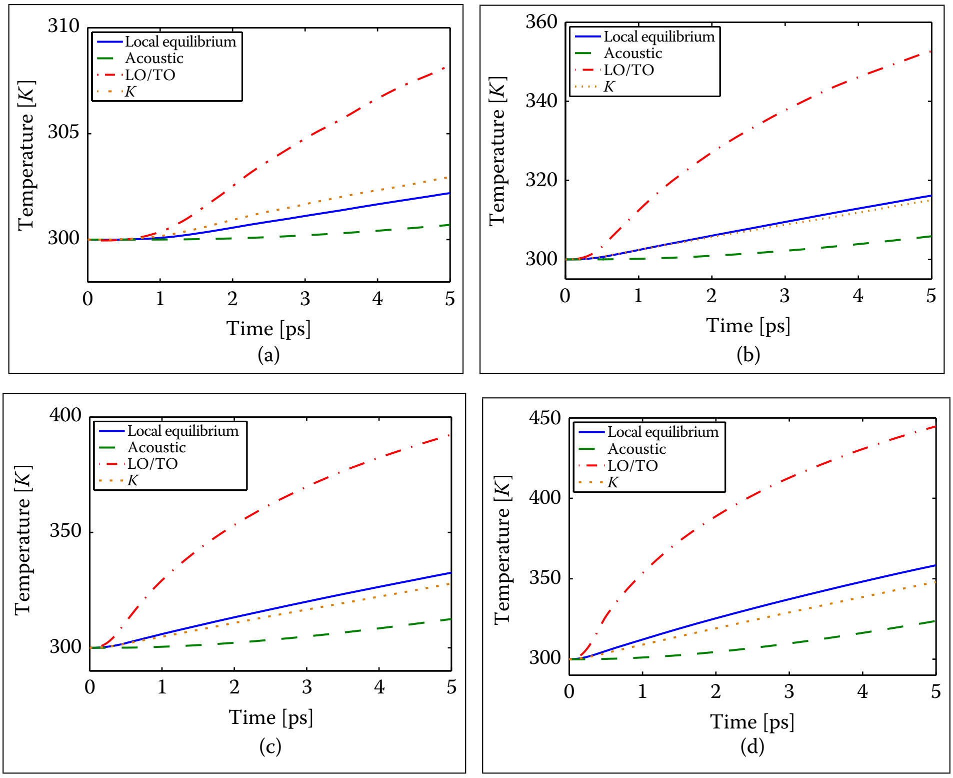


FIGURE 5.9 Phonon temperatures and local equilibrium temperature T_{LE} versus time in the case $\varepsilon_F = 0.3$ eV, when $E = 0.5$ kV/cm (a), $E = 1$ kV/cm (b), $E = 10$ kV/cm (c), and $E = 20$ kV/cm (d).

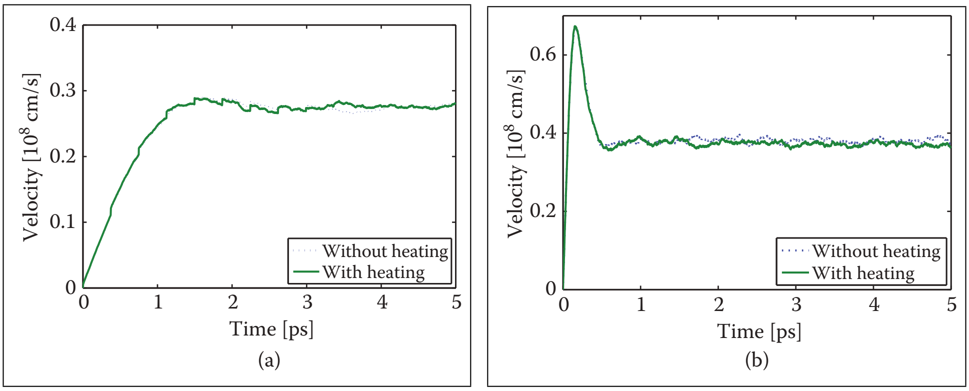


FIGURE 5.10 Average electron velocity versus time in the case $\varepsilon_F = 0.3$ eV, when $E = 1$ kV/cm (a) and $E = 20$ kV/cm (b).

in the design of future electron devices with graphene. In order to study also the influence of the heating effects on the electron transport, Figures 5.10 and 5.11 show a comparison of the average electron velocity and energy in the case when the crystal heating is considered and in the case when the phonons are at room temperature. The differences in the average electron energy and velocity are small up to fields of 20 kV/cm.

The previous cases have also been simulated with a higher Fermi energy $\mathcal{E}_F = 0.6$ eV. Now the heating effects are more evident due to the higher electron current and a consequent greater number of

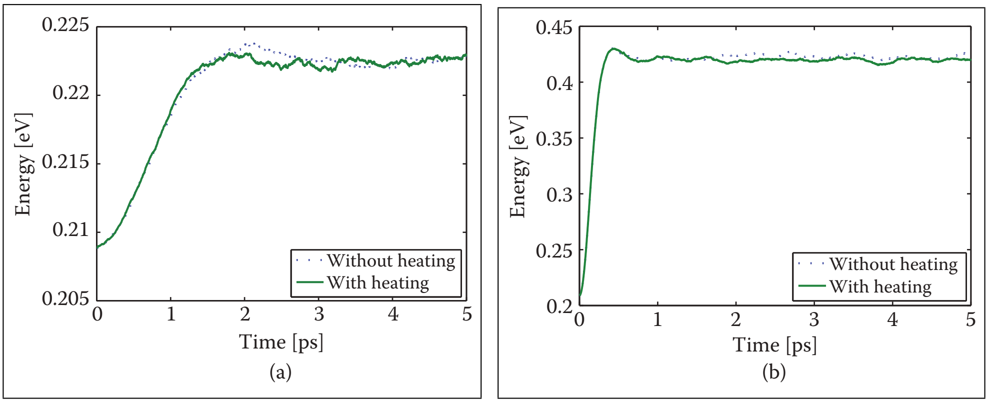


FIGURE 5.11 Average electron energy versus time in the case $\varepsilon_F = 0.3$ eV, when $E = 1$ kV/cm (a) and $E = 20$ kV/cm (b).

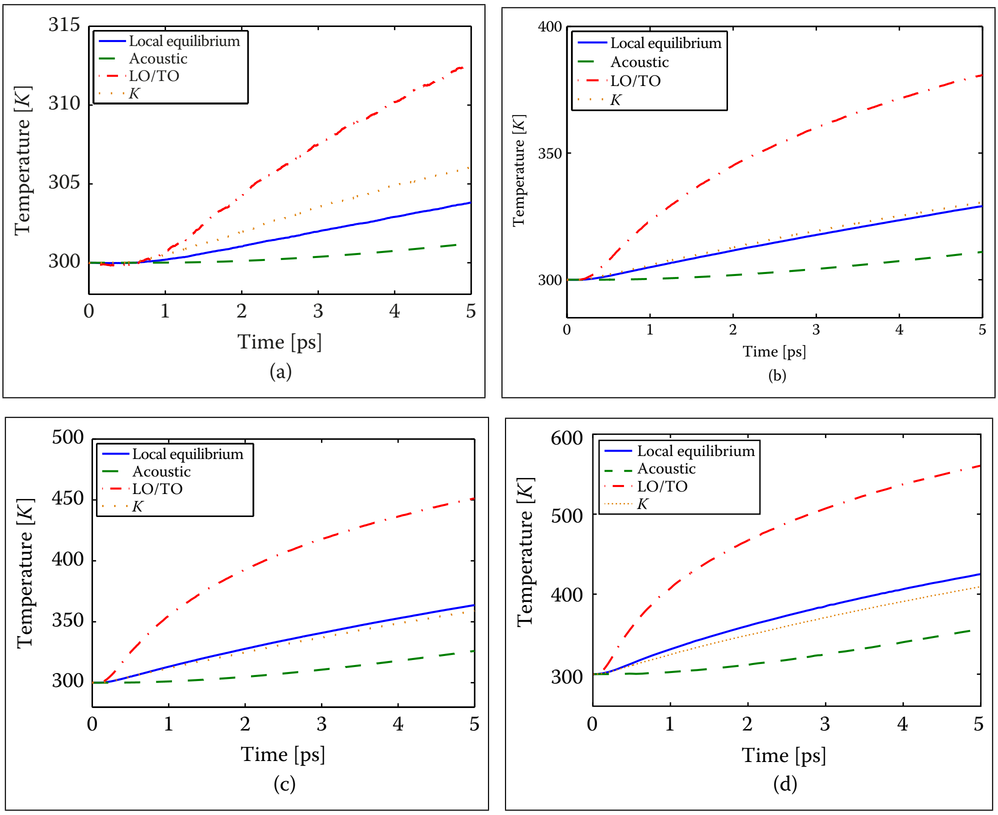
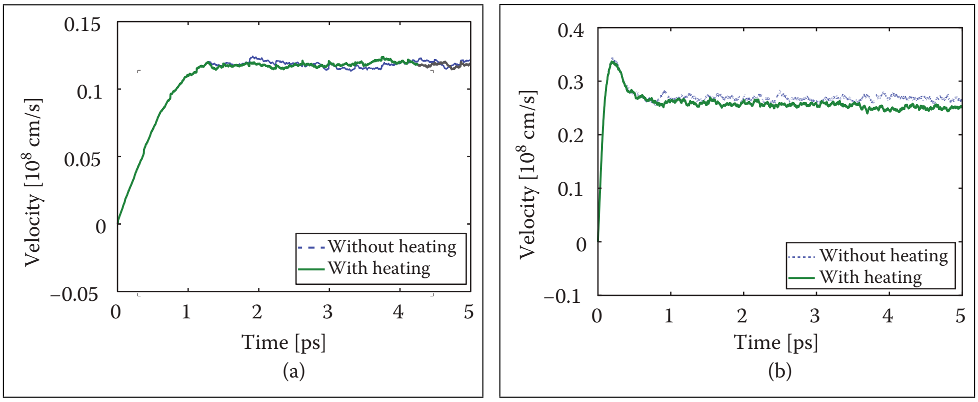


FIGURE 5.12 Phonon temperatures and local equilibrium temperature T_{LE} versus time in the case $\varepsilon_F = 0.6$ eV, when $E = 0.5$ kV/cm (a), $E = 1$ kV/cm (b), $E = 10$ kV/cm (c), and $E = 20$ kV/cm (d).

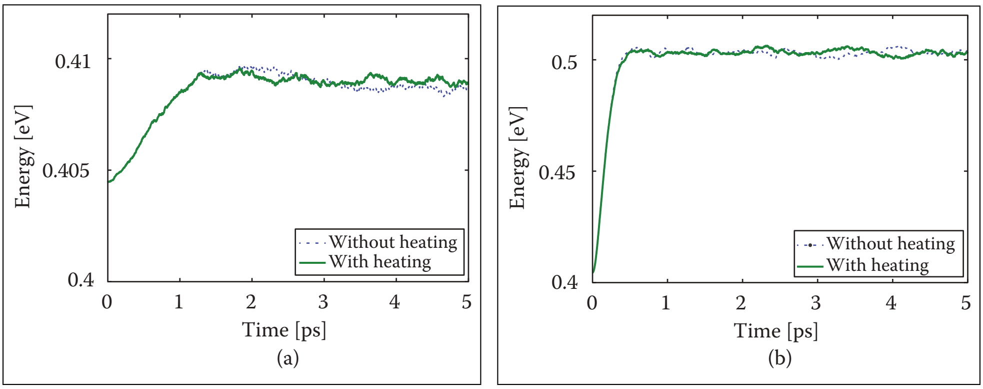
electron–phonon scatterings that transfer more energy to the lattice because the emission processes are dominant with respect to the absorption ones. The results are plotted in Figures 5.12 through 5.14 and show a qualitative trend similar to those with $\mathcal{E}_F = 0.3$. Eventually, in Figure 5.15 it is possible to notice that the effects of crystal heating on electron average energy and velocity become considerable in the presence of high electric fields.



(a)

(b)

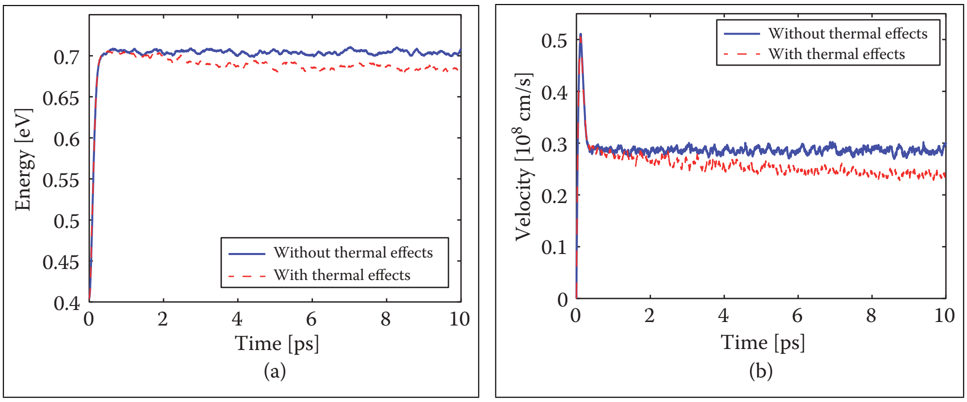
FIGURE 5.13 Average electron velocity versus time in the case $\varepsilon_F = 0.6$ eV, when $E = 1$ kV/cm (a) and $E = 20$ kV/cm (b).



(a)

(b)

FIGURE 5.14 Average electron energy versus time in the case $\varepsilon_F = 0.6$ eV, when $E = 1$ kV/cm (a) and $E = 20$ kV/cm (b).



(a)

(b)

FIGURE 5.15 Average electron energy and velocity versus time in the case $\varepsilon_F = 0.6$ eV, when $E = 50$ kV/cm.

From the analysis of the simulations, it appears that the thermal effects significantly affect the electrical characteristics at high electric fields and cause a considerable rise of the lattice temperature even at medium electric fields and therefore must be taken into account if graphene is employed as semiconductor material in electron devices.

References

1. G.K. Wachutka. Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling, *IEEE Trans. Comput. Aided Design*, **9**, 1141–1149, (1990).
2. D. Jou, J. Casas-Vazquez, G. Lebon. *Extended Irreversible Thermodynamics*. Berlin: Springer-Verlag, (1993).
3. I. Müller, T. Ruggeri. *Rational Extended Thermodynamics*. Berlin/Heidelberg/New York, NY: Springer, (1998).
4. G. Mascali. A hydrodynamical model for silicon semiconductors including crystal heating, *Eur. J. Appl. Math.*, **26**, 477–496, (2015).
5. G. Mascali. A new formula for thermal conductivity based on a hierarchy of hydrodynamical models, *J. Stat. Phys.*, **163**, 1268–1284, (2016).
6. G. Mascali, V. Romano. Simulation of Gunn oscillations with a non-parabolic hydrodynamical model based on the maximum entropy principle, *COMPEL*, **24**(1), 35–54, (2005).
7. G. Mascali, V. Romano. A hydrodynamical model for hole transport in silicon semiconductors: The case of non-parabolic warped bands, *Math. Comp. Mod.*, **53**(1–2), 213–229, (2011).
8. G. Mascali, V. Romano. A hydrodynamical model for hole transport in silicon semiconductors: The case of parabolic warped bands, *COMPEL*, **31**(2), 552–582, (2012).
9. V. Romano. Non parabolic band transport in semiconductors: Closure of the production terms in the moment equations, *Cont. Mech. Thermodyn.*, **12**, 31, (2000).
10. V. Romano. Non parabolic band hydrodynamical model of silicon semiconductors and simulation of electron devices, *Math. Methods Appl. Sci.,* **24**, 439–471, (2001).
11. J.M. Ziman. *Electrons and Phonons: The Theory of Transport Phenomena in Solids*. Oxford: Clarendon Press, (1960).
12. N.W. Ashcroft, N.D. Mermin. *Solid State Physics*. Philadelphia, PA: Sounders College Publishing International Edition, (1976).
13. L.D. Landau, E.M. Lifschitz. *Statistical Physics,* part 1, Vol. 5, Oxford: Pergamon Press, (1980).
14. C. Kittel. *Introduction to Solid State Physics*. New York, NY: John Wiley & Sons, (1996).
15. W.J. Dreyer, H. Struchtrup. Heat pulse experiment revisited, *Continuum Mech. Therm.*, **5**, 3, (1993).
16. Z. Aksamija, U. Ravaioli. Anharmonic decay of g-process longitudinal optical phonons in silicon, *Appl. Phys. Lett.*, **96**, 091911, (2010).
17. P.G. Klemens. Anharmonic decay of optical phonons, *Phys. Rev.*, **148**, 845, (1966).
18. J. Callaway. Model for lattice thermal conductivity at low temperatures, *Phys. Rev.*, **113** (4), 1046, (1959).
19. C. Jacoboni. *Theory of Electron Transport in Semiconductors*. Heidelberg: Springer, (2010).
20. G. Alí, G. Mascali, V. Romano, C.R. Torcasio. A hydrodynamic model for covalent semiconductors with a generalized energy dispersion relation, *Eur. J. Appl. Math.*, **25** (2), 255–276, (2014).
21. E. Pop, R.W. Dutton, K.E. Goodson. Analytic band Monte Carlo model for electron transport in Si including acoustic and optical phonon dispersion, *J. App. Phys.*, **96**, 4998–5005, (2004).
22. C. Auer, F. Schürrer, W. Koller. A semicontinuous formulation of the Bloch–Boltzmann–Peierls equations, *SIAM J. Appl. Math.*, **64**, 1457, (2004).
23. G. Mascali, V. Romano. Charge transport in graphene including thermal effects, *SIAM J. Appl. Math.,* **77**(2), 593, (2017).
24. G. Alí, G. Mascali, V. Romano, C.R. Torcasio. A hydrodynamical model for covalent semiconductors, with applications to GaN and SiC, *Acta Appl. Math.*, **122** (1), 335–348, (2012).
25. P. González, J.A. Carrillo, F. Gámiz, F., Deterministic Numerical Simulation of 1d kinetic descriptions of Bipolar Electron Devices. In A.M. Anile, G. Ali, G. Mascali (Eds.), *Scientific Computing in Electrical Engineering*, Series: Mathematics in Industry, Subseries: The European Consortium for Mathematics in Industry, Vol. 9, pp. 339–344, Berlin: Springer, (2006).

26. J. Piprek. *Semiconductor Optoelectronic Devices: Introduction to Physics and Simulation*. Cambridge, MA: Academic Press, (2013).
27. S.L. Chuang. *Physics of Optoelectronics Devices.* New York, NY: John Wiley & Sons, (1995).
28. E. Pop, S. Sinha, K.E. Goodson. Heat generation and transport in nanometer-scale transistors, *Proc. IEEE,* **94**, 1587–160, (2006).
29. S. Selberherr. *Analysis and Simulation of Semiconductor Devices*. Vienna: Springer-Verlag, (1984).
30. G. Mascali. Maximum entropy principle in relativistic radiation hydrodynamics II: Compton and double Compton scattering, *Continuum Mech. Thermodyn.*, **14**(6), 549, (2002).
31. E.T. Jaynes. Information theory and statistical mechanics, *Phys. Rev.*, **106**, 620, (1957).
32. C.D. Levermore. Moment closure hierarchies for kinetic theories, *J. Stat. Phys.*, **83**, 1021–1167, (1996).
33. K. Wang. *Statistical Mechanics*. New York, NY: John Wiley & Sons, (1987).
34. A.M. Anile, G. Mascali. Theoretical foundations for tail electron hydrodynamical models in semiconductors, *Appl. Math. Lett.*, **14** (2), 245–252, (2001).
35. A.M. Anile, V. Romano. Non parabolic band transport in semiconductors: Closure of the moment equations, *Continuum Mech. Thermodyn.*, **11**, 307, (1999).
36. S. La Rosa, G. Mascali, V. Romano. Exact maximum entropy closure of the hydrodynamical model for Si semiconductors: The 8-moment case, *SIAM J. Appl. Math.*, **70**, 710, (2009).
37. O. Muscato, V. Di Stefano. Hydrodynamic modeling of the electro-thermal transport in silicon semiconductors, *J. Phys. A: Math. Theor.*, **44**, 105501, (2011).
38. O. Muscato, V. Di Stefano. Local equilibrium and off-equilibrium thermoelectric effects in silicon semiconductors, *J. Appl. Phys.*, **110**(9), 093706, (2011).
39. V. Romano. Quantum corrections to the semiclassical hydrodynamical model of semiconductors based on the maximum entropy principle, *J. Math. Phys.*, **48**, 123504, (2007).
40. M. Brunk, A. Jungel, Numerical coupling of electric circuit equations and energy-transport models for semiconductors, *SIAM J. Sci. Comput.*, **30**, 873–894, (2008).
41. V. Romano. 2D simulation of a silicon MESFET with a nonparabolic hydrodynamical model based on the maximum entropy principle, *J. Comput. Phys.*, **176**, 7092, (2002).
42. V. Romano, A. Rusakov. 2d numerical simulations of an electron–phonon hydrodynamical model based on the maximum entropy principle, *Comput. Methods Appl. Mech. Eng.*, **199**, 2741–2751, (2010).
43. V. Romano, M. Zwierz. Electron–phonon hydrodynamical model for semiconductors, *Z. Angew. Math. Phys.*, **61**, 1111–1131, (2010).
44. V. Romano. 2D numerical simulation of the MEP energy-transport model with a finite difference scheme, *J. Comp. Phys.,* **221**, 439–468, (2007).
45. N. Ben Abdallah, P. Degond, S. Genieys. An energy-transport model for semiconductors derived from the Boltzmann equation, *J. Stat. Phys.*, **84**, 205–231 (1996).
46. P. Degond, A. Juengel, P. Pietra. Numerical discretization of energy-transport models for semiconductors with nonparabolic band structure, *SIAM J. Sci. Comput.,* **22**, 986–1007, (2000).
47. G.I. Marchuk, Splitting and alternating direction method. In *Handbook of Numerical Analysis*, Vol. 1, P. G. Ciarlet, J. L. Lions (eds.), North-Holland, Amsterdam: Wiley, pp. 197–462, (1990).
48. H. Zhao, Q. Guo, F. Xia, H. Wang. Two-dimensional materials for nanophotonics application, *Nanophotonics.*, **4**, 28–142, (2015).
49. M. Coco, G. Mascali, V. Romano. Monte Carlo analysis of thermal effects in monolayer graphene, *J. Comput. Theor. Transport*, 45 (7), 540—553, (2016).
50. L. Barletti. Hydrodynamic equations for electrons in graphene obtained from the maximum entropy principle, *J. Math. Phys.*, **55** (8), 083303, (2014).
51. A.H. Castro Neto, F. Guinea, N.M.R. Peres, K.S. Novoselov, A.K. Geim. The electronic properties of graphene, *Rev. Modern Phys.*, **81**, 109, (2009).

52. M. Coco, A. Majorana, G. Mascali, V. Romano. Comparing kinetic and hydrodynamical models for electron transport in monolayer graphene, *Proceedings of the 6th International Conference on Coupled Problems in Science and Engineering*, B. Schrefler, E. Oñate, M. Papadrakis (eds.), Venice: International Center for Numerical Methods in Engineering (CIMNE), 1003–1014, (2015).
53. P. Lichtenberger, O. Morandi, F. Schürrer. High-field transport and optical phonon scattering in graphene, *Phys. Rev. B*, **84**, 045406, (2011).
54. O. Morandi, L. Barletti. Particle dynamics in graphene: Collimated beam limit, *J. Comput. Theor. Transport.*, **43**(1–7), 418–432, (2015).
55. R. Rengel, C. Couso, M.J. Martin. A Monte Carlo study of electron transport in suspended monolayer graphene, *Spanish Conference on Electron Devices (CDE)*, Valladolid, Spain, IEEEXplore, (2013).
56. K.M. Borysenko, J.T. Mullen, E.A. Barry, S. Paul, Y.G. Semenov, J.M. Zavada, M. Buongiorno Nardelli, K.W. Kim. First-principles analysis of electron–phonon interactions in graphene, *Phys. Rev. B*, **11**, 121412(R), (2010).
57. E.N. Koukaras, G. Kalosakas, C. Galiotis, P. Konstantinos. Phonon properties of graphene derived from molecular dynamics simulations, *Sci. Rep.,* **5**, 12923, (2015).
58. D.L. Nika, A.A. Balandin. Two-dimensional phonon transport in graphene, *J. Phys. Condens. Matter,* **24**, 233203, (2012).
59. X. Li, E.A. Barry, J.M. Zavada, M. Buongiorno Nardelli, K.W. Kim. Surface polar phonon dominated electron transport in graphene, *Appl. Phys. Lett.*, **97**, 232105, (2010).
60. V.D. Camiola, G. Mascali, V. Romano. Numerical simulation of a double-gate MOSFET with a subband model for semiconductors based on the maximum entropy principle, *Continuum Mech. Therm.*, **24** (4–6), 417, (2012).
61. V.D. Camiola, V. Romano. Hydrodynamical model for charge transport in graphene, *J. Stat. Phys.*, **157**, 1114–1137, (2014).
62. G. Mascali, V. Romano. A comprehensive hydrodynamical model for charge transport in graphene, *IWCE–2014*, IEEE, Paris, June 3–6, (2016).
63. O. Morandi, F. Schürrer. Wigner model for quantum transport in graphene, *J. Phys. A: Math. Theor.*, **44**, 265301, (2011).
64. P. Lugli, D.K. Ferry. Degeneracy in the ensemble Monte Carlo method for high-field transport in semiconductors, *IEEE Trans. Elect. Dev.,* **ED-32** (11), 2431–2437, (1985).
65. V. Romano, A. Majorana, M. Coco. DSMC method consistent with the Pauli exclusion principle and comparison with deterministic solutions for charge transport in graphene, *J. Comp. Phys.*, **302**, 267–284, (2015).
66. O. Muscato, V. Di Stefano, W. Wagner. A variance-reduced electrothermal Monte Carlo method for semiconductor device simulation, *Comput. Math. Appl.*, **65**: 520–527, (2013).
67. T. Feng, X. Ruan, Z. Ye, B. Cao. Spectral phonon mean free path and thermal conductivity accumulation in defected graphene: The effects of defect type and concentration, *Phys. Rev. B*, 91, 224301-01-12, (2015).

6

Process Simulation

6.1 Introduction 167
6.2 MOCVD Reactor Simulation 167
Basic Transport Equations • Chemical Reactions • Surface–Gas Reactions and Film Deposition • Example of InGaN Multiple Quantum Well Growth Stimulation
6.3 Implant Models 178
Gauss Model • Pearson Model • Dual Pearson Model
6.4 Dopant Diffusion 181
Vacancies • Interstitials • Active Impurities • Inactive Impurities • Neutral Impurities
6.5 Oxidation 184
6.6 Summary 185

Simon Z. M. Li
Changsheng Xia
and
Yue Fu

6.1 Introduction

With further device simulation in mind, device process simulation aims to predict the fabrication process outcome and create the simulation mesh necessary for two-dimensional (2D) or three-dimensional (3D) simulation. The combination of process and device simulation for semiconductor technology modeling is often termed technology computer-aided design (TCAD). The pioneering work in semiconductor process simulation was the SUPREM code created by Stanford University in the 1980s and early 1990s (Hansen and Deal, 1994), mainly concerning silicon process. Commercial versions of SUPREM up to this date are still used. Compound semiconductors are the main materials used in semiconductor optoelectronics applications, and this is the focus of this chapter. Since silicon photonics (in combination with compounds) is an important research subject, we shall describe the basic theories and principles related to silicon as well. It is hoped that the elaborate theories and methods built for silicon will find use in compound semiconductors.

This chapter is organized as follows. Modeling of compound semiconductor epitaxial growth by metalorganic chemical vapor deposition (MOCVD) is described in detail. This is followed by a brief summary of implantation, diffusion, and oxidation as commonly used for silicon processes.

6.2 MOCVD Reactor Simulation

MOCVD is widely used in the epitaxial growth of compound semiconductors such as Gallium nitride (GaN) and related ternaries and quaternaries. Simulation of the MOCVD reactor not only allows understanding of the chemical reactions involved but also offers a convenient computer-aided design (CAD) tool for process optimization. The models and theories in this section have been implemented in the MOCVD reactor simulation program PROCOM from Crosslight Software Inc. (Crosslight Software Inc., 2015) and the results are generated from PROCOM simulation.

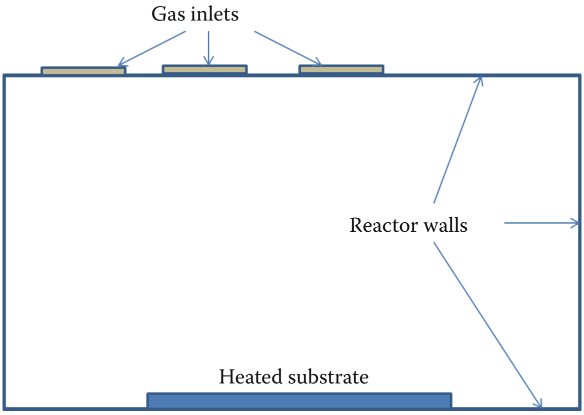


FIGURE 6.1 Schematic of a metalorganic chemical vapor deposition (MOVCD) reactor being simulated. MOVCD, metalorganic chemical vapor deposition.

We consider a simulation domain consisting of 2D or 3D space with gas, surrounded by metal walls. Gas inlets and outlets, and heated substrates are special boundary conditions. The general MOCVD reactor simulation domain is indicated in Figure 6.1. The reactor model aims to model the gas chemical reactions that result in deposition of semiconductor thin films.

6.2.1 Basic Transport Equations

For a reactor simulation program, detailed chemical kinetics and mass/heat transport models can be used to predict the growth rate, composition distribution, thickness uniformity, dopant incorporation, and defect distribution of the semiconductor films if the reactor geometry, chemical species, and growth condition parameters are given.

The processes in an MOCVD reactor belong to a special type of reactive flow: chemical species react with one another in different phases during mass transport, which has been described in the book by Oran and Boris (2001). The basic partial differential equations for the reactive flow processes in the MOCVD reactor are (Oran and Boris, 2001)

$$\frac{\partial \rho}{\partial t} = -\nabla \cdot (\rho v) \tag{6.1}$$

$$\frac{\partial n_i}{\partial t} = -\nabla \cdot \left(n_i v\right) - \nabla \cdot \left(n_i v_{di}\right) + Q_i, \quad i = 1, \ldots\ldots, N_s \tag{6.2}$$

$$\frac{\partial \rho v}{\partial t} = -\nabla \cdot (\rho v v) - \nabla \cdot \mathrm{P} \tag{6.3}$$

$$\frac{\partial E}{\partial t} = -\nabla \cdot (E v) - \nabla \cdot (v \cdot \mathrm{P}) - \nabla \cdot \left(q + q_r\right), \tag{6.4}$$

which represent conservation laws of the total mass, individual species, momentum, and energy, respectively. The basic physical variables are explained in Table 6.1.

In the reactive flow processes, the motion of a species i is usually divided into convective (v) and diffusive (v_{di}) terms:

$$v_i = v + v_{di}. \tag{6.5}$$

TABLE 6.1 Definitions of Basic Variables

Symbol	Definition
$\boldsymbol{E}$	Total energy density (erg/cm^{-3})
$\boldsymbol{n}_i$	Number density of species i (cm^{-3})
$\boldsymbol{v}$	Fluid velocity (cm/s)
$\boldsymbol{N}_s$	Number of chemical species present
$\mathbf{P}$	Pressure tensor (dynes/cm^2)
$\boldsymbol{Q}_i$	Chemical production rate of species i (cm^{-3}/s)
$\boldsymbol{q}$	Heat flux (erg/cm^2/s)
$\boldsymbol{q}_r$	Radiative heat flux (erg/cm^2/s)
$\boldsymbol{v}_{di}$	Diffusion velocity of species i (cm/s)
ρ	Fluid mass density (g/cm^3)

The convection velocity is the mass averaged velocity (or the fluid velocity):

$$v = \frac{\sum_i \rho_i v_i}{\rho}, \tag{6.6}$$

where ρ_i is the mass density of individual species i. It should be noted that Equation 6.6 leads to a zero average diffusion velocity:

$$\frac{\sum_i \rho_i v_{di}}{\rho} = 0. \tag{6.7}$$

So, diffusion is used to describe relative motion of two or more species.

The diffusion velocity between species i and k is related to the change in species number density, pressure gradient, and temperature gradient:

$$\sum\nolimits_k \frac{n_i n_k}{ND_{ik}} \left(v_{dk} - v_{ik}\right) = \nabla\left(\frac{n_i}{N}\right) - \left(\frac{\rho_i}{\rho} - \frac{n_i}{N}\right)\frac{\nabla P}{P} - K_i^T \frac{\nabla T}{T}, \tag{6.8}$$

where N is the total species number density, P is the scalar pressure, and K_i^T is the thermal diffusion coefficient of species i. The diffusion coefficient D_{ik} of a pair of species (i, k) is estimated by the following formula (Hirschfelder et al., 1954):

$$D_{ik} = \frac{2.628 \times 10^{-3}}{P\sigma_{ik}^2 \Omega_{ik}(T_{ik}^*)} \left[\frac{T^3 \left(m_i + m_k\right)}{2m_i m_k}\right]^{1/2}, \tag{6.9}$$

where m is the species mass and the $\Omega_{ik}\,(T_{ik}^*)$ is a collision integral normalized to its rigid sphere value, which in turn depends on the reduced temperature:

$$T_{ik}^* = \frac{k_B T}{\varepsilon_{ik}}. \tag{6.10}$$

The intermolecular parameters σ_{ik} and ε_{ik} can be approximated with collision diameter (σ) and potential parameter (ε) of individual species within the Lennard-Jones potential model:

$$\sigma_{ik} = \frac{1}{2}\left(\sigma_i + \sigma_k\right), \quad \varepsilon_{ik} = \sqrt{\varepsilon_i \varepsilon_k}. \tag{6.11}$$

TABLE 6.2 Some Commonly Used Species

Gas Phase	Surface Phase	Bulk Phase
H_2	H<s>	InP<d>
N_2	In<s>	InN<d>
NH_3	As<s>	AlN<d>
PH_2	SV<s>	GaP<d>
PH_3	AlN<s>	GaN<d>
AsH	PH_2<s>	InAs<d>
AsH_2	AsH<s>	GaAs<d>
$InCH_3$	GaN<s>	
$GaCH_3$	CH_3<s>	
$Ga(CH_3)_3$	$InCH_3$<s>	
$In(CH_3)_3$	$AlCH_3$<s>	
TMG*NH_3	$GaCH_3$<s>	
DMG*NH_2		
(DMG*NH_2)$_3$		

The pressure tensor ($\mathbf{P}$) is expressed in terms of scalar pressure, shear viscosity, and bulk viscosity (Oran and Boris, 2001):

$$\mathbf{P} = P(N, T)\mathbf{I} - \mu_m \left[\nabla v + (\nabla v)^T\right] + \left(\frac{2}{3}\mu_m - \eta\right)(\nabla \cdot v)\,\mathbf{I}, \tag{6.12}$$

where I is the unit tensor, μ_m is the coefficient of shear viscosity for a mixture of species, and η is the bulk viscosity coefficient. The last two terms are the stress tensors which are used to include the diffusive effects of viscosity.

The heat flux can be written as (Oran and Boris, 2001)

$$q = -\lambda_m \nabla T + \sum_i \rho_i h_i v_{\mathrm{di}} + P \sum_i K_i^T v_{di}, \tag{6.13}$$

where λ_m is the thermal conductivity of the species mixture and h_i is the specific enthalpy of species i.

The total energy E consists of the kinetic energy and the internal energy (Oran and Boris, 2001):

$$E = \frac{1}{2}\rho v^2 + E_{\mathrm{int}}, \tag{6.14}$$

where E_{int} is the internal energy, which can be achieved using the temperature dependent enthalpy:

$$E_{\mathrm{int}} = \sum_i \rho_i h_i(T) - P. \tag{6.15}$$

The enthalpy as a function of temperature can be found in many thermal dynamic tables. The reactor simulator, PROCOM, tabulates the enthalpy function for many species in its species macro library so that the simulation program can use it directly. Some commonly used species are listed in Table 6.2.

6.2.2 Chemical Reactions

A full simulation of the MOCVD reactor requires an accurate description of the detailed chemical reactions which are characterized by the chemical production rate Q_i for species i. Consider N_s species involved in

N_r reactions as follows:

$$\sum_{i=1}^{N_s} \nu'_{ji} S_i \Leftrightarrow \sum_{i=1}^{N_s} \nu''_{ji} S_i \; (j = 1, \ldots, N_r), \tag{6.16}$$

where the left-hand side represents reagents and the right-hand side products for either reversible or irreversible reactions. S_is are the reactant and resultant in every reaction process, and ν_{ji} parameters are the mole numbers before them. The production rate Q_i can be expressed in terms of the rate of progress r_j for the jth reaction:

$$Q_i = \sum_{j=1}^{N_r} \left(\nu''_{ji} - \nu'_{ji} \right) r_j, \tag{6.17}$$

where

$$r_j = k_j^f \prod_{k=1}^{N_s} [S_k]^{\nu'_{jk}} - k_j^r \prod_{k=1}^{N_s} [S_k]^{\nu''_{jk}}, \tag{6.18}$$

which is signed such that an excess amount of reagents results in a positive production rate. k_j^f and k_j^r are the forward and reverse reaction rates for the jth reaction, respectively.

The forward reaction rate is usually expressed in the form of a modified Arrhenius function:

$$k_j^f = A_j T^{B_j} \exp\left[\left(-E_j * 4.184 * 1000 \big/ R \right) \big/ T \right] (j = 1, \ldots, N_r), \tag{6.19}$$

where the parameters A, B, and E should be provided from experiments or theoretical calculations. The unit for E is kcal/mol. R is the mole gas constant. *In reactor simulation involving gas reactions, the three numbers A, B, and E are appended to the end of the chemical reaction formulas.*

For example, the gas phase reactions for GaN growth are written as:

$Ga(CH_3)_3$ = $Ga(CH_3)_2$ + CH_3	3.5e+15	0	5.95e+4
$Ga(CH_3)_3$ + NH_3 => TMG*NH_3	1.e+12	0	0
TMG*NH_3 => $Ga(CH_3)_3$ + NH_3	9.5e+9	0.	19
TMG*NH_3 => DMG*NH_2 + CH_4	1.e+13	0	32
3DMG*NH_2 => (DMG*$NH_2)_3$	1.e+21	0	0
(DMG*$NH_2)_3$ => 3GaN<g> + 6CH_4	4.e+15	0	60

For pyrolysis of TMG*NH_3 reaction process, all ν_{ji} parameters are 1, and S_is are $Ga(CH_3)_3$ and CH_3. The last three numbers are the parameters A, B, and E used in Equation 6.19 to determine the gas phase reaction rate. Figure 6.2 shows the reaction rate of every gas phase reaction for GaN growth as a function of temperature, determined by Equation 6.19. The results show that TMG*NH_3 tends to decompose into $Ga(CH_3)_3$ and DMG*NH_2 at high temperature.

For reversible reactions, the reverse reaction rate is related to the forward rate by the equilibrium constant K_j^e:

$$k_j^r = \frac{k_j^f}{K_j^e}, \tag{6.20}$$

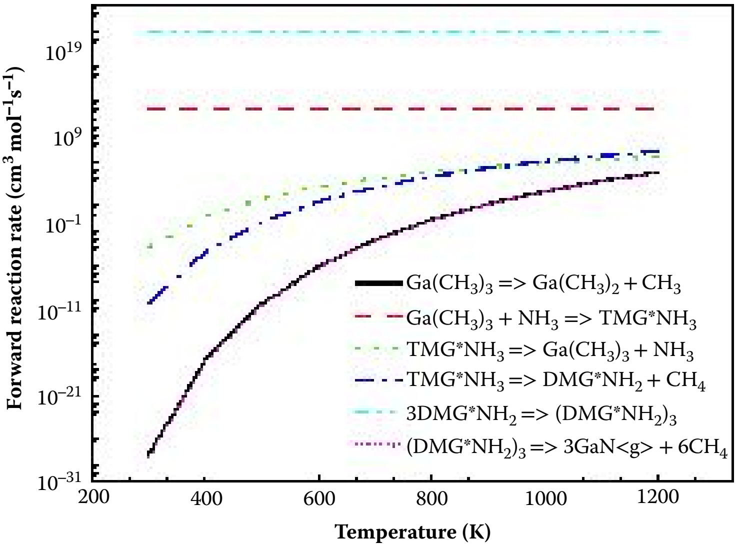


FIGURE 6.2 Reaction rates of every gas phase reaction for GaN growth as a function of temperature.

where the equilibrium constant may be obtained by setting the progress rate in Equation 6.18 to zero under the equilibrium condition (labeled with superscript 0):

$$K_j^e = \prod_{k=1}^{N_s} \left[S_k^0\right]^{v''_{jk} - v'_{jk}} \tag{6.21}$$

By elementary thermodynamics, the equilibrium constant can be calculated by the Gibbs energies:

$$\ln\left(K_j^e\right) = -\frac{1}{RT}\sum_{k=1}^{N_s}\left(v''_{jk} - v'_{jk}\right)G_k^0(T), \tag{6.22}$$

where $G_k(T)$ is the Gibbs energy of pure state species k at standard pressure P_0.

6.2.3 Surface–Gas Reactions and Film Deposition

MOCVD works by letting the chemicals in gas phase be introduced into the reactor chamber to react with one another. The purpose is to allow the gas phase to interact with the substrate to produce solid or bulk deposits of the desirable compound semiconductors. In addition to gas reactions, surface reactions between the gas species and surface vacancy are indispensable to generate the bulk material. Moreover, these reactions are usually irreversible reactions, which can be divided into four different types:

1. Adsorption reaction: A gas species collides with one or two free surface sites forming the adsorbed surface species, and, in the case of dissociative adsorption, releasing gas-phase species as well. Adsorption reaction has the form

   CH_3 + SV<s> => CH_3<s>,

   where SV denotes a surface vacancy and <s> labels a surface species.
2. Abstraction reaction: A gas species reacts with a surface species leading to the release of gas-phase species. Abstract reaction has the form

   CH_3 + H<s> => CH_4 + SV<s>.

3. Recombination reaction: Two surface species react, leading to the formation of gas-phase species. The recombination reaction has the form

   H<s> + CH_3<s> => CH_4 + 2SV<s>.

4. Desorption reaction: A surface or bulk species becomes a gas species. It has the form

   CH_3<s> => CH_3 + SV<s>.

Viscosity and thermal conductivity are very important parameters for the gases and their mixtures in MOCVD. For the most common gases such as H_2, He, and N_2, for which the experimental data on transport properties as a function of temperature are available, the viscosity μ_i and thermal conductivity λ_i usually are expressed in the following form:

$$\lambda_i = A_i T^{b_i}, \quad \mu_i = A_i T^{b_i}, \tag{6.23}$$

where A_i and b_i are parameters determined from fitting the experimental results. Table 6.3 lists the parameters we obtained for H_2, He, N_2, and Ar gases.

However, for the other gas species for which experimental data are not available, we can evaluate the viscosity of the *i*th species by the kinetic theory (Hirschfelder et al., 1954):

$$\mu_i = 8.441 \times 10^{-5} \frac{\left(m_i T\right)^{1/2}}{\sigma_i^2 \Omega\left(T_i^*\right)}, \tag{6.24}$$

and its thermal conductivity is determined by

$$\lambda_i = \left[\frac{15}{4} + 1.32\left(\frac{c_p m_i}{R} - \frac{5}{2}\right)\right] \frac{R}{m_i} \mu_i, \tag{6.25}$$

where c_p is the heat capacity.

For gas mixtures, the following relations are usually used (Wilke, 1950):

$$\mu_m = \sum_i^N \left(\frac{f_i \mu_i}{\sum_j f_j \Phi_{ij}}\right), \tag{6.26}$$

where f is the mole fraction of species, and

$$\Phi_{ij} = \frac{1}{\sqrt{8}}\left(1 + \frac{m_i}{m_j}\right)^{-1/2}\left[1 + \left(\frac{\mu_i}{\mu_j}\right)^{1/2}\left(\frac{m_j}{m_i}\right)^{1/4}\right]^2. \tag{6.27}$$

A similar relation for the thermal conductivity of gas mixture can be used.

TABLE 6.3 Parameters for the Conductivity and Viscosity of H_2, He, N_2, and Ar

	Conductivity (W/m · k)	Viscosity (Kg/m · s)
	A_i b_i	A_i b_i
H_2	1.256e−3 0.8701	1.88e−7 0.676
He	3.96e−3 0.64	5.09e−7 0.64
N_2	2.11e−4 0.836	1.99e−7 0.78
Ar	1.65e−4 0.814	2.14e−7 0.809

6.2.4 Example of InGaN Multiple Quantum Well Growth Stimulation

InGaN multiple quantum well (MQW) light-emitting diodes (LEDs) have attracted considerable attention for general lighting applications and hold promise for replacing the conventional incandescent and fluorescent lamps. However, there are still some problems, which have a significant influence on further improvement of the optical and electrical performances of these LEDs such as film thickness uniformity, epitaxy quality, composition, and impurity distribution. That is to say, growth process decides the performances of the LEDs to some extent. In this section, InGaN quantum well MOCVD growth in a vertical reactor is simulated. The full chemistry model is used to describe the chemical reaction inside the reactor. The temperature, velocity, composition, and mass transport properties are obtained by solving the coupled equations described above.

In the simulation, the vertical reactor used is cylindrical. To take advantage of this symmetry, only half of the reactor is simulated, and the other half is symmetric to the simulated one with the center line being the symmetric axis. A snapshot of the simulated vertical reactor with grid mesh is given in Figure 6.3. In this reactor model, the gas inlet is at the top, the bottom right is the outlet, and the substrate is at the bottom left. The left edge is the symmetric axis and the right edge is set to be the hot wall. The widths for the inlet, substrate, and outlet are 4.98, 3.15, and 1.63 cm, respectively.

In order to grow InGaN quantum wells, the species and their flux which enter the reactor from the inlet at different times should be controlled. This can be done by defining the inlet boundary conditions of the reactor. In this example, the species N_2, NH_3, $Ga(CH_3)_3$, and $In(CH_3)_3$ with the flux of 1359.483 sccm, 453.161 sccm, 1.631 sccm, and 0.05 sccm, respectively, are used to grow InGaN barriers. When growing InGaN wells, the flux of $In(CH_3)_3$ is increased to 0.725 sccm. The growth time for each barrier is set to 20 seconds and the growth time for each well is set to 10 seconds. The simulation starts with a well-mixed gas solution as an initial solution followed by a time-dependent solution of all the equations described previously. By changing the inlet conditions with increased growth time, after 170 seconds, the InGaN MQW structure with five wells and six barriers can be deposited. The detailed input files for the reactor structure, the boundary conditions, chemical reactions, and numerical calculation parameters of this example can be found on the Crosslight website (Crosslight Software Inc., 2015).

Indium composition distribution of the InGaN MQW on the substrate along the X-direction as a function of growth time is shown in Figure 6.4. It can be seen clearly that the indium composition in every barrier and well is very uniform which is mainly relative to the flux of $In(CH_3)_3$ injected into the reactor at different growth times. We also find that the indium composition near the center of the substrate is nearly uniform along the X-direction, but at the edge of the substrate, it is increased. This can be confirmed from

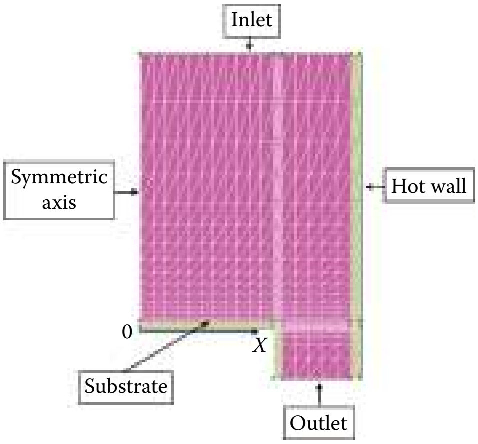


FIGURE 6.3 Snapshot of the simulated vertical reactor with grid mesh.

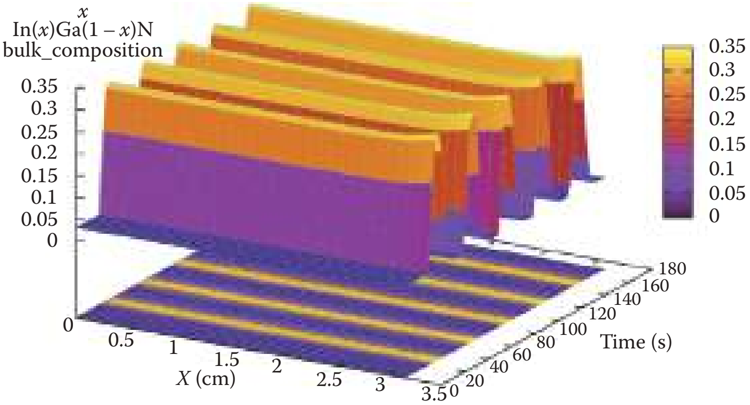


FIGURE 6.4 Indium composition distribution of InGaN multiple quantum wells on the substrate along the X-direction as a function of growth time.

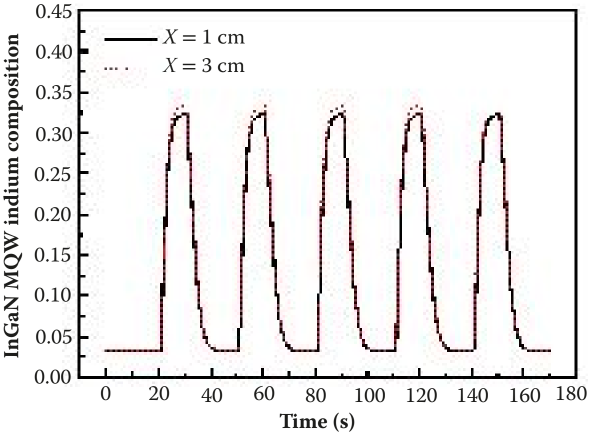


FIGURE 6.5 5 Indium composition distribution of InGaN multiple quantum wells as a function of growth time at $X = 1$ cm and $X = 3$ cm.

Figure 6.5, which shows the indium composition distributions of InGaN MQW as a function of growth time at different positions of the substrate. The indium composition in the barrier is 0.031 and that in the well is 0.32 at $X = 1$ cm. However, at $X = 3$ cm, the indium composition in the barrier is increased to 0.033 and that in the well is increased to 0.33, which indicates that the composition in the wells is more easily affected by the size of the substrate. It should be noted that the indium composition profile at the barrier/well interface is graded and that in the well is curved, which sufficiently reflects the diffusion effects of indium atoms between the barriers and wells. This result implies that graded indium composition should be used for the barrier/well interfaces and the wells in order to get good simulated results in TCAD research of GaN-based LEDs.

Figure 6.6 illustrates the thickness change of InGaN MQW on the substrate along X-direction as a function of growth time. It is obvious that the thickness of InGaN MQW is very nonuniform on the substrate and increase markedly from the center of the substrate to its edge. Figure 6.7 shows the thickness of InGaN MQW at $X = 1$ cm and $X = 3$ cm. As shown in this figure, at $x = 1$ cm, the barrier thickness is 11 nm and the well thickness is 8 nm. However, the barrier thickness is 38 nm and the well thickness is 28 nm at $x = 2$ cm. The total thickness of the MQW is increased about 3.4 times from the center of the substrate to its edge.

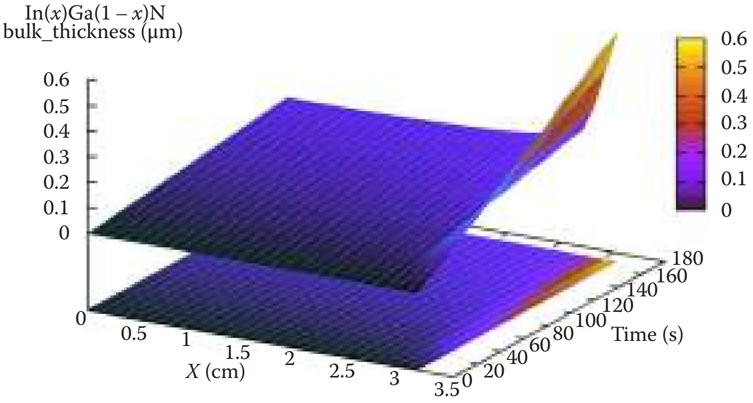


FIGURE 6.6 Thickness of InGaN multiple quantum wells on the substrate along the X-direction as a function of growth time.

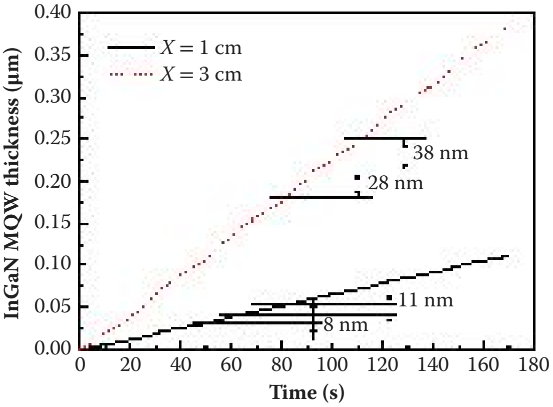


FIGURE 6.7 Thickness of InGaN multiple quantum wells as a function of growth time at $X = 1$ cm and $X = 3$ cm.

The above inhomogeneous distributions of indium composition and thickness on the substrate indicate that most species tend to aggregate at the substrate edge near the outlet. This may be caused by the nonuniform mass transport of the gas species in the reactor. So, in order to grow high-quality InGaN MQW structures on the substrate, it is significantly important to get uniform mass transport of the gas species on the substrate by controlling the flux of injected species, adjusting the pressure and temperature in the reactor, or using a rotating substrate (Cho et al., 1999; Kadinski et al., 2004).

Figures 6.8 through 6.10 illustrate the mass fraction distribution of $Ga(CH_3)_3$, TMG^*NH_3, and DMG^*NH_2 in the reactor, respectively. We can see that the mass fraction of TMG^*NH_3 focused mainly at the gas inlet and those of $Ga(CH_3)_3$ and DMG^*NH_2 are located at the region of the substrate. The results indicate the following gas reaction mechanism: Because the reaction rates are dependent on the temperature, the input gas TMG and NH_3 combine quickly to produce TMG^*NH_3 as soon as they meet at the entrance gas inlet and remain combined. They reach the substrate and then decompose according to the reactions

$$
\begin{aligned}
&TMG^*NH_3 \quad => \quad Ga(CH_3)_3 \quad + NH_3 \\
&TMG^*NH_3 \quad => \quad DMG^*NH_2 \quad + CH_4
\end{aligned}
$$

due to the high reaction rates at the high temperature, which indicates the important role of the TMG^*NH_3 adduct in the growth of GaN material.

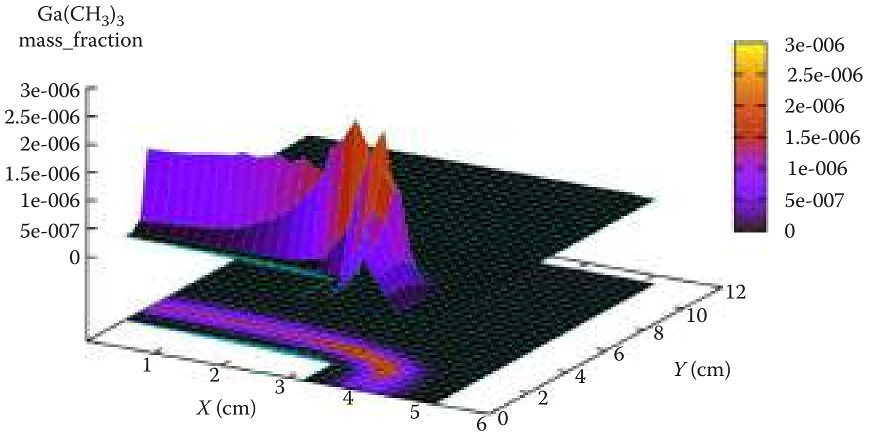


FIGURE 6.8 Mass fraction distribution of $Ga(CH_3)_3$ in the reactor.

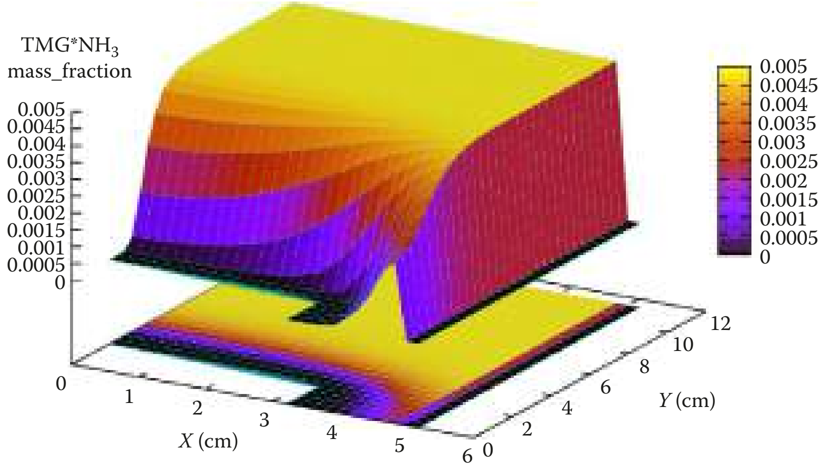


FIGURE 6.9 Mass fraction distribution of $TMG*NH_3$ in the reactor.

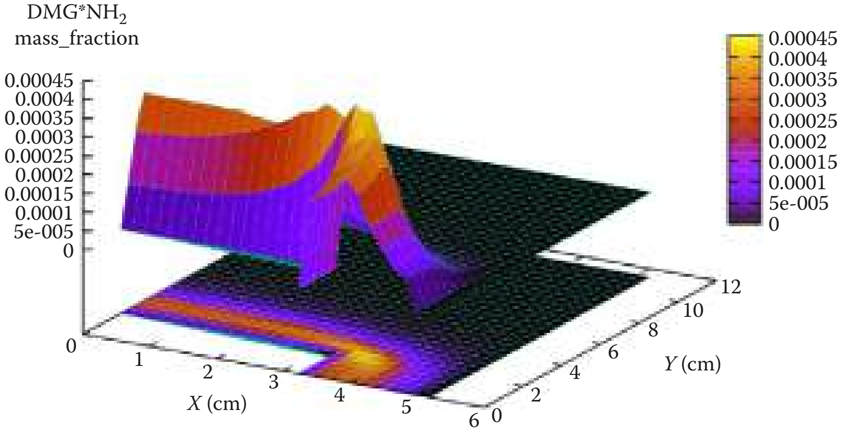


FIGURE 6.10 Mass fraction distribution of $DMG*NH_2$ in the reactor.

In conclusion, the InGaN MQW growth process was simulated using PROCM. The thickness and composition distributions of InGaN MQW were analyzed. The process simulation can provide detailed information about the growth process of multiple layers directly such as layer thickness, layer composition, and their uniformity distribution on the substrate, which are very important for the researchers or the producers to optimize the structure of the LED devices and, therefore, improve the optical and electrical performances of these devices.

6.3 Implant Models

Section 6.2 dealt with epitaxial growth based on physical and chemical models. Ion implantation is, however, the most important method of introducing dopants into semiconductors, especially silicon. This section describes the basics of the ion implantation model based on the SUPREM4 code (Hansen and Deal, 1994) as implemented in the CSUPREM simulation program.

There are two approaches to ion implantation modeling. The most accurate approach is based on atomic scale interaction using Monte Carlo simulation techniques. This is usually well documented for amorphous materials with simple layer structure (Srim, 2015). The second approach is based on analytical fit to experimental secondary ion mass spectrometry (SIMS) data (or Monte Carlo data) to build up a large database. Interpolation is used for ion energies not specifically included in the SIMS database. We shall detail the second approach since this is more efficient and more commonly used in practical TCAD projects.

The theoretical basis for the SIMS-interpolation approach is to decompose the implant scattering source S as a product of two functions:

$$S(x_1) = P(d)\,R(r), \tag{6.28}$$

where x_1 is the entering position (see Figure 6.11 schematic). d is the depth of implant path, and r is the lateral scattering distance. Here we only consider vertical implant entry because any angle in implant may be eliminated by performing a rotation of the coordinate system.

The doping profile at any observation point (x, y) may be expressed as a contribution from lateral scattering of all implant paths coming at the same y-value. That is,

$$D(x, y) = \int P\left(d\left(x_1, y\right)\right) R\left(x_1 - x\right) dx_1. \tag{6.29}$$

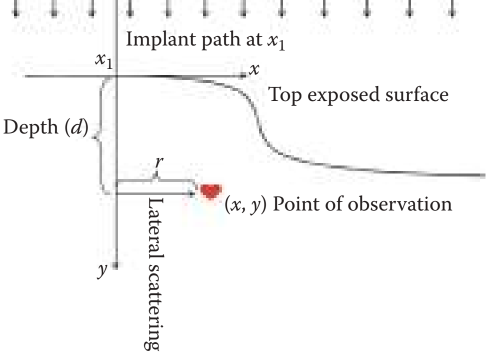


FIGURE 6.11 Schematic for ion implant model.

Please note that the lateral function $R(r)$ is normalized so that in the case of flat exposed surface and uniform material, the doping is simply given by $P(d)$, as expected. The lateral function $R(r)$ has always been assumed to be a simple Gaussian, although some sophisticated process simulator assumes it varies with depth.

Much work in the early days of TCAD focused on how best to use analytical function to construct the depth function $P(d)$ to represent the real SIMS data. Gaussian functions, Pearson functions, and dual-Pearson functions with various parameter settings and sophisticated SIMS fitting strategies have been used.

6.3.1 Gauss Model

Gaussian distribution is the most popular model based on the central moments (Selberherr, 1984). The probability density (or frequency function) $f(x)$ is given by

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma_p} \exp\left[-\frac{\left(x - R_p\right)^2}{2\sigma_p^2}\right], \tag{6.30}$$

where R_p and σ_p are the central moment characteristic parameters, projected range and standard deviation, respectively:

$$R_p = \int_{-\infty}^{+\infty} x f(x)\, dx \tag{6.31}$$

$$\sigma_p = \sqrt{\int_{-\infty}^{+\infty} \left(x - R_p\right)^2 f(x)\, dx}\,. \tag{6.32}$$

6.3.2 Pearson Model

Another approach followed by Hofker (1975) for fitting a frequency function to experimental data is to use a Pearson type IV distribution function. The whole family of Pearson distributions (Gibbons and Mylroie, 1973) is based on the solutions of the Pearson differential equation:

$$\frac{df}{dy} = \frac{y - a}{b_0 + b_1 y + b_2 y^2}, \quad y = x - R_p. \tag{6.33}$$

The Pearson coefficients a, b_0, b_1, and b_2 can be determined by the first four characteristic parameters as follows:

$$a = b_1 \tag{6.34}$$

$$b_0 = -\frac{\sigma_p^2\left(4\beta - 3\gamma^2\right)}{10\beta - 12\gamma^2 - 18} \tag{6.35}$$

$$b_1 = -\frac{\gamma\sigma_p(\beta + 3)}{10\beta - 12\gamma^2 - 18} \tag{6.36}$$

$$b_2 = -\frac{\left(2\beta - 3\gamma^2 - 6\right)}{10\beta - 12\gamma^2 - 18}, \tag{6.37}$$

where γ and β are the other two central moment characteristic parameters, skewness and the kurtosis or excess, respectively:

$$\gamma = \frac{\int_{-\infty}^{+\infty} \left(x - R_p\right)^3 f(x)\, dx}{\sigma_p^3} \tag{6.38}$$

$$\beta = \frac{\int_{-\infty}^{+\infty} \left(x - R_p\right)^4 f(x)\, dx}{\sigma_p^4}. \tag{6.39}$$

The solutions of Equation 6.33 depend on the roots of the equation

$$b_0 + b_1 y + b_2 y^2 = 0, \tag{6.40}$$

which ultimately depend on the values of γ and β. Eight different solutions—Gaussian, Pearson type I, type II, ... and type VII—can be obtained.

If $\gamma = 1$ and $\beta = 3$, we obtain the Gaussian distribution, which is the limited case for all types of Pearson distributions.

If $\gamma \neq 1$ and $\beta < 3 + 1.5\gamma^2$, we get the Pearson type II.

The case when the Equation 6.40 does not have real roots corresponds to Pearson type IV. This case arises when

$$0 < \gamma < 32 \tag{6.41}$$

$$\beta < \frac{39\gamma^2 + 48 + 6\left(\gamma^2 + 4\right)^{3/2}}{32 - \gamma^2}. \tag{6.42}$$

For the description of implantation, only Pearson type IV and type VII distribution generally can be applied. These frequency functions have a single maximum at $y = a$ and decay monotonously to zero on both sides. The type VII distribution is not skewed, which limits its application. The general solution of the Pearson differential Equation 6.33 under the condition 6.41 which characterizes the Pearson IV is given by

$$f(x) = K\left[-\left(b_0 + b_1\left(x - R_p\right) + b_2\left(x - R_p\right)^2\right)\right]^{\frac{1}{2b_2}} \exp\left[-\frac{\frac{b_1}{b_2} + 2a}{\sqrt{4b_2 b_0 - b_1^2}} \operatorname{atan}\left(\frac{2b_2\left(x - R_p\right) + b_1}{\sqrt{4b_2 b_0 - b_1^2}}\right)\right], \tag{6.43}$$

where the normalization constant K is given by

$$\int_{-\infty}^{+\infty} f(x)\, dx = 1. \tag{6.44}$$

If no values for the kurtosis are available, the following universal expression is used:

$$\beta = 2.8 + 2.4\gamma^2. \tag{6.45}$$

6.3.3 Dual Pearson Model

The Pearson type IV involves the first four moments and is thus more accurate for predicting a doping profile in many applications. However, it is not really suitable to predict the typical channeling tail observed when channeling phenomena are present. In this case, the Dual Pearson IV is used,

$$f(x) = N_d\left[(1-\zeta)f_1\left(x, R_p, \sigma_p, \gamma_p, \beta_2\right) + \zeta f_2\left(x, R_p', \sigma_p{}', \gamma_p{}', \beta_2{}'\right)\right], \tag{6.46}$$

where f_1 is the Pearson IV for the surface region and f_2 is the Pearson IV for the bulk region. R_p, σ_p, γ_p, and β_2 are the projected range, standard deviation, skewness, and the kuitosis, respectively, for the first Pearson distribution. Similarly, those with $'$ are for the second Pearson function distribution, describing the channeling tail. More details can be found in the works by Park et al. (1990) and Yang et al. (1995).

The use of various analytical functions originated from limitations in speed and memory of early microcomputers and these analytical functions were the only way to quickly evaluate the implant dose years ago. Moreover, the sophisticated fitting means very little since there is no atomic physics contained in those Pearson functions. However, with today's faster computers, there is no longer such a need. We can simply smooth out the SIMS data or Monte-Carlo data and import them into process simulation as depth profile. Today's fast computers will handle the interpolation easily.

Many process simulators such as Crosslight's CSUPREM have an option to use an implant table that is built using SIMS data from different energies and for different materials. No fitting of dual-Pearson function is needed and one need not worry whether there is any error from fitting analytical functions.

The above discussion is based on implant through a single material. In processing of semiconductors, it is common to implant ions through different materials. For example, a pad oxide layer is commonly formed on top of silicon before implantation. Since the implant depth profiles of the same ions in oxide and silicon are different, it is necessary to consider the situation of implant through multiple layers of materials with different implant properties. The problem is that, given implant profiles of single materials, how to combine the different profiles to form a composite doping profile for a multiple layer material system. Two different methods commonly used in the TCAD industry for such combination are the dose matching and range matching methods. Interested readers are referred to Li and Fu (2011).

6.4 Dopant Diffusion

A key simulation task in conventional TCAD process simulation is to solve the diffusion equations over the simulation mesh to predict the impurity doping profile after thermal processing. This is an essential task for silicon-based materials; the theoretical background has been detailed in Li and Fu (2011). For compound semiconductor material systems, diffusion is not often used and we shall only give a brief overview of the relevant models here with tips and hints when using commercial process simulators.

Impurity diffusion models are often based on the concept of pair-diffusion. These models not only account for the effect impurity interactions via space charge, but also accurately describe the interaction between impurities and lattice point-defects such as interstitials (I) and vacancies (V). The impurity atoms cannot diffuse by themselves and require neighboring point-defects as a diffusing vehicle. It is common to refer to impurities involved in a point-defect–related diffusion mechanism as dopant-defect pairs. We label a dopant A paired with a vacancy V as an AV pair, and a dopant A paired with interstitial I as an AI pair. So, the diffusivities of dopants are represented by the diffusivities of dopant-defect pairs. Impurity diffusion in process simulation includes several types, which we briefly describe.

6.4.1 Vacancies

Vacancies (V) can diffuse with their own diffusion constants and obey the following diffusion equation (Xia et al., 2007):

$$\frac{\partial C_V}{\partial t} = \vec{\nabla}\cdot\left(-\vec{J_V} - \sum_{\mathrm{imp}}\vec{J_{AV}}\right) - R, \tag{6.47}$$

where C_V is the vacancy concentration, and $\vec{J}_{AV}$ is the flux of the dopant-defect pair AV described in a subsequent section. $\vec{J}_V$ is the unpaired vacancy flux, which accounts correctly for the effect of an electric field on the charged portion of the defect concentration and can be written as (Xia et al., 2007):

$$-\vec{J}_V = D_V C_V^* \vec{\nabla} \frac{C_V}{C_V^*}. \tag{6.48}$$

C_V^* is the equilibrium vacancy concentration. R is the recombination of vacancies representing simple interaction between interstitials and vacancies, which can be expressed as

$$R = K_R \left(C_I C_V - C_I^* C_V^* \right), \tag{6.49}$$

where K_R is the bulk recombination coefficient, and C_I and C_I^* are the interstitial and interstitial equilibrium concentration, respectively.

The flux balance boundary condition, which the defects obey, can be described by Hu (1985) as

$$\vec{J}_V \cdot \vec{n} + K_V \left(C_V - C_V^* \right) = g, \tag{6.50}$$

where $\vec{n}$ is the surface normal, K_V is the surface recombination constant, and g is the generation, if any, at the surface.

6.4.2 Interstitials

Interstitials can also diffuse by themselves and obey the diffusion equation expressed by (Xia et al., 2007)

$$\frac{\partial (C_I - C_{ET})}{\partial t} = \vec{\nabla} \cdot \left(-\vec{J}_I - \sum_{\text{imp}} \vec{J}_{AI} \right) - R, \tag{6.51}$$

where C_I is the interstitial concentration, C_{ET} is the number of empty interstitial traps, and J_{AI} is the flux of paired interstitials diffusing with impurity A. J_I is the flux of the unpaired interstitials, which accounts correctly for the effect of an electric field on the charged portion of the defect concentration and can be written as (Sze, 1981):

$$-\vec{J}_I = D_I C_I^* \vec{\nabla} \frac{C_I}{C_I^*}. \tag{6.52}$$

C_I^* is the equilibrium interstitial concentration. R is all sources of bulk recombination of interstitials which is the same as described for vacancies in Section 6.4.1. Like vacancies, interstitial defects obey the following flux balance boundary condition (Hu, 1985):

$$\vec{J}_I \cdot \vec{n} + K_I \left(C_I - C_I^* \right) = g, \tag{6.53}$$

where $\vec{n}$ is the surface normal, K_I is the surface recombination constant, and g is the generation, if any, at the surface.

The trap reaction explains some of the wide variety of diffusion coefficients extracted from several different experimental conditions and is described by Griffin and Plummer (1986) as

$$\frac{\partial C_{ET}}{\partial t} = -K_T \left[C_{ET} C_I - \frac{e^*}{1 - e^*} C_I^* \left(C_T - C_{ET} \right) \right], \tag{6.54}$$

where C_T is the total trap concentration, K_T is the trap reaction coefficient, and e^* is the equilibrium trap occupancy ratio.

6.4.3 Active Impurities

Some dopants diffuse only when they are activated and the diffusion equation depends on the activated concentration of the impurities. Here, activation is used to describe how well the impurity atoms are incorporated into the host lattice. In process simulation, dopant activation is achieved via thermal annealing and this depends on solid solubility and other factors. The total dopant (before or after activation) concentration is labeled C_{T} while the part that has been activated is denoted C_{A}.

For n-type dopants, the following model applies to arsenic and selenium. The diffusion flux densities paired with V and I are given, respectively, by

$$\vec{J}_{\mathrm{AV}} = -D_{\mathrm{V}} C_{\mathrm{A}} \frac{C_{\mathrm{V}}}{C_{\mathrm{V}}^*} \vec{\nabla} \ln \left(C_{\mathrm{A}} \frac{C_{\mathrm{V}}}{C_{\mathrm{V}}^*} \frac{n}{n_i} \right) \tag{6.55}$$

$$\vec{J}_{\mathrm{AI}} = -D_{\mathrm{I}} C_{\mathrm{A}} \frac{C_{\mathrm{I}}}{C_{\mathrm{I}}^*} \vec{\nabla} \ln (C_{\mathrm{A}} \frac{C_{\mathrm{I}}}{C_{\mathrm{I}}^*} \frac{n}{n_i}). \tag{6.56}$$

The p-type dopants boron, beryllium, and magnesium follow similar equations:

$$\vec{J}_{\mathrm{AV}} = -D_{\mathrm{V}} C_{\mathrm{A}} \frac{C_{\mathrm{V}}}{C_{\mathrm{V}}^*} \vec{\nabla} \ln \left(C_{\mathrm{A}} \frac{C_{\mathrm{V}}}{C_{\mathrm{V}}^*} \frac{p}{n_i} \right) \tag{6.57}$$

$$\vec{J}_{\mathrm{AI}} = -D_{\mathrm{I}} C_{\mathrm{A}} \frac{C_{\mathrm{I}}}{C_{\mathrm{I}}^*} \vec{\nabla} \ln (C_{\mathrm{A}} \frac{C_{\mathrm{I}}}{C_{\mathrm{I}}^*} \frac{p}{n_i}). \tag{6.58}$$

In the above, n and p refer to the electron and hole concentrations, respectively, and n_i refers to the intrinsic carrier density.

6.4.4 Inactive Impurities

In some cases, inactive impurities can diffuse when they are paired with point-defects. For example, n-type dopants such as phosphorus and antimony would show this behavior. Their flux density can be defined by

$$\vec{J}_{\mathrm{AV}} = -D_{\mathrm{V}} C_{\mathrm{T}} \frac{C_{\mathrm{V}}}{C_{\mathrm{V}}^*} \vec{\nabla} \ln (C_{\mathrm{T}} \frac{C_{\mathrm{V}}}{C_{\mathrm{V}}^*} \frac{n}{n_i}) \tag{6.59}$$

$$\vec{J}_{\mathrm{AI}} = D_{\mathrm{I}} C_{\mathrm{T}} \frac{C_{\mathrm{I}}}{C_{\mathrm{I}}^*} \vec{\nabla} \ln (C_{\mathrm{T}} \frac{C_{\mathrm{I}}}{C_{\mathrm{I}}^*} \frac{n}{n_i}). \tag{6.60}$$

6.4.5 Neutral Impurities

In process simulation, many charge neutral impurities such as O_2, H_2O, gold, cesium, and germanium usually take part in the diffusion process. Some of them may result in significant influence on the performances of the devices. For example, interdiffusion of germanium in silicon can lead to mobility enhancement in nanoscale CMOS because germanium tends to produce the mechanical stress. Diffusion of charge-neutral impurities can be defined simply by Fick's law:

$$\vec{J}_{\mathrm{T}} = -D \cdot \vec{\nabla} C_{\mathrm{T}} \tag{6.61}$$

$$\frac{\partial C_{\mathrm{T}}}{\partial t} = -\vec{\nabla} \cdot \vec{J}_{\mathrm{T}}. \tag{6.62}$$

6.5 Oxidation

The oxidation model is unique for silicon, and when treated fully, it is a complicated model in process simulation because it involves consideration and coupling of oxidant diffusion, viscous fluid flow, injection of V/I defects, diffusion of existing dopants, and moving oxide/silicon interface.

We shall describe a simple 1D oxidation model proposed by Deal and Grove in the 1960s (Deal and Grove, 1965) in some detail and explain qualitatively how such a simple model can be upgraded to fully coupled 2D/3D treatments.

The idea behind the theory of the Deal–Grove model is that the oxidant is transported from the gas phase into the oxide, then diffuses through the oxide until it reach the oxide/silicon interface where chemical reaction converts the oxidant into oxide while consuming solid bulk silicon. The basic assumption of the Deal–Grove model is that chemical reaction occurs at the oxide/silicon interface instead of at the gas/oxide interface.

Referring to Figure 6.12 and assuming the concentration of oxidant in the gas phase C_g is a known quantity (proportional to gas pressure), the model assumes three density fluxes in the three regions, which can be written as

$$J_g = h(C_g - C_s) \tag{6.63}$$

$$J_o = \frac{D(C_s - C_i)}{X_o} \tag{6.64}$$

$$J_i = kC_i, \tag{6.65}$$

where C_s and C_i are oxidant concentrations at the oxide surface and oxide/silicon interface, respectively. X_o is the oxide thickness. J is the flux density in units of 1/area/s. The first equation is the gas law (Henry's law), and the equation in region two is the Fick's diffusion law, while the third equation is the chemical reaction rate equation.

For the last equation, the reaction must result in oxide deposit, and thus

$$J_i = N_i \frac{dX_o}{dt}, \tag{6.66}$$

where N_i is the number of oxidant molecules per volume.

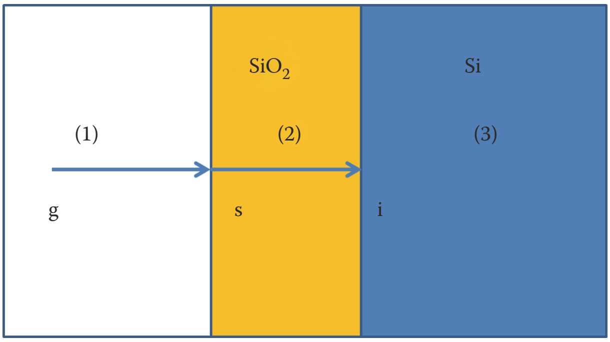


FIGURE 6.12 Schematic showing the three regions in the Deal–Grove theory of oxidation.

Usually the gas concentration is given and if steady-state condition is assumed, $J_{\mathrm{g}} = J_{\mathrm{o}} = J_{\mathrm{i}}$ and the above three equations reduce to

$$h(C_{\mathrm{g}} - C_{\mathrm{s}}) = \frac{D(C_{\mathrm{s}} - C_{\mathrm{i}})}{X_{\mathrm{o}}} \tag{6.67}$$

$$\frac{D(C_{\mathrm{s}} - C_i)}{X_{\mathrm{o}}} = kC_{\mathrm{i}}. \tag{6.68}$$

This enables us to solve for C_{i} and C_{s} after some lengthy algebra. Using Equation 6.66, we obtain

$$\frac{dX_{\mathrm{o}}}{dt} = \frac{B}{A + 2 * X_{\mathrm{o}}}, \tag{6.69}$$

where

$$B = \frac{2DC_{\mathrm{g}}}{N_{\mathrm{i}}} \tag{6.70}$$

$$A = 2D\left(\frac{1}{h} + \frac{1}{k}\right). \tag{6.71}$$

There is an analytical solution to Equation 6.69 for thickness versus oxidation time. It is sufficient to examine the growth behavior here. At small oxide thickness, the growth rate is a constant. As X_{o} becomes thicker, it is harder for the oxidant to reach the interface and growth slows down as it should.

The 1D oxidation model works very well if the device geometry is simple. The only complication would be oxidation enhanced diffusion (OED) in region 3 of Figure 6.12. It is well known that interface chemical reaction is accompanied by injection of V/I defects, which enhance the diffusivity as indicated by flux density formulas we discussed previously.

The real challenge is when the surface is not flat. Oxide behaves like a viscous fluid at high temperature and when the surface is not flat, mechanical stress equations must be solved to predict the shape of the boundary surrounding region 2 in Figure 6.12. Since diffusion and defect injection is dependent on stress, a fully coupled solution of (I, V, oxidant, impurity, and stress) must be solved under the condition of moving boundary. Due to such complexity, most process simulators can handle 2D viscous oxidation well but 3D oxidation is still considered a subject of research. The simulator CSUPREM uses quasi-3D approach with mesh plane stacking in arbitrary orientation to mimic full 3D oxidation for complicated 3D geometries such as deep 3D trench.

As a demonstration of the basic models in silicon process simulation, there is a simplified light-doped drain (LDD) NMOS example. The detailed input files can be found on the Crosslight website (Crosslight Software Inc., 2015). The net doping (defined as the log10 of n-doping minus p-doping) is displayed in Figure 6.13. Please note that the final structure contains three different impurities, As, P, and B, vacancies (V), and interstitials (I). The system contains both the total and the activated components, the latter with a letter a suffix, such as B and Ba.

6.6 Summary

We have given an overview of the theories and principles of semiconductor growth process modeling mainly for compound semiconductors. With the growing importance of compound semiconductor in both optoelectronic and microelectronic applications, we expect MOCVD reactor modeling will gain increasing importance in the near future.

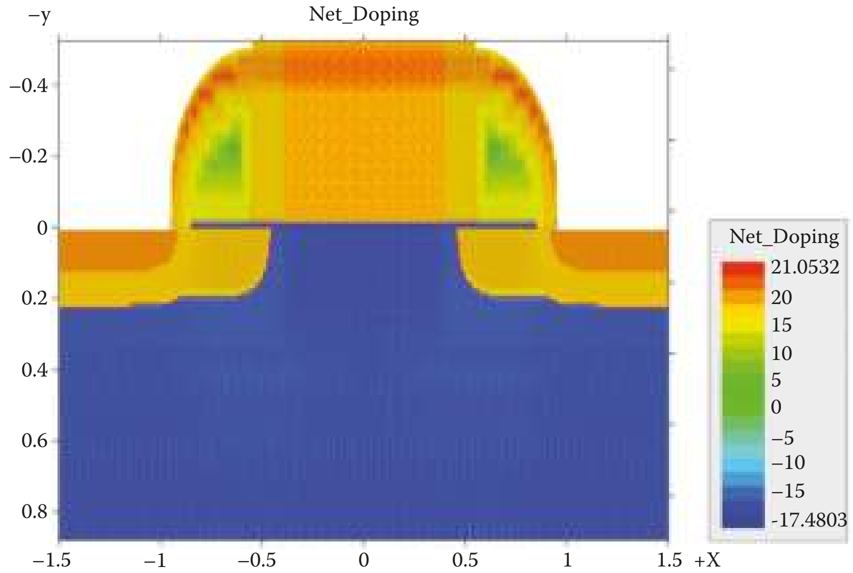


FIGURE 6.13 Results of final net-doping from CSUPREM process simulation for a lightly doped NMOS.

Silicon-related optoelectronic devices, such as CMOS image sensors and hybrid lasers, are important optoelectronic devices. We have presented a brief overview of theories behind commercial process simulators of silicon regarding ion implantation, dopant diffusion, and oxidation. In addition, other processes such as layer deposition, etching, and material-changing are typically implemented in our process simulators as simple geometric manipulations. We hope this chapter about process simulation will help users or potential users of such simulators better understand the physics behind the simulator commands.

References

Cho WK, Choi DH, Kim MU (1999) Optimization of the inlet velocity profile for uniform epitaxial growth in a vertical metalorganic chemical vapor deposition reactor. *International Journal of Heat and Mass Transfer* 42: 4143–4152.

Crosslight Software Inc. (2015) Process Simulation of Compound Semiconductors. http://www.crosslight.com.

Deal BE, Grove AS (1965) General relationship for the thermal oxidation of silicon. *Journal of Applied Physics* 36: 3770–3778.

Gibbons JF, Mylroie S (1973) Estimation of impurity profiles in ion implanted amorphous targets using joined half-Gaussian distribution. *Applied Physics Letters* 22: 568–569.

Griffin PB, Plummer JD (1986) Process physics determining 2-D impurity profiles in VLSI devices. *IEEE International Electron Devices Meeting* 32: 522–525.

Hansen SE, Deal MD (1994) SupremIV.GS, *Two Dimensional Process Simulation for Silicon and Gallium Arsenide*. http://www-tcad.stanford.edu/tcad/programs/suprem-IV.GS/Book.html.

Hirschfelder JO, Curtis CF, Bird RB (1954) *Molecular Theory of Gases and Liquids*. New York, NY: John Wiley & Sons.

Hofker WK (1975) Concentration profiles of boron implanted in amorphous and polycrystalline silicon. *Philips Research Reports Supplements* 8: 41–57.

Hu SM (1985) On interstitial and vacancy concentrations in presence of injection. *Journal of Applied Physics* 57: 1069–1075.

Kadinski L, Merai V, Parekh A, Ramer J, Armour EA, Stall R, Gurary A, Galyukov A, Makarov Y (2004) Computational analysis of GaN/InGaN deposition in MOCVD vertical rotating disk reactors. *Journal of Crystal Growth* 261: 175–181.

Li S, Fu Y (2011) *3D TCAD Simulation for Semiconductor Processes, Devices and Optoelectronics*. New York, NY: Springer.

Oran ES, Boris JP (2001) *Numerical Simulation of Reactive Flow*, 2nd Ed. Cambridge: Cambridge University Press.

Park C, Klein KM, Tasch AF (1990) Efficient modeling parameter extraction for dual Pearson approach to simulation of implanted impurity profiles in silicon. *Solid-State Electronics* 33: 645–650.

Selberherr S (1984) *Analysis and simulation of semiconductor devices*. Vienna: Springer-Verlag

Srim (2015) *The Stopping and Range of Ions in Matter*. http://www.srim.org.

Sze SM (1981) *Physics of Semiconductor Devices*, 2nd Ed. New York, NY: John Wiley & Sons.

Wilke CR (1950) A viscosity equation for gas mixtures. *Journal of Chemical Physics* 18: 517–519.

Xia G, Hoyt JL, Canonico M (2007) Si–Ge interdiffusion in strained Si/strained SiGe heterostructures and implications for enhanced mobility metal-oxide-semiconductor field-effect transistors. *Journal of Applied Physics* 101: 044901(1–11).

Yang SH, Morris S, Parab K, Tasch AF (1995) *An Accurate Monte Carlo Model for the Simulation of Arsenic and Boron Implants into Silicon. UT-MARLOWE Version 2.0*. Austin, TX: University of Texas.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

II

Novel Materials

7 Organic Semiconductors *Roderick C. I. MacKenzie* 191
Introduction to Organic Semiconductors • Electrical and Structural Differences between Organic and Inorganic Semiconductors • Modeling of Electronic Materials • Transport, Trapping, and Recombination • Full Device Models • Geminate and Nongeminate Recombination—Role of Excitons in Organic Devices • Numerical Aspects of Organic Device Simulations • Conclusions

8 Polarization in III-N Semiconductors *Max A. Migliorato, Joydeep Pal, Xin Huang, Weiguo Hu, Morten Willatzen, and Yousong Gu* 219
Introduction • Piezoelectric Effect • Piezoelectricity in Electronic Materials • Mathematical Formalism • Piezoelectric Coefficients of Wurtzite III-N Semiconductors • Electromechanical Coupling in Wurtzite Semiconductors • Nonpolar Orientations in Wurtzite Semiconductors • Summary and Future Directions

9 Dilute Nitride Alloys *Christopher A. Broderick, Masoud Seifikar, Eoin P. O'Reilly, and Judy M. Rorison* 251
Introduction • Theory of the Electronic Structure of Dilute Nitride Alloys • Theory of GaAs-Based 1.3 μm InGaNAs Quantum Well Lasers • Alternative Applications of Dilute Nitride Alloys • Theory of the Density of States and Carrier Transport in Dilute Nitride Alloys • Summary and Conclusions

10 Dilute Bismide Alloys *Christopher A. Broderick, Igor P. Marko, Eoin P. O'Reilly, and Stephen J. Sweeney* 313
Introduction • Electronic and Optical Properties of $GaBi_xAs_{1-x}$ Alloys: Atomistic Theory • Modeling Dilute Bismide Band Structure: The $\mathbf{k}\cdot\mathbf{p}$ Method • $GaBi_xAs_{1-x}$/(Al)GaAs Quantum Wells Lasers • Emerging Directions in Dilute Bismide Research • Summary and Conclusions

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

7

Organic Semiconductors

7.1 Introduction to Organic Semiconductors 191
7.2 Electrical and Structural Differences between Organic and Inorganic Semiconductors 193
7.3 Modeling of Electronic Materials 194
Calculating the Electronic Structure of Organic Semiconductors • Electronic Structure of Organic Semiconductors • Microscopic Charge Transport in Organic Semiconductors • Molecular Dynamics • Purity and Doping of Organic Semiconductors: Where Physicists and Chemists Disagree
7.4 Transport, Trapping, and Recombination 202
Transport, Trapping, and Recombination in Thin Films • Band Structure of Organic Materials • Concept of Mobility in Organic Semiconductors • The Limitations of Microscopic Charge Transport Simulations
7.5 Full Device Models 207
Early Device Models for Organic Semiconductors • Including Carrier Trapping in Organic Device Models • SRH Recombination—Splitting Up Energy Space
7.6 Geminate and Nongeminate Recombination—Role of Excitons in Organic Devices 214
7.7 Numerical Aspects of Organic Device Simulations 215
7.8 Conclusions 216

Roderick C. I. MacKenzie

7.1 Introduction to Organic Semiconductors

The majority of this book focuses on classical crystalline semiconductors such as silicon and gallium arsenide. Indeed, almost all electronic devices we use today from transistors to quantum well lasers (Lim 2007) are based on crystalline semiconductors, and it was the development of these materials and devices made from them that has laid the foundation of the Internet and information age. When one thinks of the atomic structure of silicon, one thinks of a highly ordered, regular arrangement of atoms spanning the physical dimensions of the sample. One could visualize this arrangement of atoms on a more human scale, as marbles of the same size neatly packed in a tin. Bloch's theorem (Bloch 1928) tells us that this regular arrangement of atoms will lead to a well-defined conduction and valance band, with a forbidden region neatly positioned between them called the band gap. Another important characteristic of crystalline semiconductors used in semiconductor devices is that they are highly pure with typical impurity concentrations of one part in ten million (Delannoy 2012). This high purity means that conducting electrons and holes are unlikely to scatter off defects such as oxygen molecules embedded in the lattice. This results in very high-charge carrier mobility values (typically ~1000 $cm^2 V^{-1} s^{-1}$ at room temperature [Canali et al. 1975]). Practically, this high mobility means that very little waste heat is generated as the device is used; this in turn enables millions of transistors to be integrated into tiny chips and the development of highly complex information processing units. From the theoretical perspective, this lack of defect states means that we can fairly successfully imagine conducting electrons and holes in these materials as quasi-free particles that obey classical Newtonian mechanics.

While these crystalline semiconductors have excellent electrical properties, this comes at a cost (New Scientist 1979). To form pure silicon from naturally occurring quartz (SiO_2), one must break the highly stable Si–O bonds and remove all impurities; this requires a sophisticated and highly energetic multistage purification process. The process starts with the liquefaction of silicon to trichlorosilane ($HSiCl_3$) by application of HCl and heating to 300–400°C (Delannoy 2012), followed by the application of hydrogen at 1100°C to further purify the liquid. Numerous distillation steps follow, which end in the conversion of $HSiCl_3$ back into pure silicon (Delannoy 2012); the result is a 99.9999999% pure material. However, due to the high-energy input of this process, the resulting material is an expensive commodity. Indeed, it was recently calculated that if one added up all the energy used to make all silicon solar panels since 1970, then calculated how much energy they all produced, solar cells have only started generating net energy since around 2010 (Dale and Benson 2013). More worrying, almost all of this energy came from burning fossil fuels (Dale and Benson 2013). Therefore just by considering these simple energy arguments, we can think of typical crystalline semiconductors as excellent materials for small high-value electronic items such as CPUs and laser diodes, but not such good materials for large-area devices such as displays, solar cells, and large lighting panels. It is therefore clear that for large-area devices another approach is needed using a different, lower-energy, material system.

Over the last 20 years, the search for a low-cost alternative to Silicon has been intense with a great deal of both academic and industrial attention focused on using carbon-based conducting polymers and small organic molecules as semiconducting materials. Conducting carbon molecules are, however, not new; the synthesis of the first conducting polymer occurred by accident in 1862 while H. Letheby at the College of London Hospital was investigating two cases of fatal poisoning (Letheby 1862). He produced via electrolysis a "brilliant blue" colored pigment, which we know today to be polyaniline (or PANI for short), a well-known conducting polymer (see Figure 7.1).

The modern era of organic electronics began in 1977 when Chiang et al. (1977) reported that when the polymer polyacetylene was doped with iodine, its conductivity increased by a factor of 10 million. This resulted in a Nobel Prize in Chemistry, and the modern age of organic electronics was born. Since 1977, the field of organic electronics has exploded, resulting in a plethora of new devices and materials being developed from carbon-based molecules; indeed, today there is no inorganic device without an organic analog. Examples of such devices are the organic light-emitting diodes (OLEDs) (Yersin 2008), organic field effect transistors (OFETs) (Zhao et al. 2016), biosensors, and organic photovoltaic devices (OPVs) (Tress et al. 2011). In fact, if you have a modern smart phone in your pocket, the screen is almost certainly made from light-emitting diodes fabricated from small organic molecules (such as the one in Figure 7.2); you can tell if the screen is an OLED screen by trying to view the screen from an angle close to 90°: if at angles >85°, the image remains visible and the colors remain true, it's an OLED screen; if not, then it's an older liquid crystal display.

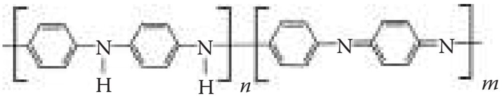


FIGURE 7.1 Polyaniline or PANI for short. The electrical properties of the material can be tuned by varying the ratio of *n* to *m* units. The material H. Letheby found probably had a high number of *m* units, making it blue.

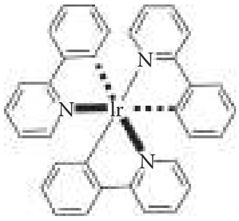


FIGURE 7.2 Tris[2-2,N]iridium(III) or IR(ppy)$_3$ for short, a green-emitting molecule used in organic light-emitting diodes (OLEDs).

The big advantage of using organic semiconductors over inorganic semiconductors is that they can be made using low-energy (read low-cost) wet chemistry and the deposition of these materials to form devices can be done at low temperature (<200°C), saving energy (read money) (Voigt et al. 2012). Another very important property of organic semiconductors is that they are flexible, and if devices are fabricated on flexible substrates, such as plastic, one can produce fully flexible devices which can easily be integrated into clothing and buildings. Although having flexible devices could be attractive from a product integration standpoint (think bendable cell phones), the real advantage of having a flexible device is that when one comes to fabricate it, one can use traditional low-cost, high-speed roll-to-roll printing techniques such as ink-jet printing. Imagine, for example, replacing the newspaper in a newspaper printing press with a flexible plastic substrate (such as indium tin oxide [ITO]-coated polyethylene terephthalate [PET]) and the newspaper ink with an organic polymer dissolved in a solvent to form an ink. One would then be able to fabricate electronic devices as cheaply and as easily as one can produce newspapers. Indeed, such processes have already been used to produce organic solar cells (Voigt et al. 2011, 2012; Adams et al. 2015). In summary, organic semiconductors offer a potential route to low-cost, low-energy, and large-area devices.

7.2 Electrical and Structural Differences between Organic and Inorganic Semiconductors

Although organic semiconductors offer many advantages over their traditional inorganic counterparts, they are intrinsically less ordered than crystalline semiconductors. If one can think of the atomic structure of a crystalline semiconductor, such as silicon, as marbles stacked neatly in a tin, the equivalent picture for an organic material would be a plate of cooked spaghetti, where each piece of pasta represented a polymer chain which can conduct charge along its backbone (see Figure 7.1 for an example of a conducting polymer). In fact, generally speaking, in organic semiconductors, there is no long-range order, with molecules very often being orientated more or less randomly. If we recall, Bloch said that the periodicity of atoms in a semiconductor is the reason for the well-defined conduction and valance band edges (Bloch 1928); we can anticipate that this lack of long-range order has important consequences for how we think of charge transport in organic materials. Indeed, in organic materials, rather than thinking of charge being transported in a band, we think of charge being localized on individual molecules; and charge is only transported from one molecule to another when a charge carrier has gained enough energy to do so, usually through thermal excitation. This process of thermally excited movement is usually referred to as hopping. Another factor limiting conduction of charge carriers in organic materials is that charge can only be transported away from a molecule, if there is another molecule (state) in range on which to hop. If two molecules are packed closely together, it will be relatively easy for charge to transfer from one molecule to the other; however, if the molecules are far apart, it will be less probable that charge will transfer. Thus, in these materials, microscopic packing of the molecules will dictate the macroscopic material mobility. In the next few pages, we explore these concepts in some depth.

7.3 Modeling of Electronic Materials

7.3.1 Calculating the Electronic Structure of Organic Semiconductors

To properly understand charge transport in these organic materials, one must first understand how molecules and charge carriers interact on a quantum level. During our undergraduate studies, we all solved Schrödinger's equation ($E\Psi = H\Psi$) in an infinite potential well to obtain the first three or so wave functions of an electron. The results of such a calculation are presented in pictorial form in Figure 7.3. If instead of solving the equation in a 1D potential well, we change the potential to represent a point charge in space,

$$V = \frac{Q}{4\pi_0 r}, \tag{7.1}$$

such as a single proton, and solved Schrödinger's equation in the 3D space surrounding this point charge, we will again be able to calculate where electrons could exist. The resulting wave functions are in fact the atomic orbitals of atomic hydrogen, which electrons occupy while orbiting the atom. The results from such calculations are depicted in pictorial form in Figure 7.4. By convention, each solution to Schrödinger's equation is labeled 1s, 2s, $2p_x$, $2p_y$, and $2p_z$. Notice, just as in Figure 7.3, the wave functions depicted in Figure 7.4 have positive and negative regions.

The hydrogen atom is the simplest example of applying Schrödinger's equation to calculate the electronic orbitals of an atom. By simply rewriting the Hamiltonian to include the necessary potentials, the same basic approach can be used to calculate the electronic orbitals of other atoms, such as carbon or even whole systems of atoms such as molecules, by including multiple point charges. One could for example use this approach to calculate the delocalized electronic orbitals on polyaniline shown in Figure 7.1 or the small

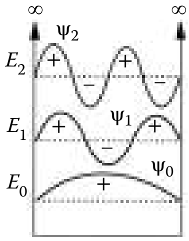


FIGURE 7.3 A diagram of the first three solutions for Schrödinger's equation in an infinite potential well.

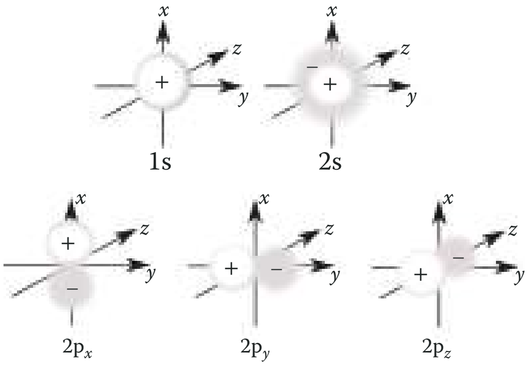


FIGURE 7.4 The 1s, 2s , $2p_x$, $2p_y$, and $2p_z$ orbitals of a hydrogen atom calculated by solving Schrödinger's equation.

molecule shown in Figure 7.2. By taking the square of the calculated wave function $|\Psi|^2$, we can then calculate the probability density function of a charge carrier, which will tell us where we are most likely to find an electron on the polymer/molecule. Clearly, once one moves away from atomic hydrogen with one proton and one electron, one not only has to include the electrostatic potentials of more protons, but also the electrostatic potentials of the other electrons in the system and the influence they will have on the electronic orbital of interest. For a molecular system containing N electrons and M nuclei we can write the general Hamiltonian as

$$H = -\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{A=1}^{M} \frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{r_{iA}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}}, \tag{7.2}$$

where M_A is the ratio of the mass of the nucleus A to the mass of an electron and Z_A is the atomic number of the nucleus. The first term in Equation 7.2 represents the kinetic energy of the electrons, the second term represents the kinetic energy of the nuclei, the third term represents the Coulomb attraction between nuclei and electrons, the fourth term represents electron–electron repulsion, and the fifth term represents nuclei–nuclei repulsion (Szabo 1996). In order to make the problem more tractable, it is common to make the Born–Oppenheimer approximation (Szabo 1996), whereby one assumes that the electrons move much faster than the nuclei so one may consider the nuclei stationary. Finding a solution to the above equation is a major challenge for complex molecular systems with many atomic nuclei because each electron and proton applies a force to all other electrons and protons in the system. One must therefore not only calculate the wave functions of the electrons around the position of known nuclear potentials, but one must also first find the minimum energy positions of the nuclei of the molecule in a 3D space containing the charge clouds, and clearly every update to the positions of the nuclei will affect the distribution of the charge clouds. It is therefore a complex iterative problem requiring significant computing power. The search for efficient methods to do this has led to an entire field called computational quantum chemistry. There are also many complementary methods available today, which are all, in essence, approximations to the solution of Equation 7.2 in one form or another. These include one of the very earliest quantum chemical methods, the Hartree–Fock method (Fock 1930), dating back to 1927, and the more modern and efficient derivatives of such methods such as density functional theory (DFT) (Hohenberg and Kohn 1964; Kohn and Sham 1965), which tries to make the problem more tractable by assuming that electrons do not interact with one another. Writing such codes is not a trivial task, and fortunately today, there are many packages available that have been written to perform these calculations efficiently, such as Gaussian, General Atomic and Molecular Electronic Structure System (GAMESS) and Quantum ESPRESSO. Section 7.3.2 discusses the general results of such calculations and their importance for the field of organic electronics.

7.3.2 Electronic Structure of Organic Semiconductors

We now look at some wave functions (orbitals) of molecules commonly used in organic electronics. In Figure 7.5, a diagram of the ethene molecule is presented where only the wave functions for the 2p-orbitals have been drawn. If one compares these orbitals to those of atomic hydrogen in Figure 7.4, it can be seen that the shape of the orbitals directly above and below the carbon atoms have more or less the same form. It can also be seen from Figure 7.5 that rather than the orbitals being confined to one atom as in the case of atomic hydrogen, they stretch across the carbon bond, forming what is called an MO. A molecular orbital (MO) is the same as an atomic orbital except that wave function spans over a region of molecule rather than just one atom. The MO depicted in Figure 7.6 is called a p-orbital as it is formed from two atomic p-orbitals.

P-orbitals are of significant technological importance for organic electronics because they enable the electron wave function to spread (delocalize) over multiple carbon atoms. This means that if an electron transfers into a very delocalized p-orbital of a molecule, it will spread out in space, which in turn will

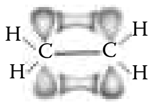
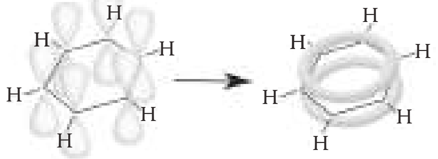


FIGURE 7.5 The wave functions of ethene.



6 atomic p-orbitals

1 delocalized p-orbital

FIGURE 7.6 (a) The six individual atomic p-orbitals of benzene. (b) The six orbitals formed into one large delocalized orbital.

allow it to interact with many of its neighboring molecules. Thus, if an empty state exists on a neighboring molecule, a delocalized electron is likely to find it and therefore hop into it. This means that delocalized MOs are associated with high-charge carrier mobilities and good charge carrier transport.

Another example of a delocalized orbital is that found above and below the carbon ring in benzene (see Figure 7.6), and an even more extreme example can be found in Buckminsterfullerene (often called C_{60} or fullerene. This is depicted in Figure 7.7, where the MO is spread out over the entire ball-shaped carbon cage.

Fullerene is a very important organic electronic molecule that has found uses in solar cells, OLEDs, and sensors. The reason for fullerene's usefulness and popularity can be attributed to its very high electron mobility of 1 $\mathrm{cm}^2\mathrm{V}^{-1}\mathrm{s}^{-1}$ (this mobility is considered high for organic electronic molecules), which can in turn be traced back to its very delocalized p-orbitals. Once an electron hops onto fullerene, it is very quickly delocalized all over the cage. Thus, if any neighboring molecule has an empty electronic state, the electron on the fullerene is highly likely to have some electronic coupling to this state and hop to it. The ball shape of fullerene also helps it pack closely with other fullerene molecules, increasing the probability of charge transfer and thus high mobility.

7.3.3 Microscopic Charge Transport in Organic Semiconductors

Many readers of this chapter will have spent time, while studying quantum mechanics, calculating the probability of an electron transferring from one quantum state to another using Fermi's golden rule. A good example of this problem is calculating the probability of stimulated emission from a quantum well laser diode. In such a case, one first calculates the wave function of an electron in the conduction band of the quantum well (Ψ_i), then calculates the wave function of a hole in the valance band of the quantum well (Ψ_f). One then uses Fermi's golden rule to give the probability of electron transitioning from its initial state to its final state

$$\Gamma = \frac{2\pi}{\hbar} |\langle \Psi_f | H | \rangle|^2 \rho_f, \tag{7.3}$$

where H is the Hamiltonian, $\langle \Psi_f | H | \Psi_i \rangle$ represents the overlap of the initial and final state, ρ_f is the density of final states, and other symbols take their usual meaning. Much the same approach is used to calculate the hopping rate of an electron between two molecules. The framework which is often used is the

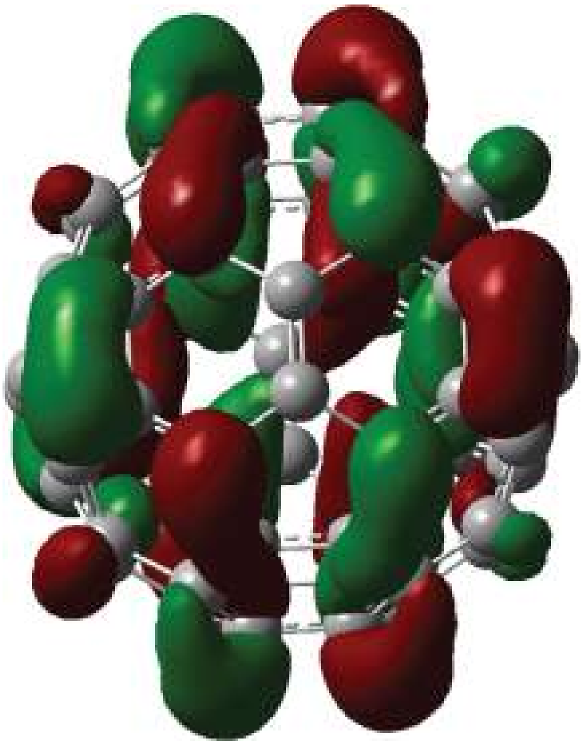


FIGURE 7.7 The delocalized electron orbital of a fullerene cage, Calculated by Dr. Anne A.Y. Guilbert of Imperial College London, using the Gaussian and the hf/6-311g(d,p) method.

semiclassical Marcus theory,

$$\Gamma = \frac{2\pi}{\hbar} | \langle \Psi_{\mathrm{f}} | H | \Psi \rangle |^{2} \exp\left(\frac{-\Delta E + \lambda}{4\lambda \kappa_{B} T} \right), \tag{7.4}$$

for which Rudolph A. Marcus won the Nobel Prize in Chemistry in 1992. If one compares Equations 7.3 and 7.4, one can see that they look more or less identical, except in the case of Marcus theory, $\Psi_{\mathrm{i}}o/\Psi_{\mathrm{f}}$ represent the wave function of the initial and final MO rather than a wave function in a quantum well. On the right-hand side of Equation 7.4, one can see an extra exponential term that does not appear in Fermi's golden rule. This term is in effect an activation energy, which describes the energy difference between the initial and final states, where λ is the reorganization energy, T is the temperature of the molecular system, κ_{B} is Boltzmann's constant, and ΔE is the energy corresponding to the driving force. These terms are discussed in detail later.

To understand how Marcus equation applies to organic molecules, for the moment let's neglect the exponential term on the right-hand side and only consider the overlap of the initial and final states. If we reexamine Figure 7.7 where the highest unoccupied orbital of C_{60} has been depicted, one can see that the MOs will in general hug the bonds between the carbon atoms. If we now consider the pair of fullerene cages in Figure 7.8, we can imagine that in some orientations the p-orbitals on fullerene-a will be well aligned to the p-orbitals on fullerene-b, and in some other orientations the p-orbitals on fullerene-a will be poorly aligned to the p-orbitals on fullerene-b. The former case represents $\langle \Psi_{\mathrm{f}} | H | \Psi_{\mathrm{i}} \approx 1 \rangle$, a high probability of charge transfer, and the latter case represents $\langle \Psi_{\mathrm{f}} | H | \Psi_{\mathrm{i}} \approx 0 \rangle$, a low probability of charge transfer.

This can be seen in Figure 7.9, where fullerene-a has been rotated around the y-axis and fullerene-b held still, and the overlap between the MOs of fullerene-a and -b, $\langle \Psi_{\mathrm{f}} | H | \Psi_{\mathrm{i}} \rangle$, calculated every few degrees. It

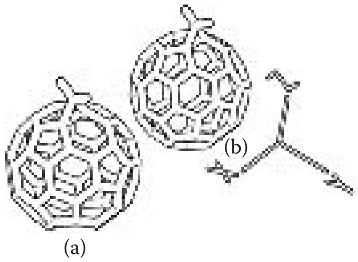


FIGURE 7.8 Two fullerene cages plotted next to each other. They are denoted fullerene-a and fullerene-b in the text. (From MacKenzie et al., *J. Chem. Phys.*, 132, 064904, 2010.)

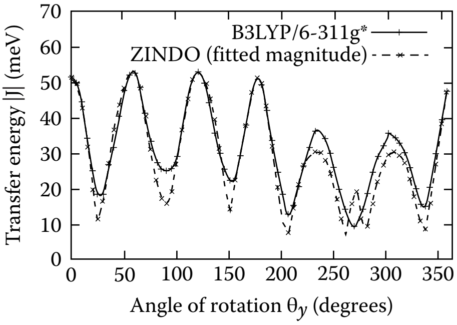


FIGURE 7.9 *J* as a function of rotation around the *y*-axis (see Figure 7.8). (From MacKenzie, R. C. I. et al., *J. Chem. Phys.,* 132, 064904, 2010.)

can be seen that at some angles of rotation, *J* is high and at some other angles *J* is almost zero. This demonstrates how very position-dependent charge transfer is between fullerenes (and all organic molecules in general) and how important it is to have well-aligned p-orbitals to obtain efficient charge transfer. One could imagine a theoretical thin film where for some reason all fullerenes were well aligned, which would have a very high-charge carrier mobility and another theoretical thin film where the fullerenes were poorly aligned resulting in a very low charge carrier mobility. Indeed, when chemists and engineers design new molecules for organic electronic applications, one key consideration influencing design is how likely the molecules are to stack in a way such that the p-orbitals line up to enable high-charge carrier mobilities

Figure 7.10 depicts *J* as a function of distance from the center of fullerene-a to fullerene-b. When the fullerenes are more or less touching (at a center–center distance of 10 Å) a maximum *J* of around 100 meV is calculated. By moving the molecules only 1 Å further apart, charge transfer probability drops by around an order of magnitude. This again highlights the need for close packing of organic molecules and good alignment to obtain efficient charge transfer.

The final thing to note about Figures 7.9 and 7.10, from a computational point of view, is that there are two lines drawn on each figure; one labeled Zerner's intermediate neglect of differential overlap (ZINDO), and one labeled B3LYP/6-311g*. ZINDO is a relatively fast quantum chemical method used to calculate MOs, which uses semiempirical parameters to speed up the calculation (Ridley and Zerner 1973). Each data point on the ZINDO lines took around 10 minutes to compute on an 8-core compute node. In comparison, the B3LYP method is a DFT method and much slower with each data point, taking around 2–3 hours to calculate on an 8-core compute node. One can see from these computational times that simulating any more than a few tens of molecules is computationally too expensive to be done with quantum chemical methods on a regular basis.

Finally, before we leave this topic let's turn our attention to the exponential term in Equation 7.4. The key terms remaining to be explained are λ and ΔE. As stated above, ΔE is the energy difference between the final

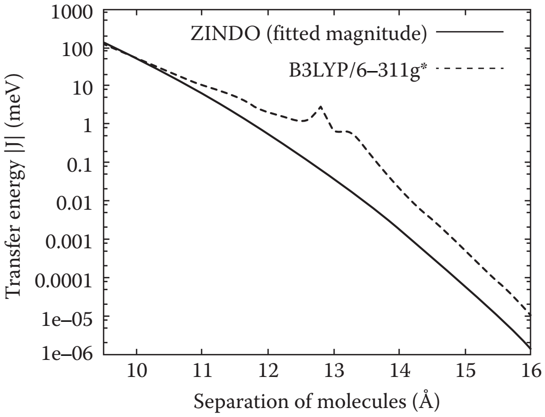


FIGURE 7.10 Transfer energy as a function of center–center fullerene separation.

and initial state. This can be caused by the molecular pair sitting in a potential field or by a dipole being induced between the two molecules. The reorganizational energy, λ, is a little more complex to explain and a full mathematical explanation can be found in Kwiatkowsk (2009). Briefly summarized, the Franck–Condon principle states that charge transfer between two molecules is much more likely if no change in momentum is required for the transition. This means for efficient transfer to occur from one molecule to another both molecules have to be in an energetic state with the same momentum. Usually both states don't have the same momentum, and λ energy is required to move one molecule into a state where no momentum transfer is required for the transition. All the parameters needed for the evaluation of Equation 7.4 can be obtained directly from quantum mechanical calculations (Kwiatkowsk 2009).

7.3.4 Molecular Dynamics

Figure 7.9 demonstrated that charge transport is highly dependent upon molecular orientation and packing. Therefore, when trying to understand microscopic organic materials, it is important to understand how molecules pack, and it is also important to be able to simulate film formation. The computational science that attempts to model how molecules pack together is called molecular dynamics (MD) (Guilbert et al. 2012, 2014, 2015; Dattani 2014). Outside of organic electronics, it is a heavily used tool to understand how biological problems, such as how proteins fold and interact with other molecules such as drugs. The general principle is to try to model the atomic bonds between atoms, and the electrostatic interactions between each molecule. One approach to doing this is by simply solving Schrödinger's equations in time domain and allowing the positions of the nuclei and charge clouds to evolve over time (Jiao et al. 2013); this is called first-principles molecular dynamics (FPMD). This approach, although attractive due to its conceptual simplicity and good approximation to reality, is extremely computationally intensive, requiring a significant cluster of computers (and electricity!) to tackle even the simplest of problems, such as one molecule's interaction with a metallic surface.

A more tractable approach is to first perform accurate quantum chemical calculations on a single test molecule (or pair of test molecules) to understand how a molecule will behave when it is stretched, put under torsion, or encounters the van der Waals attraction forces of another identical molecule, and then to use this understanding based on quantum chemistry to build simple Newtonian models of the molecules of interest. In these models, atoms are represented by Newtonian masses and molecular bonds are represented by Hook-type springs. In doing this, a complex quantum mechanical problem requiring significant CPU time, such as modeling how 200–300 molecules interact to form a film, can be sped up by many orders of magnitude.

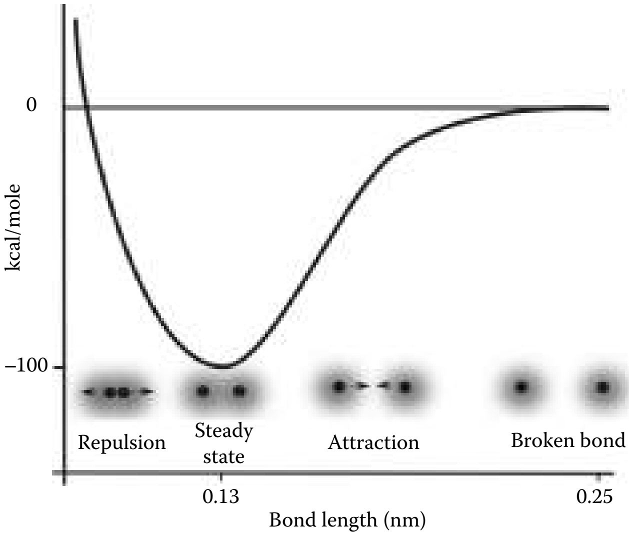


FIGURE 7.11 Potential energy surface for a carbon–carbon bond being compressed, in a relaxed state, and being broken.

Figure 7.11 depicts a typical energy surface that could have been produced by a quantum chemical calculation for the stretching and compression of a C=C double bond in a molecule of interest. One could parameterize this energy surface by fitting it to a standard Lennard-Jones potential to the curve, which is commonly used to parameterize bond energies,

$$V_{\mathrm{LJ}} = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right], \tag{7.5}$$

where ε is the depth of the potential well and σ is the distance at which the finite potential is zero. If one now assumes that all molecular bonds in the molecule of interest can be represented by mechanical springs of form,

$$F = kx, \tag{7.6}$$

where the Hook's constant k can be calculated by taking the second derivative of Equation 7.5, one can start to construct the Newtonian molecular models.

As well as the change in potential energy as a function of bond length, one must also consider angular bond stretching (see Figure 7.12a) and the rotational potential of the bonds (see Figure 7.12b). The potential profile of angular bond stretching in Figure 7.12a is fairly self-explanatory; however, the reason for the hill-shaped potential in Figure 7.12b is less self-evident. It is in fact due to the need to break pi-bonds (see Figure 7.6) as one rotates a C=C bond.

The sets of spring constants, which describe bond stretching and rotation, along with the values used in equations such as Equation 7.5, are called force fields. Very often, one does not need to compute the various spring constants from quantum chemistry calculations directly, as one can use various libraries of precalculated force fields already available in the literature. Examples of these force fields are Assisted Model Building with Energy Refinement (AMBER), Chemistry at HARvard Macromolecular Mechanics (CHARMM), Groningen Molecular Simulation (GROMOS), and Optimized Potentials for Liquid Simulations (OPLS), but there are many more. One should, however, not use these force fields blindly, as the addition of very electronegative ions to a molecule can draw the electron clouds to one end of the molecule and alter how all bonds behave.

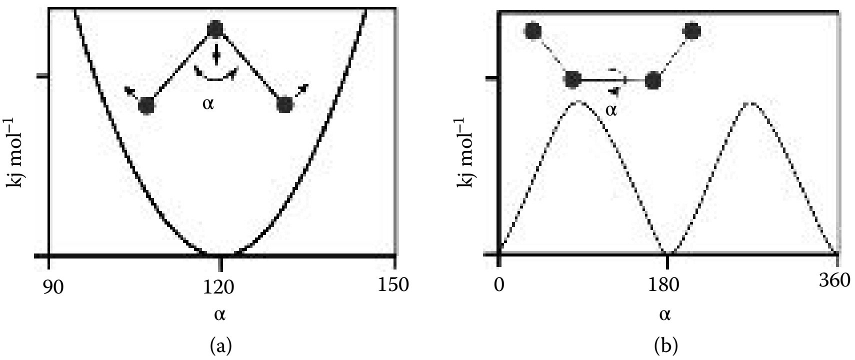


FIGURE 7.12 (a) Angular bond stretching. (b) Rotational potential.

Using MD, one can with relative ease model ensembles of a few hundred molecules without requiring a cluster of computers. To further speed up MD simulations, it is common to use a united atom approach where smaller atoms such as hydrogen are included in the weight of the carbon atoms to which they are attached. It is also possible to approximate entire units of molecules with representative weights; that is, it is common to represent a single fullerene as a single weight of 60 carbon atoms (Chen et al. 2003) rather than simulate each carbon atom in its cage. This is called coarse graining. Another example of coarse graining would be to represent a single monomer unit on a polymer as a single mass with springs connecting it to the other monomers in the polymer (To and Adams 2014). As with quantum chemistry, MD is a large established field with many of the best tools being open source packages, such as GROningen MAchine for Chemical Simulations (GROMACS) (www.gromacs.org n.d.) and Large-scale Atomic/Molecular Massively Parallel Simulator (LAMPS) (LAMPS n.d.). Indeed, the GROMACS manual is a very good starting point for learning about the practicalities of MD simulation.

7.3.5 Purity and Doping of Organic Semiconductors: Where Physicists and Chemists Disagree

Before moving away from detailed materials modeling, it is worth considering what organic semiconducting materials actually consist of. In Section 7.1, I stated that one of the advantages of organic semiconductors was that they can be made by wet chemistry without the need for expensive and energy-consuming purification steps. If one asked a device engineer or physicist what the definition of a pure semiconductor is, he would say a material that has a purity of 99.9999999% (or nine "nines" purity). The atomic density of silicon is around 5×10^{28} atoms m^{-3}; this means in nominally undoped silicon there are 5×10^{19} dopant atoms m^{-3}. In contrast, a highly doped inorganic semiconductor has around 1×10^{25} dopant atoms m^{-3}. If one then went and asked a chemist if the material he had just made is pure, he would say "Yes, it's 99.9% pure." Indeed, if one tries to buy any organic semiconductor, one will very often only be able to obtain 99.9% pure material. Thus, in general, there is a six order of magnitude disagreement between what is considered "pure" in the field of organic and inorganic electronics. It is useful to consider what this means in terms of doping and trap states. If we consider fullerene and assume it is a square box with the volume of 1 nm × 1 nm × 1 nm, it will have a density of 1×10^{27} molecules m^{-3}; if we then assume it is 99.9% pure, we can then calculate that it will have dopant density of 1×10^{24} atoms m^{-3}. Thus, from these simple calculations it can be seen that even a "pure" organic semiconductor is doped almost as much as a highly doped inorganic semiconductor. This simply highlights how contaminated organic semiconductors are when viewed from the inorganic semiconductor standpoint, which is in turn a result of the low-energy production method.

7.4 Transport, Trapping, and Recombination

7.4.1 Transport, Trapping, and Recombination in Thin Films

In Section 7.3.1, we discussed charge transport between two molecules in terms of Markov theory. Now, rather than just considering a pair of molecules, let's consider how charge moves throughout a thin film of organic material. Figure 7.13 depicts in cartoon form a film of organic semiconductor. The round balls represent fullerenes, while the filled-in squares connected by lines represent polymers, with each monomer unit being represented by a single square. Fullerene is a good electron conductor, while many polymers such as poly(3-hexylthiophene-2,5-diyl) (P3HT) (see Figure 7.14) are good hole conductors. It is therefore common to mix fullerenes with polymers to obtain a material which can transport both electrons and holes. An example of a device where such a film is used is an organic solar cell (Deschler et al. 2013)

Let's now consider the transport of an electron from one side of the material to the other. In our cartoon, an electron starts off on the left-hand side of the diagram. It is injected by an electron-injecting contact and hops between fullerenes until it reaches the far side of the film (path-a, indicated by dashed arrows). At this point, the electron will probably recombine with a hole, on the hole-injecting contact. As in an inorganic semiconductor, transport can be driven by drift or diffusion.

Path-a is the optimum route that an electron could take through the medium. Another possible path would be path-b, where after a few hops, instead of the charge proceeding across the material, it starts to travel down a dead end. This is known as a configurational dead end. If a potential were being used to drive charges across the material (see the arrow at bottom of Figure 7.13), there would be a significant

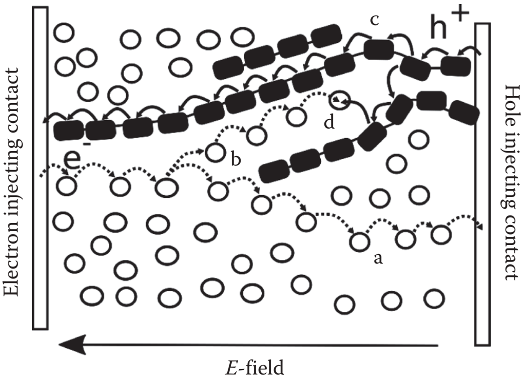


FIGURE 7.13 A cartoon of electron-conducting fullerene (balls) mixed with a hole conducting polymer (square segments). Movement of electrons is represented with dashed arrows, while movement of holes is represented with solid arrows.

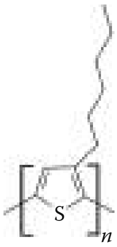


FIGURE 7.14 The hole-conducting polymer poly(3-hexylthiophene-2,5-diyl) (P3HT) is a well-understood polymer commonly used in organic solar cells.

potential gradient between the end of the dead end and the beginning of the dead end. Thus, once a charge had become trapped in the dead end, it would be very improbable that it would be able to escape. This is called an energetic dead end, and the charge carrier would be considered trapped. Now, let's consider the movement of a hole across our device, this time traveling from the right- to the left-hand side of the diagram. The charge carrier is first injected by the hole-injecting contact and hops along a polymer backbone (path-c), traveling from monomer to monomer until it reaches the left-hand side of the diagram where the hole would recombine on the electron-injecting contact. (*Note*: Sometimes when monomer units are well aligned, p-orbitals can delocalize across multiple monomer units. If this happens very fast, transport over multiple monomers can be observed.) Path-c represents an optimum hole transport path across the medium. Now, let's consider the case (path-d), where the hole nears the trapped electron that took path-b. As the hole nears the trapped electron, a finite probability exists for recombination to occur between a quasi-free hole (still moving) and the trapped electron. This free-carrier to trapped-carrier recombination process is a very important way for charges to recombine in organic semiconductors. We will discuss this process in detail later in this chapter.

As the cartoon in Figure 7.13 suggests, understanding trap states in organic semiconductors is very important for understanding both transport and recombination. In fact, in a working organic solar cell well, over 90% of the charge in the device is trapped in trap states with only around 10% being mobile at any one time (MacKenzie et al. 2011). The more disordered a material is (spaghetti like, rather than like marbles in a tin), the more trap states there will be, and the more likely it will be that a charge carrier will become trapped in configurational and energetic dead ends. Charge carrier traps also have a significant influence on net mobility because the more charge-carrier traps that exist in a material, the more probable it will be that a carrier will become trapped at some point while moving across the device, and its movement will be retarded. If the carrier does not detrap after becoming trapped, it will ultimately recombine as seen in Figure 7.13, path-d. Therefore, as a general rule one can say that the more trap states a device has, the higher the recombination rate and the lower the mobility. Very often, organic devices are thermally annealed after fabrication as this helps improve the crystallinity of the films and reduce the number of trap states, increasing mobility and decreasing the recombination rate.

7.4.2 Band Structure of Organic Materials

Figure 7.15a shows the band structure for a typical inorganic semiconductor such as silicon. It has a conduction band (E_c) and a valance band (E_v). A very small number of trap states caused by impurities is shown in the diagram as short black lines in the band gap. Figure 7.15b shows a typical band structure of an organic semiconductor. First, let's deal with the nomenclature, rather than the electron carrying states being referred to as the conduction band as for inorganic semiconductors. In organic semiconductors, they are referred to as the lowest unoccupied molecular orbital (LUMO). Rather than having a valance band, organic semiconductors have a highest occupied molecular orbital (HOMO). In Figure 7.15b, each horizontal line represents a single trap state. It can be seen that both the HOMO and LUMO consist of many individual trap states (typically about 1×10^{26}–1×10^{27} m^{-3}). Each of these trap states originates from a single MO of a molecule. The energetic distribution of these states can be attributed to a variety of factors. To understand this let's consider a film comprised entirely of fullerene molecules. Imagine that, rather than the fullerenes being packed in a film, each molecule is separated from the other by a distance large enough that the molecules cannot electrostatically interact. In this situation, all the LUMO/HOMO levels of fullerene would be identical at 4.5 and 6.1 eV. Now imagine that we slowly bring these fullerene molecules closer together to form our film. As we do this each fullerene cage starts to feel the electrostatic force of its neighboring cages. This electrostatic repulsion/attraction induces dipoles on the molecules which distort the shape of the fullerene cages. Thus, one can imagine that the solutions to Schrödinger's equation will change and so will their eigenvalues, thus the energy levels of the MOs on each molecule will also change. This is one reason for the distribution of trap states as seen in Figure 7.15b. The second reason for this distribution of trap states is the configurational/energetic disorder shown in Figure 7.13.

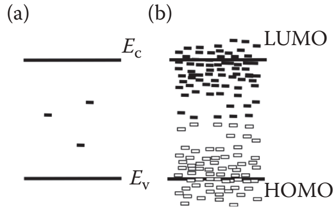


FIGURE 7.15 (a) The band structure of an inorganic semiconductor such as silicon. (b) The "band structure" of an organic semiconductor. Note the density of trap states in an organic semiconductor is typically around 1×10^{26} m^{-3}.

If one considers the distribution of trap states in Figure 7.15b, it can be seen that my cartoon follows no well-described mathematical distribution. This is deliberate as there is some debate in the literature about the exact nature of this energetic distribution. If one opens an undergraduate chemistry book describing the band structure of organic solids, one will often find a Gaussian distribution describing the distribution of trap states around the HOMO/LUMO of the form

$$\rho_{\mathrm{LUMO}}(E) = \frac{N}{\sqrt{2\pi\sigma^2}} \exp\left[-\left(\frac{E - E_{\mathrm{LUMO}}}{2\sigma^2}\right)^2\right], \tag{7.7}$$

where $\rho(E)$ is the density of states (DoS) at any given energy, σ is the spread of the distribution, E_{LUMO} is the center of the distribution (typically taken as the LUMO of an isolated molecule in its ground state), and N is the total number of traps in the distribution. A similar expression can be written for the HOMO. A Gaussian distribution of states is depicted in Figure 7.16a. However, many have found (e.g., To and Adams 2014; Nelson 2003; Kirchartz et al. 2012; Street 2011; MacKenzie et al. 2011) that a Gaussian distribution of traps often cannot account for all of the electrical properties of organic devices and does not always represent the experimentally measured DoS very well (Street 2011). Therefore, when modeling devices, a significant number of authors (Nelson 2003; Kirchartz et al. 2012; MacKenzie et al. 2014, n.d.) have suggested using an exponential distribution of trap states of the form

$$\rho_{\mathrm{LUMO}}(E) = N_{\mathrm{LUMO}}\left(\frac{E}{E^{\mathrm{u}}_{\mathrm{LUMO}}}\right), \tag{7.8}$$

where N_{LUMO} is the density of trap states at the LUMO edge, and $E^{\mathrm{u}}_{\mathrm{LUMO}}$ stands for the Urbach energy and dictates the slope of the energetic distribution of trap states (see Figure 7.16b). If E_{u} is large, the exponential distribution of trap states will slowly decay into the band gap, resulting in a large number of states deep in the band gap. If E_{u} is small, there will be few deep states in the band gap. A slope of around 30 meV is considered a small value of E_{u} and 150 meV is considered a very large value.

In reality, the exact distribution of states is probably a more complex function than a pure Gaussian or a pure exponential; there is evidence for defect states (possibly caused by impurities) deep in the band gap (MacKenzie et al. 2012; Schafferhans et al. 2010; Carr and Chaudhary 2013), meaning the DoS is probably best described as a mixture of an exponential tail and Gaussian deep states (see Figure 7.16c). However, by introducing a complex DoS distribution into any model, one introduces multiple free parameters defining the shape, which may or may not be known. Therefore, it is to some extent impractical to use such complex DoS distributions for day-to-day device simulation. It should also be noted that the exact DoS distribution will be affected by how the molecules pack and the density/type of impurities. Molecular packing will be a function of deposition technique, solvent, annealing steps, and the organic molecule under study, while impurities will be strongly dependent upon which manufacturer produced the semiconductor and from which batch it came. It is therefore hard to recommend a universal DoS function for all organic materials (Gaussian, exponential, or a more complex function), and the modeler should probably use the

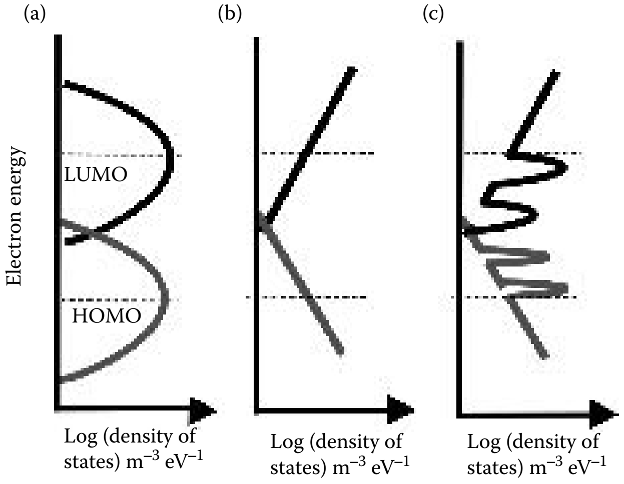

(a)
(b)
(c)
Electron energy
LUMO
HOMO
Log (density of states) m^{-3} eV^{-1}
Log (density of states) m^{-3} eV^{-1}
Log (density of states) m^{-3} eV^{-1}

FIGURE 7.16 (a) A Gaussian distribution of trap states. (b) An exponential distribution of trap states. (c) A combination of both Gaussians and exponentials.

DoS function that fits his experimental data best (Hodgkiss et al. 2014). In my own experience, however, I have never failed to reproduce experimental results by approximating the DoS with a single exponential tail for the HOMO and another for the LUMO (see Figure 7.16b) Finally, it should be noted that here we are only discussing the states in the DoS that are deep enough to influence the electrical properties of the device, such as transport, recombination, and carrier trapping. The higher states may well have more of a Gaussian form (Boix et al. 2009); if one examines Figure 7.3 in this paper, the high-energy states look more Gaussian-like, while the lower-energy states look as if they have a more exponential form.

Before leaving this section, it is worth discussing briefly the emission and absorption models for organic materials. In inorganic devices, once one has the DoS (conduction/valance bands), it is quite common to attempt to calculate absorption and emission for the material using Fermi's golden rule. An example of this is using $\mathbf{k}\cdot\mathbf{p}$ theory to calculate the band structure, and thus emissions from a laser diode. As should be clear from the preceding text, the exact DoS for organic semiconductors is often not well known. Therefore, when modeling devices such as solar cells and photo detectors, this author always tries to get the experimentalists to measure the actual absorption spectra from the exact batch of organic material from which the devices were made.

7.4.3 Concept of Mobility in Organic Semiconductors

If you asked a device engineer to find out what the value of electron mobility was in a silicon wafer, without too much difficulty he/she could use a hall-effect measurement to give you *a single* value of mobility for your material. At the heart of the hall-effect measurement is the definition of mobility, which states electron velocity is equal to mobility times the electric field or

$$v = \mu E. \tag{7.9}$$

However, in organic materials, the situation is more complex because the total time it takes for a charge carrier to travel across a medium is dependent upon the density of trap states into which the charge carrier can become trapped. This is depicted in Figure 7.17 where an electron–hole pair is photogenerated on the left of the device while a potential field is applied across it. After photogeneration, the hole is quickly extracted from the left contact, while the electron travels across the medium. As it travels, it becomes

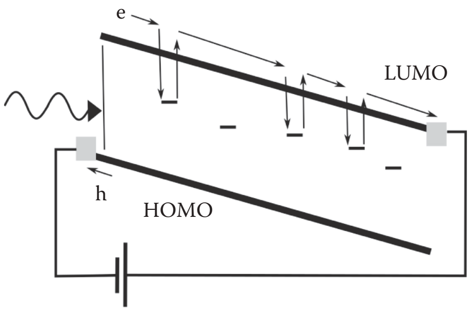


FIGURE 7.17 Photogeneration of an electron hole pair. In this example, the electron undergoes multiple capture escape events into trap states as it travels across the device. The higher the trap density the slower its net speed will be across the device.

trapped and then escapes multiple times. The more time the electron spends being trapped rather than traveling, the longer it takes to cross the device, and the lower the net mobility.

Furthermore, the velocity is dependent not only upon the density of the trap states, but also upon how filled the trap states are. If a trap state is filled, this will reduce the overall probability of a charge carrier becoming trapped (Pauli exclusion principle). Thus, generally speaking, if one deposits an organic semiconductor on a glass slide and keeps it out of the light, the trap states will be mostly empty (broadly speaking and neglecting doping), and thus its mobility will be low. If one now places this same sample in sunlight so charges can be photogenerated, the traps will start to fill and the mobility of the sample will increase. If one now deposits the same semiconductor not on glass but on a conducting transparent metal oxide contact, such as ITO, charge will be injected into the semiconductor from the contact and the semiconductor's mobility will again be higher than when it is deposited on nonconducting glass. If one adds a second contact to the device, say an aluminum top contact to form a diode, more charge carriers will be injected into the device and the mobility will again be increased. Then finally, if one injects or extracts charge by applying a bias voltage, the carrier density within the device will again change, and so will the mobility. Thus when one sees a value of mobility in the literature for an organic semiconductor, one must always consider under what conditions it was measured, and to some extent with which method. One way of defining mobility in organic semiconductors is as follows:

$$\mu_n\left(T_e, E_{fn}\right) = \frac{\int_{E_e^{edge}}^{\infty} \mu_e^0 \rho_e(E) f\left(T_e, E, E_{fn}\right) dE}{\int_{-\infty}^{\infty} \rho_e(E) f\left(T_e, E, E_{fn}\right) dE} = \mu_e \frac{n_{free}}{n_{trap} + n_{free}}, \tag{7.10}$$

where f is the occupation function of a state, ρ is the DoS (Equation 7.8 or 7.7), and μ_e^0 is the mobility of free (not trapped) charge carriers. Equation 7.10 assumes the existence of an imaginary energetic line, or mobility edge E_{edge}, above which all charge carriers are assumed to be free and mobile and below which charge carriers are assumed to be trapped and immobile. Thus by calculating the ratio of movable carriers to nonmovable carriers, one can calculate mobility as a function of charge carrier density. One can understand this equation by considering extreme cases: when all the carriers are highly excited and above energy E_{edge}, the net mobility will be μ_e^0 and when all the carriers are in deep trap states below E_{edge}, the net mobility will be zero. Although, the existence of a sharp mobility edge E_{edge} is somewhat of a theoretical concept, it theoretically accounts for the fact deep carriers will be immobile while highly energetic carriers will be very free. In reality, there will be a gradient of "freeness" between these two extremes. As such the concept of a sharp mobility edge has proved very useful in many models and allows mobility to vary as a function of charge-carrier density as observed experimentally.

7.4.4 The Limitations of Microscopic Charge Transport Simulations

Much of the above discussion has focused on microscopic materials and modeling of organic semiconductors. All these methods focus on how molecules interact on the molecular level and attempt to understand how molecular form affects electronic molecular function. Performing quantum chemical calculations on individual molecules can take between a few seconds to a few hours. Generally speaking, if one wants to efficiently study the motion of molecules and how they pack together one needs to parameterize the results of quantum chemical calculations to form Newtonian MD force fields. Even with simplifying the quantum mechanical problem to a Newtonian problem and the use of coarse graining the computational problem is still significant. To give a general idea of the computational task, to simulate an ensemble of 200 fullerenes for 5 ns, it would take a 16-core compute node approximately a week of compute time. This huge amount of compute time needed to simulate even simple systems means that one cannot model even simple events such as the formation of a film through evaporation because evaporation takes place on the timescale of many seconds and thus occurs on an almost unreachable timescale for these methods. Thus, even with future advances in computing power, it is difficult to see how MD/quantum chemistry can be easily used to simulate whole devices or experimentally significant timescales in the foreseeable future. It should also be noted that setting up efficient MD force fields for a given material system requires significant human input and is therefore time-consuming. If one now considers how the inorganic community models devices, they very rarely attempt to model the individual atoms to understand a device; instead they try to write general equations which approximate the movement of charges and the potential within the device as a whole. Section 7.5 deals with the adaptation of classical device models which have proved so successful for inorganic semiconductors to organic semiconductors.

7.5 Full Device Models

7.5.1 Early Device Models for Organic Semiconductors

The first attempts at device modeling of organic semiconductors borrowed significantly from inorganic device modeling (Koster et al. 2005). Rather than trying to describe the complex molecular structure of the device where electrons were transported in the LUMO of one material (usually phenyl-C61-butyric acid methyl ester [PCBM]) and holes in the HOMO of another material (often a polymer) (see Figure 7.13), the authors of the first device models used an effective medium approximation. In this approximation, one assumed that there is a conduction "band" that spans the entire device. Transport in this band represents charge carriers hopping from one fullerene molecule to another in the LUMO. They also assumed there is a single valance band that spans the entire device and transport in this band accounts for the movement of holes between the polymer chains. Note I put the word band in quotes to highlight that in organic semiconductors band-like transport does not really happen. Transport in these bands was modeled using the standard drift diffusion equations for electrons,

$$J_{\mathrm{n}} = q\mu_{\mathrm{e}} n \frac{\partial E_{\mathrm{LUMO}}}{\partial x} + qD_{\mathrm{n}} \frac{\partial n}{\partial x}, \tag{7.11}$$

and for holes,

$$J_{\mathrm{p}} = q\mu_{\mathrm{h}} p \frac{\partial E_{\mathrm{HOMO}}}{\partial x} - qD_{\mathrm{p}} \frac{\partial p}{\partial x}. \tag{7.12}$$

Poisson's equation describes how space charge alters the potential profile within the device

$$\frac{d}{dx} \cdot \varepsilon_0 \varepsilon_{\mathrm{r}} \frac{d}{dx} \varphi = q\left(n - p\right), \tag{7.13}$$

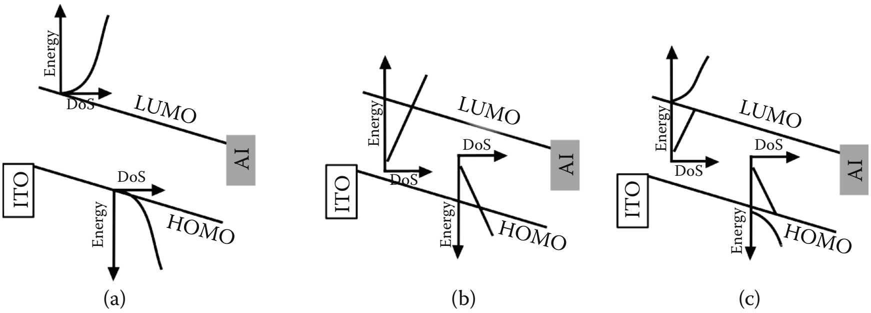


FIGURE 7.18 (a) A "band" diagram of the first organic solar cell device models using a parabolic band approximation with no traps, (b) the introduction of an exponential tail of states into device models. (c) device models based on Shockley–Read–Hall (SRH)—trapping and recombination—these have proved successful.

where ε_0 and ε_r represent the permittivity of free space and the relative permittivity of the medium, respectively, and φ represents the electrostatic potential. Such a model is depicted in Figure 7.18a. Recombination in these early device models was modeled assuming that the recombination rate is proportional to the product of the electron and hole density

$$R = knp \tag{7.14}$$

and that the rate constant could be described by a rate constant k defined as

$$k = q\left(\frac{\mu_n + \mu_p}{2}\right). \tag{7.15}$$

This description of recombination is called Langevin recombination and assumes that electrons and holes both move by Brownian motion (hence the dependence of the equation on mobility) and are annihilated when they meet.

If one contrasts this description of an organic electronic device to the general description of organic semiconductors in the first half of the chapter, one can identify a few major flaws in the model. The first flaw is that band-like transport is assumed and hopping transport via localized states is completely neglected. Furthermore, bearing in mind around 90% of the charge in organic semiconductor resides in very immobile trap states, the result is that the model will never reproduce the correct charge density, and thus the recombination rate will always be wrong (see Equation 7.14). Furthermore, the model can also never account for the change of mobility as a function of carrier density (see Equation 7.10) because mobility is a constant in this model. And more fundamentally, Langevin recombination itself assumes that electrons and holes are free and can participate in a Brownian type motion, with recombination only occurring when the charge carriers get close enough that they are sucked into each other's potential. Brownian type motion neglects carrier trapping (see Figure 7.13). It is therefore clear that to have a better device model carrier trapping must be considered.

7.5.2 Including Carrier Trapping in Organic Device Models

In an attempt to rectify the problems presented by the effective medium model, some authors (MacKenzie et al. 2011) attempted to replace the parabolic band structure with an exponential band structure starting deep in the band gap and continuing past the mobility edge (Roichman and Tessler 2002), as shown in Figure 7.18b. For simplicity, the occupation probability of this DoS was modeled with a single Fermi–Dirac distribution (Roichman and Tessler 2002). This assumed that all carriers in the DoS were able to exchange energy to relax into a distribution with a single quasi-Fermi level. If one considers the arguments above about carriers becoming trapped in deep states, it is clear that this is an over-simplification as there is no obvious reason why two deeply trapped carriers should easily be able to exchange energy However, by assigning carriers above the mobility edge a finite mobility and those below the mobility edge a mobility of zero as done in Equation 7.10, the model did allow the mobility to vary as a function of carrier density. Recombination in this model was accounted for using the Langevin expression given in Equations 7.14 and 7.15; however because now the mobility is allowed to vary as a function of carrier density (Equation 7.10), so does the Langevin prefactor (k), as was observed experimentally (see Figure 7.19),

$$k\left(n,p\right)=\frac{q\left(\mu_{\mathrm{e}}(n)+\mu_{\mathrm{h}}\left(p\right)\right)}{2_{0\mathrm{r}}}. \tag{7.16}$$

Thus this model represented an improvement over the standard Langevin effective medium models which assumed parabolic bands and a constant recombination prefactor (Braun 1984; Koster et al. 2005). It also proved conceptually easy to understand and quite successful in reproducing steady-state current–voltage curves. Incidentally, it is possible to show that Equation 7.16 is comparable to Shockley–Read–Hall (SRH)-type recombination, where free carriers are annihilate trapped carriers (MacKenzie et al. 2011)

Because this book is about numerical modeling, I will now make a few notes about some practical aspects of implementing this model. With this model, it is important to note that because mobile charge carriers no longer have parabolic bands (see Figure 7.18b) the standard Einstein relation,

$$D=\mu_{\mathrm{n}}\frac{k_{\mathrm{B}}T}{q}, \tag{7.17}$$

no longer holds. This is because in the derivation of this relation, the assumption of parabolic bands is made (Azoff 1986; Roichman and Tessler 2002). Instead, the full form of the relation must be used namely

$$D=\frac{\mu_{\mathrm{q}}n}{q(dn/dF_{\mathrm{n}})} \tag{7.18}$$

Evaluating both n and dn/dF_{n} requires numerical integration over energy space, which can lead to errors in the evaluation of D, if one is not careful. If numerical errors enter into D, the drift and diffusion currents will not match each other at equilibrium and dark current–voltage curves will not go through zero. The other thing to note is that quite a number of integrals over energy space are needed to make this model work. For

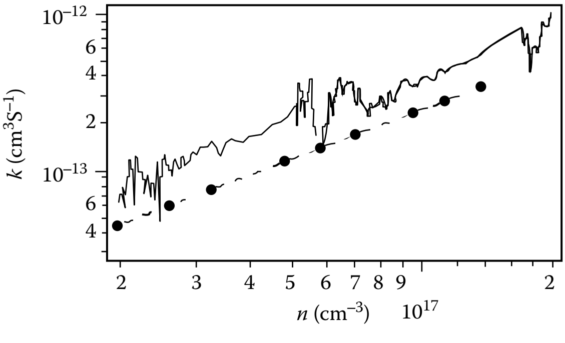


FIGURE 7.19 Variation of Langevin prefactor (k) as a function of carrier density. This shows that the recombination prefactor must vary as a function of carrier density. Solid line indicates data from Transient Absorption Spectroscopy (TAS), and dashed line indicates data from Transient Photovoltage (TPV). (From Shuttle, C. G. et al., *Phys. Rev. B.,* 78, 113201, 2008.)

example, Equation 7.10 where mobility is calculated also simply calculating the carrier density *n* requires another integral over energy space. Numerical integration is the bane of any engineer who wants to write efficient models because it involves loops and is therefore slow. Therefore, the most flexible and efficient way of dealing with these integrals is to tabulate them before the simulation starts. On the other hand most derivatives of the Scharfetter–Gummel current equations should be done analytically for numerical stability.

7.5.3 SRH Recombination—Splitting Up Energy Space

Even though the model described in Section 7.5.2 assumed that all carriers in the DoS at any given point are in quasi-thermal equilibrium (i.e., have a single quasi-Fermi level), the model performed well in the steady state where one could accept (if one is forgiving) that the carriers would have enough time to reach some sort of thermal equilibrium. However, this assumption of a single quasi-Fermi level describing both the carriers in the very deep traps and the very energetic carriers becomes harder to justify once one starts to perform time-domain simulations where it is clear that for a material with many trap states—not all carriers will be able to react instantaneously upon the application of a light/voltage pulse—especially if it was applied on the fs timescale. Rather than inventing new models for organic materials, the organic community borrowed the SRH model (Kirchartz et al. 2012; Shockley and Read 1952; Foertig et al. 2012), which has been used for years in the inorganic community to describe charge transport/trapping and recombination in disordered inorganic materials such as amorphous silicon.

This model is depicted in Figure 7.18c and expanded in Figure 7.20. From the band structure point of view the model assumes that free carriers move in a parabolic band (represented by parabolas in Figure 7.20, one for the free electrons one for the free holes), while trapped carriers reside in an exponential distribution of trap states (represented by two exponentials in Figure 7.20, one for the LUMO trap states one for the HOMO trap states).

From Figure 7.20 it can clearly be seen that the exponential trap distributions are broken up into a series of trap states depicted by squares. For each of these trap states the following rate equation is solved

$$\frac{dn_t}{dt} = r_{\text{ec}} - r_{\text{ee}} - r_{\text{hc}} + r_{\text{he}}, \tag{7.19}$$

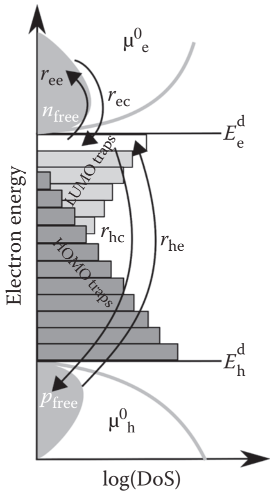


FIGURE 7.20 The SRH model depicted with a parabolic band describing free carriers and an exponential band describing trapped carriers.

where r_{ec} is the rate of electron capture into a trap, r_{ee} is the rate of electron escape from a trap, r_{hc} is the rate of hole capture into an electron trap, and r_{hc} is the rate of hole escape from an electron trap. The equations describing these are given in Table 7.1.

Where v_{th} is the thermal velocity of the carriers, n is the free electron density, σ_{n} is the capture cross section of free electrons into the electron trap, σ_{p} is the capture cross section of free holes into the electron trap, N_{t} is the density of trap states, and f is the occupation function of the trap state. If the capture rates for the electrons (r_{ec}) and holes (r_{hc}) into a single electron trap are examined, it can be seen that they consist of the product of the density of free charge carriers, the occupation probability of the trap state, the capture cross section, and the carrier thermal velocity. The capture rate does not, however, depend upon the depth of the energetic trap; thus, capture of charge carriers is not energetically assisted. If one examines the escape rates of carriers from the traps, $r_{\mathrm{ee}}$and r_{he}, it can be seen that they depend upon the terms $e_{\mathrm{n}}$and e_{p} which are defined as

$$e_{\mathrm{n}} = v_{\mathrm{th}}\sigma_{\mathrm{n}}N_{\mathrm{c}}\exp\left(\frac{E_{\mathrm{t}} - E_{\mathrm{c}}}{kT}\right) \tag{7.20}$$

and

$$e_{\mathrm{p}} = v_{\mathrm{th}}\sigma_{\mathrm{p}}N_{\mathrm{v}}\exp\left(\frac{E_{\mathrm{v}} - E_{\mathrm{t}}}{kT}\right). \tag{7.21}$$

It can be seen that these terms depend exponentially on the depth of a trap. Thus, the deeper the trap the harder it is for charge carriers to escape. In this model, there is no energetic barrier/assistance to relaxation but there is an energetic barrier to escape.

One can think of electron capture into an electron trap as simply trapping, and as free hole capture into an electron trap as charge carrier recombination. Equations 7.19 through 7.21 only describe capture/escape events for an electron trap (in this case the very highest LUMO trap level in Figure 7.20) an analogous set of equations can be written for capture and escape from hole traps. To couple this model to a drift-diffusion transport model, one must be able to calculate the total recombination rate from all trap levels. If one were solving the set of capture/escape equations in steady state, one could simply analytically integrate the equations over an exponential distribution of trap states and obtain the well-known standard steady-state SRH recombination rate equation,

$$R_{\mathrm{SRH}} = \frac{pn - n_{\mathrm{i}}^2}{\tau_{\mathrm{n}}(p + n_{\mathrm{i}}) + \tau_{\mathrm{p}}(n + n_{\mathrm{i}})}. \tag{7.22}$$

However, because organic devices are often studied using transient experiments to learn about charge-carrier dynamics, this approach would be limiting. We, therefore, divide up the distribution of trap states into evenly spaced traps (see Figure 7.20) and solve a rate equation describing trapping/detrapping for each energetic level at each spatial mesh point in the device. Usually it is sufficient to have around 8–15 traps

TABLE 7.1 The SRH Capture and Escape Equations

Mechanism	Rate	Description
Electron capture rate	r_{ec}	$nv_{\mathrm{th}}\sigma_{\mathrm{n}}N_{\mathrm{t}}(1-f)$
Electron escape rate	r_{ee}	$e_{\mathrm{n}}N_{\mathrm{t}}f$
Hole capture rate	r_{hc}	$pv_{\mathrm{th}}\sigma_{\mathrm{p}}N_{\mathrm{t}}f$
Hole escape rate	r_{he}	$e_{\mathrm{p}}N_{\mathrm{t}}(1-f)$

in each band, although sometimes when investigating the structure of the DoS, up to 80 can be required. It is assumed that within each trap we can use a single quasi-Fermi level to describe the occupation of the trapped carriers. One final advantage of the SRH-based mode over the model described in Section 7.5.2 is that the SRH model uses parabolic bands to describe the movement of free carriers. Although an over-simplification, it enables the use of the usual Einstein relations which helps the stability of the model.

The SRH model, when applied to organic semiconductors, has been very successful at being able to describe a range of experimental data measured using a variety of different electrical/optical measurement techniques over a wide range of timescales. These include light/dark steady-state JV measurements (Sims et al. 2004) (see Figure 7.21), the slow transient current response of a device to a rising voltage waveform (see Figure 7.22), the microsecond transient current responses of a device due to optical excitation (see Figure 7.23), picosecond timescale pump probe responses of devices due to a femtosecond excitation (see Figure 7.24), and even frequency domain measurements (MacKenzie et al. 2016). The SRH-based drift diffusion model used to produce Figures 7.22 through 7.24 is available free to download at www.gpvdm.com.

Before we leave recombination, it is worth coming back to the Langevin recombination expression in Equations 7.14 and 7.15 once more. Equation 7.14 describes a process which is dependent upon the density of electrons and on the density of holes; this is almost always understood as *free* electrons and *holes* colliding and annihilating. This process is referred to in the literature as bimolecular recombination. Some authors understand this to be recombination between an electron on one molecule and a hole on another molecule, while others understand it to mean the process is of order two (a process depending upon the density of both electrons and holes). The latter definition seems to make more sense when compared to the inorganic literature. Although the Langevin prefactor to Equation 7.14 (7.15) comes from a derivation which is probably not valid for organic semiconductors, this does not mean free electron to free hole recombination does not happen in organic materials; it just means the prefactor is probably not correct. For the special case of highly ordered organic crystals with very few defect states, there is no reason why bimolecular recombination should not dominate SRH recombination. When modeling devices where one suspects bimolecular recombination is important, it is therefore fairly safe to assume k is simply an arbitrary model constant and neglect Equation 7.16.

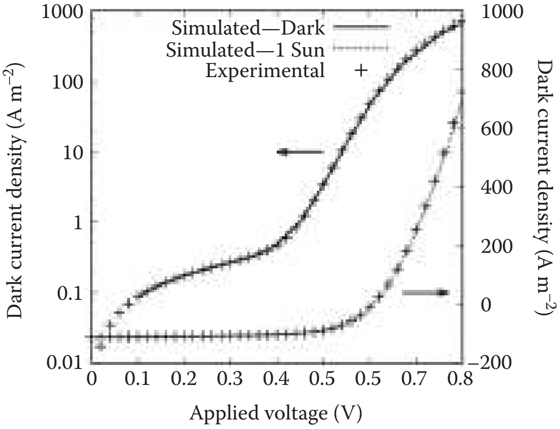


FIGURE 7.21 Steady-state JV curve measurements for a P3HT: phenyl-C61-butyric acid methyl ester (PCBM) organic solar cell reproduced by an SRH model. Note for the light JV curve, the open circuit voltage (intersect of the x-axis), short circuit voltage (intersect of the y-axis), and fill factor (maximum power point) are reproduced exactly by the model. The slope of the dark curve (ideality factor) is also reproduced by the model, suggesting recombination is being modeled correctly. (From MacKenzie, Roderick C. I. et al., *Adv. Energy Mater.*, 2012.)

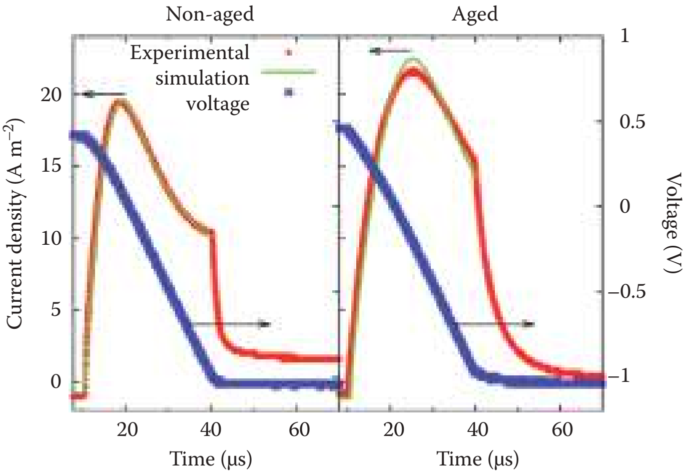


FIGURE 7.22 An example of a slow electrical transient current responses of a P3HT:PCBM diode to a 40-s-long linearly increasing voltage transient. The transient starts at 0.5 V and ends at −1 V. This measurement is called the Charge Extraction by Linearly Increasing Voltage (CELIV) measurement. By applying a voltage ramp, one can extract a significant amount of charge from a device. The time it takes the charge to leave the device can be used to calculate a mobility figure for the material. It can be seen that for this relatively slowly varying electrical transient, the SRH-based model can reproduce the results. (From Hanfland, R. et al., *Appl. Phys. Lett.*, 103, 063904, 2013.)

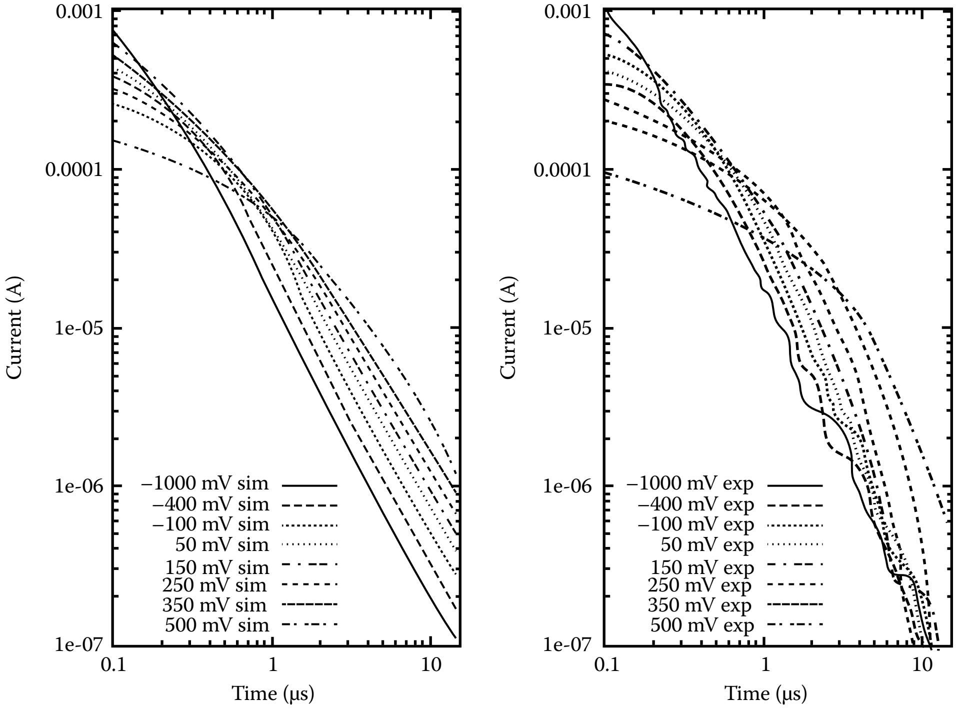


FIGURE 7.23 Right, experimental current transient responses from an organic diode after being photoexcited by a 5ns laser pulse as a function of applied bias. Left, the simulated predictions of SRH-based model of the same experiment. The model was only calibrated to the experimental data at 0 mV, but it was able to predict the transient responses up to 1V away from the calibration point. (From MacKenzie, Roderick C. I. et al., *Adv. Energy Mater.*, 2012.)

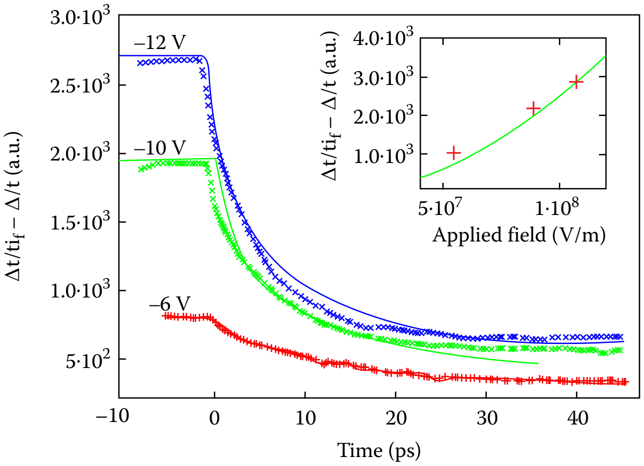


FIGURE 7.24 The simulated and experimental ultrafast (picoseconds) response of an organic diode to a 250-fs laser. The measured signal is the result of optically probing an organic diode up to 40 ps after photoexcitation. This type of pump–probe experiment can provide information on how charge carriers separate on very early timescales in these materials. It can be seen that the experiment and the result of the SRH-based model agree well. (From MacKenzie, Roderick C. I. et al., *Phys. Rev. B.*, 89, 195307, 2014.)

7.6 Geminate and Nongeminate Recombination—Role of Excitons in Organic Devices

One key difference between organic and inorganic semiconductors is that the relative permittivity of organic semiconductors is around three while the relative permittivity of inorganic semiconductors such as silicon are around 12. This becomes important when one considers photogeneration in these materials. If one considers the electrostatic force between an electron and hole,

$$F = \frac{q_1 q_2}{4\pi\varepsilon_0\varepsilon_r r^2}. \tag{7.23}$$

It can be seen that the force between the charge pairs is inversely proportional to the permittivity of the medium. Thus, the force between an electron hole pair in silicon will be three to four times weaker than in an organic material. Practically, this means after photogeneration in an inorganic semiconductor, the electron and hole can escape each other and drift off into the sea of electrons/holes. In organic semiconductors, the attractive force between newly generated electrons and holes is so strong that each can't escape the other's field. Thus, they remain bound. The bound electron and hole pair is called an exciton. Unless some external force splits this pair (such as when the pair encounters a heterojunction), the electron and hole will steadily fall into each other's field and annihilate. This annihilation process is referred to as geminate recombination in the literature. The recombination of unbound electrons/holes as described in Section 7.5 is referred to as nongeminate recombination. The charge generation process in an organic device is depicted in Figure 7.25.

Some device models (Koster et al. 2005; Fluxim n.d.) include an exciton generation/diffusion equation, then link the exciton splitting term to the standard drift diffusion equations are a charge generation rate. Although this approach is probably physically correct, exciton diffusion distances are typically quite short (5–20 nm) but usually not well known for any given material system. Therefore, when modeling photoactive diodes such as solar cells, it is usually sufficient to multiply the calculated charge generation rate (modal profile) from the optical model by a reduction parameter (0–1.0) to account for any geminate recombination which may occur.

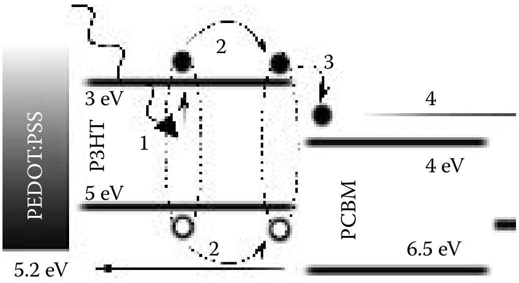


FIGURE 7.25 A diagram of the charge generation process in an organic semiconductor. A photon is absorbed in the P3HT and photoexcites an exciton (1). This bound electron hole pair diffuses to a heterojunction between the P3HT and PCBM (2) where the potential gradient splits the exciton into free charge carriers (3). Once the electron and hole are free of each other's electrostatic potential they can drift/diffuse to the contacts (4) and to the external circuit. This device is a diode with indium tin oxide (ITO) and Al contacts. The poly(3,4-ethylenedioxythiophene)-poly(styrenesulfonate) (PEDOT:PSS) layer prevents electrons from the P3HT recombining on the ITO.

7.7 Numerical Aspects of Organic Device Simulations

Many of the early papers used Gummel's method to solve the drift-diffusion equations (i.e., where one solves Poisson's equations, then solves the drift-diffusion equations, then iterates until convergence). There are quite a few papers claiming this to be a good, stable method. However, in my experience when one separates out the solution of these equations convergence is in general less stable than when one solves all sets of equations in one matrix, using Newton's method. In general, one can get away with Gummel's method when one is only solving the drift-diffusion equations, but when one starts to introduce trap states, and other nonlinearities, the problem becomes more and more nonlinear and one is forced to use Newton's method applied in a single matrix equation, of the form

$$[J] = -[\Delta]\left[f\right], \tag{7.24}$$

where J is the Jacobin containing the derivatives of the error function (f) and Δ is the update to the solution variables. Writing the equations in this form produces an overhead for the scientist in setting up the problem, but the time saved during computation pays back for the time invested. The only exception to this rule is when modeling thermal effects; because all parameters of the equations are highly dependent upon temperature it can often be very time-consuming to set up the Jacobin correctly. Finally, backtracking with these models, if one reads the art of scientific computing, one comes away with the impression that backtracking is useful to help convergence of these models. I have, however, never found that the computational overhead in setting up the backtracking speeds up the convergence enough to compensate for the computational overhead. In my experience, if you need to implement backtracking, your derivatives in the Newton solver are probably wrong. The matrix Equation 7.23 represents a very sparse system, and the best sparse matrix solver I have found to date is UMFPACK, published by Timothy A. Davis at the Texas A&M University.

Very often, one needs to fit these models to multiple sets of experimental data in order to extract device material. This involves a search of a large parameter space. I have found the most effective method for doing this to be the downhill simplex algorithm from the GNU scientific library; it avoids the need to evaluate derivatives of the error function and is robust. The most efficient strategy to fitting data sets is to start 100 copies of the model off on a cluster with randomly selected material parameters, then to let the downhill simplex algorithm improve the fits. After a while the fits will converge. If the result is acceptable, stop the fitting process. If not, then a local minima has been found and the fitting process should be restarted with another random set of parameters.

7.8 Conclusions

This chapter has given an overview of modeling organic semiconductors and highlighted the difference between classical materials and organic materials. We covered microscopic modeling of organic materials from the quantum perspective. We then discussed MD to simulate thin film formation. Finally, we discussed some aspects of device models used in the simulation of organic electronic devices. Organic materials are far more complex than traditional inorganic semiconductors due to the need to understand the form–function relationships, which govern their performance; this chapter represents only the briefest of introductions to the subject. Further reading on organic electronics can be found in Brutting and Adachi (2012).

References

Adams, J. et al. Air-processed organic tandem solar cells on glass: Toward competitive operating lifetimes. *Energy Environ. Sci.* 8 (2015): 169–176.

Azoff, E. M. Computer Modeling of Heterojunction Bipolar Transistors. PhD thesis. Sheffield: University of Sheffield (1986).

Berendsen, et al. (1995) *Comp. Phys. Comm.* 91: 43–56.

Bloch, F. Über die Quantenmechanik der Elektronen in Kristallgittern. *Z. Phys.* 52 (1928): 555–600.

Boix, P. P., G. Garcia-Belmonte, U. Munecas, M. Neophytou, C. Waldauf, and R. Pacios. Determination of gap defect states in organic bulk heterojunction solar cells from capacitance measurements. *Appl. Phys. Lett.* 95 (2009): 233302–233303.

Braun, C. L. Electric field assisted dissociation of charge transfer states as a mechanism of photocarrier production. *J. Chem. Phys.* 80 (1984): 4157.

Brutting, W. and C. Adachi. *Physics of Organic Semiconductors*, 2nd Ed. New York, NY: Wiley (2012).

Canali, C., C. Jacoboni, F. Nava, G. Ottaviani, and A. Alberigi-Quaranta. Electron drift velocity in silicon. *Phys. Rev. B* 12 (1975): 2265.

Carr, J. A. and S. Chaudhary. The identification, characterization and mitigation of defect states in organic photovoltaic devices: A review and outlook. *Energy Environ. Sci.* 6 (2013): 3414–3438.

Chen, B., I. Siepmann, S. Karaborni, and M. L. Klein. Vapor-liquid and vapor-solid phase equilibria of fullerenes: The role of the potential shape on the triple point. *J. Phys. Chem. B* 107 (2003): 12320–12323.

Chiang, C. K. et al. Electrical conductivity in doped polyacetylene. *Phys. Rev. Lett* 39 (1977): 1098.

Dale, M. and S. M. Benson. Energy balance of the global photovoltaic (PV) industry—Is the PV industry a net electricity producer? *Environ. Sci. Technol.* 47 (2013): 3482–3489.

Dattani, R. A general mechanism for controlling thin film structures in all-conjugated block copolymer: Fullerene blends. *J. Mater. Chem. A* 2 (2014): 14711–14719.

Delannoy, Y. Purification of silicon for photovoltaic applications. *J. Cryst. Growth* 360 (2012): 61–67.

Deschler, F, D. Riedel, B. Ecker, E. von Hauff, E. Da Como, and R. C. I. MacKenzie. Increasing organic solar cell efficiency with polymer interlayers. *Phys. Chem. Chem. Phys.* 15 (2013): 764–769.

Fluxim, Setfos "semiconducting emissive thin film optics simulator", http://www.fluxim.com/drift-diffusion-module/. (n.d).

Fock, V. Näherungsmethode zur Lösung des quantenmechanischen Mehrkörperproblems. *Z. Phys.* 61, no. 1(1930): 126–148.

Foertig, A., J. Rauh, V. Dyakonov, and C. Deibel. Shockley equation parameters of P3HT:PCBM solar cells determined by transient techniques. *Phys. Rev. B* 86 (2012): 115302.

Guilbert, A. A. Y. et al. Effect of multiple adduct fullerenes on microstructure and phase behavior of P3HT: Fullerene blend films for organic solar cells. *ACS Nano* 6 (2012): 3868–3875.

Guilbert, A. A. Y. et al. Spectroscopic evaluation of mixing and crystallinity of fullerenes in bulk heterojunctions. *Adv. Funct. Mater.* 24 (2014): 6972–6980.

Guilbert, A. A. Y. et al. Influence of bridging atom and side chains on the structure and crystallinity of cyclopentadithiophene–benzothiadiazole polymers. *Chem. Mater.* 26, no. 2 (2014): 1226–1233.

Guilbert, A. A. Y. et al. Temperature-dependent dynamics of polyalkylthiophene conjugated polymers: A combined neutron scattering and simulation study. *Chem. Mater.* 27, no. 22 (2015): 7652–7661.

Hanfland, R., M. A. Fischer, W. Brütting, U. Würfel, and R. C. I. MacKenzie. The physical meaning of charge extraction by linearly increasing voltage transients from organic solar cells. *Appl. Phys. Lett.* 103 (2013): 063904.

Hodgkiss, J., G. Conboy, G. Hutchings, S. Higgins, and I. Galbraith. Organic photovoltaics and energy: General discussion. *Faraday Discuss.* 174 (2014): 341–355.

Hohenberg, P. and W. Kohn. Inhomogeneous electron gas. *Phys. Rev.* 136 (1964): B864.

Jiao, Y., F. Zhangl, M. Grätzel, and S. Meng. Structure–property relations in all-organic dye-sensitized solar cells. *Adv. Funct. Mater.* 23 (2013): 424–429.

Kirchartz, T. et al. Sensitivity of the Mott–Schottky analysis in organic solar cells. *J. Phys. Chem. C* 116, no. 14 (2012): 7672–7680.

Kirchartz, T., B. E. Pieters, J. Kirkpatrick, U. Rau, and J. Nelson. Recombination via tail states in polythiophene: Fullerene solar cells. *Phys. Rev. B* 83 (2011): 115209

Kohn, W. and L. J. Sham. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* 140 (1965): A1133.

Koster, L. J. A., E. C. P. Smits, V. D. Mihailetchi, and P. W. M. Blom. Device model for the operation of polymer/fullerene bulk heterojunction solar cells. *Phys. Rev. B* 72 (2005): 085205.

Kwiatkowsk, J. *From Molecules to Mobilities: Modelling Charge Transport in Organic Semiconductors*. Thesis London: Imperial College London (2009).

LAMPS. http://lammps.sandia.gov/. (n.d.).

Letheby, H. On the production of a blue substance by the electrolysis of sulphate of aniline. *J. Chem. Soc.* 15 (1862): 161–163.

Lim, J. J. Simulation of double quantum well GalnNAs laser diodes. *IET Optoelectron.* 1, no. 6 (2007): 259–265.

MacKenzie, R. C. I. et al. Interpreting the density of states extracted from organic solar cells using transient photocurrent measurements. *J. Phys. Chem. C* 117, no. 24 (n.d.): 12407–12414.

MacKenzie, R. C. I. et al. Loss mechanisms in high efficiency polymer solar cells. *Adv. Energy Mater.* 6, no. 4 (2016): 1501742.

MacKenzie, R. C. I., J. M. Frost, and J. Nelson. A numerical study of mobility in thin films of fullerene derivatives. *J. Chem. Phys.* 132 (2010): 064904.

MacKenzie, R. C. I., A. Göritz, S. Greedy, E. von Hauff, and J. Nelson. Theory of Stark spectroscopy transients from thin film organic semiconducting devices. *Phys. Rev. B* 89 (2014): 195307.

MacKenzie, R. C. I., T. Kirchartz, G. F. A. Dibb, and J. Nelson. Modeling nongeminate recombination in P3HT: PCBM solar cells. *J. Phys. Chem. C* 115, no. 19 (2011): 9806–9813.

MacKenzie, R. C. I., C. G. Shuttle, M. L. Chabinyc, and J. Nelson. Extracting microscopic device parameters from transient photocurrent measurements of P3HT: PCBM solar cells. *Adv. Energy Mater.* 2, no. 6 (2012): 662–669.

Nelson, J. Diffusion-limited recombination in polymer-fullerene blends and its influence on photocurrent collection. *Phys. Rev. B* 67 (2003): 155209.

New Scientist. Brighter outlook for cheaper silicon solar cells. *New Scientist* (August 1979): 522.

Ridley, J. and M. Zerner. An intermediate neglect of differential overlap technique for spectroscopy: Pyrrole and the azines. *Theor. Chim. Acta* 32 (1973): 111–134.

Roichman, Y. and N. Tessler. Generalized Einstein relation for disordered semiconductors—Implications for device performance. *Appl. Phys. Lett*. 80 (2002): 1948.

Schafferhans, J., A. Baumann, A. Wagenpfahl, C. Deibel, and V. Dyakonov. Oxygen doping of P3HT: PCBM blends: Influence on trap states, charge carrier mobility and solar cell performance. *Org. Electron*. 11, no. 10 (2010): 1693–1700.

Shockley, W. and W. T. J. Read. Statistics of the recombinations of holes and electrons. *Phys. Rev.* 87 (1952): 835.

Shuttle, C. G., B. O'Regan, A. M. Ballantyne, J. Nelson, D. D. C. Bradley, and J. R. Durrant. Bimolecular recombination losses in polythiophene: Fullerene solar cells. *Phys. Rev. B* 78 (2008): 113201.

Sims, L. et al. Investigation of the s-shape caused by the hole selective layer in bulk heterojunction solar cells. *Org. Electron*. 15 (2014): 2862–2867.

Street, R. A. Localized state distribution and its effect on recombination in organic solar cells. *Phys. Rev. B* 84 (2011): 075208.

Szabo, A. *Modern Quantum Chemistry.* Mineola, NY: Dover Books, 1996.

To, T. T. and S. Adams. Modelling of P3HT: PCBM interface using coarse-grained forcefield derived from accurate atomistic forcefield. *Phys. Chem. Chem. Phys.* 16 (2014): 4653–4663.

Tress, W, K. Leo, and M. Riede. Influence of hole-transport layers and donor materials on open-circuit voltage and shape of I–V curves of organic solar cells. *Adv. Funct. Mater*. 21, no. 11 (2011): 2140–2149.

Voigt, M. M. et al. Gravure printing for three subsequent solar cell layers of inverted structures on flexible substrates. *Sol. Energ. Mat. Sol. Cells* 95 (2011): 731–734.

Voigt, M. M. et al. Gravure printing inverted organic solar cells: The influence of ink properties on film quality and device performance. *Sol. Energ. Mat. Sol. Cells* 105 (2012): 77–85.

Yersin, H. *Highly Efficient OLEDs with Phosphorescent Materials.* New York, NY: John Wiley & Sons (2008).

Zhao, K. et al. Vertical phase separation in small molecule: Polymer blend organic thin film transistors can be dynamically controlled. *Adv. Funct. Mater.* 26 (2016): 1–10.

8

Polarization in III-N Semiconductors

8.1 Introduction 219
8.2 Piezoelectric Effect 220
8.3 Piezoelectricity in Electronic Materials 220
8.4 Mathematical Formalism 222
Piezoelectric Tensor • Crystal Vectors • Classic Description of Strain-Induced Polarization • Spontaneous Polarization in Wurtzite Crystals
8.5 Piezoelectric Coefficients of Wurtzite III-N Semiconductors... 226
Proper versus Improper Coefficients • Linear Piezoelectric Coefficients of III-N Semiconductors • Nonlinearities in III-N Semiconductors • Full Second-Order Piezoelectric Tensor Formalism • Calculated Second-Order Piezoelectric Coefficients • Calculated versus Experimental Values of the Spontaneous Polarization • Continuity Equation in Superlattices and Quantum Well Structures • Schemes for Dealing with Alloys
8.6 Electromechanical Coupling in Wurtzite Semiconductors 239
Electromechanical Coupling with Linear Piezoelectricity • Electromechanical Coupling with Nonlinear Piezoelectricity
8.7 Nonpolar Orientations in Wurtzite Semiconductors 245
8.8 Summary and Future Directions 245

Max A. Migliorato
Joydeep Pal
Xin Huang
Weiguo Hu
Morten Willatzen
and
Yousong Gu

8.1 Introduction

In recent years, because of the extraordinary technological developments in solid-state lighting, there has been increased interest in polarization effects under external strain, the so-called piezoelectric polarization. GaN and its related semiconductor materials (AlN and InN) are of crucial importance for the realization of the fundamental component of solid-state lighting, the blue light-emitting diode (LED): the Nobel Prize in Physics was jointly awarded to Isamu Akasaki, Hiroshi Amano, and Shuji Nakamura (Nobel, 2014) for this work. With LEDs in mind, here we cover all aspects of modeling the effects of piezoelectric polarization in III-N alloys. A prediction was reported in the *Proceedings of the State-of-the-Art Program on Compound Semiconductors XXXIX and Nitride and Wide Bandgap Semiconductors for Sensors, Photonics and Electronics* that the output efficiency of III-N semiconductor-based LEDs was always going to be limited by the magnitude of the internal polarization fields (Kopf, 2003). In the years since, it has become universally acknowledged that such a prediction was correct. Detailed knowledge of the piezoelectric properties of III-N semiconductors is therefore an essential element in designing more efficient LEDs.

In this chapter, we review the different models that have been proposed over the years and how they compare to experimental measurements. We outline both linear and nonlinear models and discuss the recent experimental determination of values of spontaneous polarization in polymorphic crystals.

8.2 Piezoelectric Effect

The direct piezoelectric effect is the generation of an electric dipole moment in certain crystals if a stress is applied and the electric dipole moment is proportional to the applied stress. Such an effect was first demonstrated by the Curie brothers, Jacques and Pierre, in 1880 (Curie and Curie, 1880). The word "piezo" is from the Greek "piezen (πιέζειν)" which means "to press or squeeze." The effect has been understood as the charge formation in some crystals due to the ionic displacements under some mechanical pressure. The converse piezoelectric effect, where the crystals deforms under the application of external electric field in the polarization direction, was outlined mathematically in 1881 by Lippmann (1881) from fundamental thermodynamic principles. Later, in 1910, the publication *Lehrbuch der Kristallphysik* (textbook on crystal physics) by Woldemar Voigt distinguished 20 natural piezoelectric crystal classes and for the first time treated the piezoelectric coefficients using tensor analysis. The use in this work of a contracted notation for the indexes gave rise to what is conventionally known as the Voigt notation (Voigt, 1910).

The reasons the property of piezoelectricity is closely linked to the crystal symmetry of materials is that mechanical stress must break the charge symmetry for a change in polarization to be observed. This asymmetry in the charge density leads to an electric field in the crystal which is known as the piezoelectric field.

Though it is often challenging to perform direct experimental measurements of the resulting electric field, an impact can be more easily observed on both the electrical and optical properties of the crystal.

Of interest for this chapter is that piezoelectricity is commonly found in semiconductor materials typically used in optoelectronic applications, where the crystal symmetry is either cubic (e.g., GaAs and InP) or hexagonal (GaN and ZnO), as shown in Figure 8.1. All III-nitride (III-N) semiconductors typically found in the wurtzite crystal structure (Figure 8.1d), not only GaN, strong polarization along the polar axis. This phenomenon has long been known. The most important quantitative theoretical analysis was published by Bernardini et al. (1997). At the time of this writing, the parameters proposed in their work have been cited 1500 times according to the American Physical Society (APS) database.

8.3 Piezoelectricity in Electronic Materials

Though bulk materials possess the property of piezoelectricity, the changes with pressure in the total polarization, which is the sum of spontaneous (polarization that exists even in the absence of external pressure) and strain-induced piezoelectric polarization, lead to an increase in the polarization-induced electrostatic charge density. In a structurally pure bulk semiconductor, the variation in the polarization fields will be observed at the surfaces as the electrical dipoles cancel one another throughout the structure. Nonetheless, the existence of the surface states or mobile carriers can nullify the resultant surface charges. Therefore, in conventional mainstream electronics, there is little scope for forcing stress on bulk materials for any practical purpose. This is not to say that there are no applications at all, and in fact, there is an entire field (piezotronics, from piezoelectric and electronics) which is based on using nanostructured materials (typically nanowires) and their strain–voltage relationship (Wang, 2007). Furthermore, there are some notable applications in micro- and nanoelectro mechanical systems (MEMS and NEMS). However, in mainstream optoelectronics, strain is often built into the crystal layers rather than supplied externally. This is the case of heterostructures, i.e., layers of different semiconductor materials carefully deposited through an epitaxial process (typically for III-N using metal organic chemical vapor deposition [MOCVD]), such that

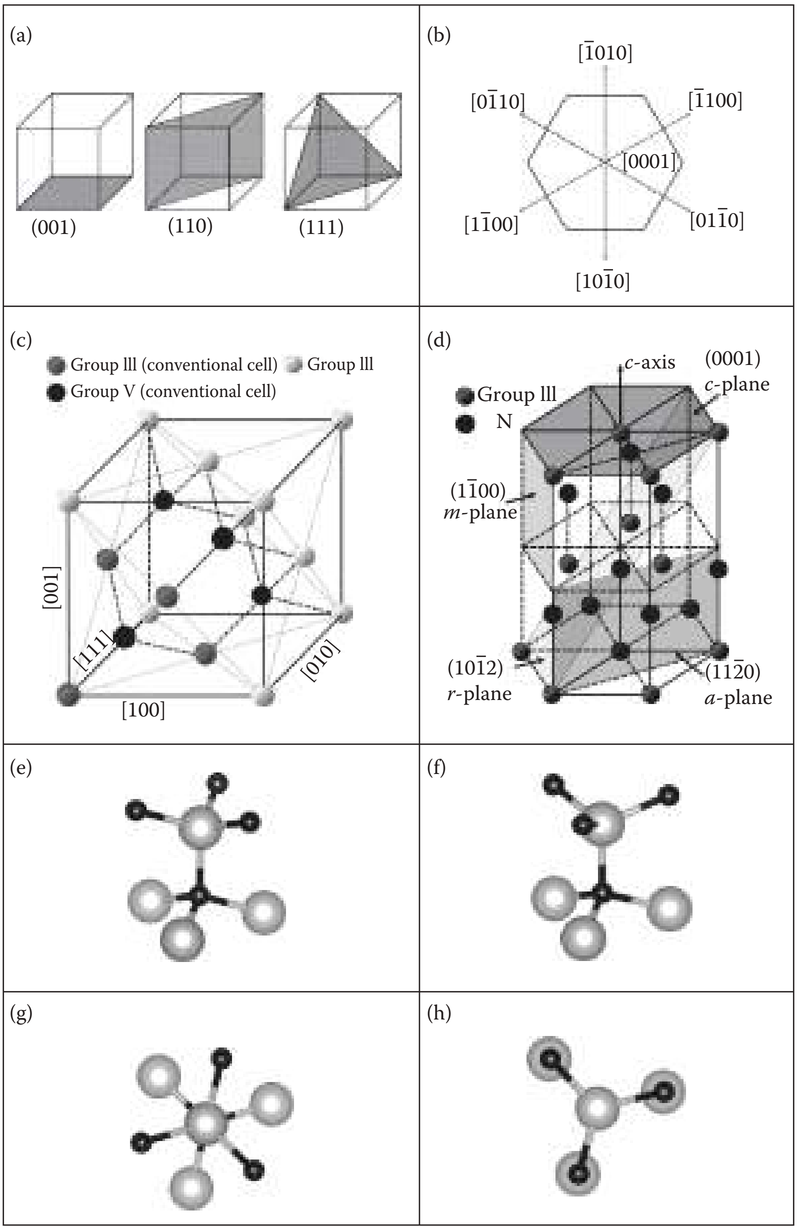

(a)
(001)
(110)
(111)
(b)
$[\bar{1}010]$
$[0\bar{1}10]$
$[\bar{1}100]$
[0001]
$[1\bar{1}00]$
$[01\bar{1}0]$
$[10\bar{1}0]$
(c)
Group lll (conventional cell)
Group lll
Group V (conventional cell)
[001]
[111]
[010]
[100]
(d)
c-axis
(0001)
c-plane
Group lll
N
$(1\bar{1}00)$
m-plane
$(10\bar{1}2)$
r-plane
$(11\bar{2}0)$
a-plane
(e)
(f)
(g)
(h)

FIGURE 8.1 Zincblende (left) and wurtzite (right) crystals: (a)–(b) high symmetry planes, (c)–(d) crystal structures and atomic positions—for zincblende (c), atoms belonging to the conventional cell are given alongside the crystal directions, while for wurtzite (d), also given are the naming conventions of the high symmetry planes—(e)–(f) atomic structure side view (along the [111] and [0001] directions for the zincblende and wurtzite crystals, respectively) and (g)–(h) atomic structure top view (from the [111] and [0001] planes for the zincblende and wurtzite crystals, respectively).

the crystal maintains a high-crystal order. When two materials in a heterostructure have different lattice parameters, then the atomic crystal cells of one tend to stretch to adapt to those of the material they are being deposited over. This is called lattice-mismatched heteroepitaxy, which can produce very high-quality crystals with a very low number of crystal defects.

In a polar heterostructure, the strained layers develop a small dipole factor in each unit cell, which cancels throughout except at the interfaces. This produces a uniform distribution of polarization, associated to a net surface charge density at the interfaces of the heterostructure, leading to a net electric field which has a significant influence on the electro-optical characteristics of the material.

In order to understand this in more detail, we first need to outline the conventional formalism used to represent such strain-induced polarization effects.

8.4 Mathematical Formalism

As implied in Section 8.3, piezoelectricity can be described in terms of microscopic dipoles, and it can therefore have an atomistic description. However, because it is also a macroscopic property, a scalable mathematical description is also practical (Singh, 1992, 2003).

8.4.1 Piezoelectric Tensor

In the case of piezoelectricity, the requirement is fulfilled by describing the resulting polarization field $\mathbf{P}_{\mathbf{pz}}$ in terms of the piezoelectric moduli $\mathbf{d}_{\mathbf{ijk}}$ or piezoelectric coefficients $\mathbf{e}_{\mathbf{ijk}}$ which are tensors coupling to either the stress tensor $\boldsymbol{\sigma}_{\mathbf{jk}}$ or the strain tensor $\boldsymbol{\varepsilon}_{\mathbf{jk}}$

$$\mathbf{P}_{\mathbf{pz,\,i}} = \mathbf{d}_{\mathbf{ijk}}\boldsymbol{\sigma}_{\mathbf{jk}} = \mathbf{e}_{\mathbf{ijk}}\boldsymbol{\varepsilon}_{\mathbf{jk}}, \tag{8.1}$$

where the stress/strain relationship is dependent on the elastic tensor $\mathbf{c}_{\mathbf{ijkl}}$:

$$\boldsymbol{\sigma}_{\mathbf{ij}} = \mathbf{c}_{\mathbf{ijkl}}\boldsymbol{\varepsilon}_{\mathbf{kl}} \tag{8.2}$$

Understanding the various subscript indexes is not as complicated as one would imagine. The four indexes i, j, k, and l all range from 1 to 3, to represent the three directions x, y, and z of a three-dimensional (3D) space. Using Equation 8.3 we illustrate how to interpret the quantities with, e.g., three subscript indexes:

$$\mathbf{P}_{\mathbf{pz,\,1}} = \mathbf{e}_{\mathbf{123}}\boldsymbol{\varepsilon}_{\mathbf{23}}. \tag{8.3}$$

This case indicates that the polarization in the x (1) direction is generated by a strain in the yz (23) direction, multiplied by the related piezoelectric coefficient, where the first index is the direction of the polarization and the other two are the direction of the strain. In this case, the strain is in a diagonal direction compared to the x-, y-, or z-axes. Assuming that the material can be taken as cubic, with sides parallel to the axes, $\boldsymbol{\varepsilon}_{\mathbf{23}}$ would turn the square face into a diamond. This is commonly referred to as a shear deformation. If instead, the two subscript indexes of the strain component were the same, then the direction of the strain would be parallel to one of the sides of the cube and the strain would simply modify one of its dimensions. As the jk indexes are interchangeable between themselves, the nine combinations of two indexes contain three redundancies. It is then common to use a reduced notation instead. The simplification (Voigt, 1910) works by describing pairs of indexes with the following convention:

$$\mathbf{xx = 1, yy = 2, zz = 3, yz = 4, zx = 5, xy = 6}. \tag{8.4}$$

This is also justified by noting that the symmetry of the tensors $\mathbf{d}_{\mathbf{ijk}}$, $\boldsymbol{\sigma}_{\mathbf{jk}}$, $\mathbf{e}_{\mathbf{ijk}}$, and $\boldsymbol{\varepsilon}_{\mathbf{jk}}$ belongs to the indexes j and k rather than i. In the reduced Voigt notation which compresses the jk indexes into a single index α, the matrix representations for the $\boldsymbol{\varepsilon}_\alpha$ and $\mathbf{e}_{i\alpha}$ tensors are therefore

$$\boldsymbol{\varepsilon}_\alpha = \begin{pmatrix} \varepsilon_{xx} & \varepsilon_{yx} & \varepsilon_{zx} \\ \varepsilon_{xy} & \varepsilon_{yy} & \varepsilon_{zy} \\ \varepsilon_{xz} & \varepsilon_{yz} & \varepsilon_{zz} \end{pmatrix} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \\ \varepsilon_4 \\ \varepsilon_5 \\ \varepsilon_6 \end{pmatrix} \tag{8.5}$$

$$\mathbf{e}_{i\alpha} = \begin{pmatrix} e_{11} & e_{12} & e_{13} & e_{14} & e_{15} & e_{16} \\ e_{21} & e_{22} & e_{23} & e_{24} & e_{25} & e_{26} \\ e_{31} & e_{32} & e_{33} & e_{34} & e_{35} & e_{36} \end{pmatrix}. \tag{8.6}$$

Note that when using Voigt notation, the shear components are given by $\boldsymbol{\varepsilon}_\alpha = \boldsymbol{\varepsilon}_{jk} + \boldsymbol{\varepsilon}_{kj} = 2\boldsymbol{\varepsilon}_{jk}$, where in the last passage the factor of 2 arises if we assume that the indexes jk are interchangeable. Omitting the factor of 2 has been a recurring problem in the literature (Romanov et al., 2006).

Although there are many different coefficients, symmetry arguments reveal that many of the coefficients of $\mathbf{e}_{i\alpha}$ are either vanishing or equal to others.

8.4.1.1 Piezoelectric Tensor in Zincblende Crystals

In a zincblende crystal, there are only three independent elastic constants ($\mathbf{c}_{11} = \mathbf{c}_{22} = \mathbf{c}_{33}, \mathbf{c}_{12} = \mathbf{c}_{13} = \mathbf{c}_{23}, \mathbf{c}_{44} = \mathbf{c}_{55} = \mathbf{c}_{66}$) that reflect the isotropic cubic symmetry. Furthermore, it was shown by Nye (1957) that there are only three nonzero independent coefficients:

$$\mathbf{e}_{i\alpha} = \begin{pmatrix} 0 & 0 & 0 & \mathbf{e}_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & \mathbf{e}_{25} & 0 \\ 0 & 0 & 0 & 0 & 0 & \mathbf{e}_{36} \end{pmatrix}, \tag{8.7}$$

which are all of equal value. In such a case, it is conventional to refer to all three of them by the coefficient with the lowest index, i.e., $\mathbf{e}_{14}$.

8.4.1.2 Piezoelectric Tensor in Wurtzite Crystals

In wurtzite crystals, there are five independent elastic constants $\mathbf{c}_{11} = \mathbf{c}_{22}, \mathbf{c}_{33}, \mathbf{c}_{12}, \mathbf{c}_{13} = \mathbf{c}_{23}, \mathbf{c}_{44}$ which reflect the isotropic hexagonal symmetry in two directions and a different symmetry along the third direction (*c*-axis).

The wurtzite phase, typical of III-N materials, is a bit more complicated than zincblende in terms of piezoelectric tensor, and instead of three there are five independent coefficients:

$$\mathbf{e}_{i\alpha} = \begin{pmatrix} 0 & 0 & 0 & 0 & \mathbf{e}_{15} & 0 \\ 0 & 0 & 0 & \mathbf{e}_{24} & 0 & 0 \\ \mathbf{e}_{31} & \mathbf{e}_{32} & \mathbf{e}_{33} & 0 & 0 & 0 \end{pmatrix}, \tag{8.8}$$

which can be reduced to three by noting that $\mathbf{e}_{31} = \mathbf{e}_{32}$ and $\mathbf{e}_{15} = \mathbf{e}_{24}$. Again, the convention is to refer to them as $\mathbf{e}_{31}$ and $\mathbf{e}_{15}$.

8.4.2 Crystal Vectors

Thus far, we have not specified what the *x*-, *y*-, and *z*-directions are for the two crystals we have considered, i.e., the zincblende and the wurtzite structures. The choices are not arbitrary.

8.4.2.1 Crystal Directions in Zincblende Crystals

For zincblende crystals, one conventionally takes x, y, and z aligned with the [100], [010], and [001] crystal directions (Figure 8.1c). This choice uniquely links the tensor formalism just described with the atomic structure, given that the crystal directions, which appear to simply describe a cubic cell, are also such that all atomic bonds are in one of four directions [111], $[\bar{1}\bar{1}1]$, $[1\bar{1}\bar{1}]$, or $[\bar{1}1\bar{1}]$. The [111] direction is known as the polar axis of the zincblende crystal, i.e., only polarization at angles less than $\pi/2$ from the polar axis is nonvanishing.

8.4.2.2 Crystal Directions in Wurtzite Crystals

In the wurtzite crystal, the crystal directions are not described as simply as in zincblende, as it is a very common convention to use a four-index representation instead of the more intuitive three-index representation. The fourth index is actually placed in third position, and instead of being independent of the other three indexes, it is the negative of the sum of the first two:

$$\vec{v} = [h, k, i = -(h + k), l]. \tag{8.9}$$

The fourth index i clearly would be redundant for describing a 3D space, but there is some advantage in using such a representation. Consider, e.g., the crystal planes (110) and $(1\bar{2}0)$. These two planes are actually equivalent, but this is not obvious in their three-index representation. If we instead use the four-index representation, then the same planes are written as $(11\bar{2}0)$ and $(1\bar{2}10)$, which are now easily identified as equivalent upon cyclic permutation of the indexes.

With the four-index representation (Figure 8.1b), the convention is that [0001], the c-axis, is the one perpendicular to the plane with hexagonal symmetry, while the vector $[10\bar{1}0]$ (and the other five vectors obtained by cyclic permutation of the first three indexes) is orthogonal to one side of the hexagon. Hence, it represents the vector normal to one of the reticular planes.

In terms of atomic bonds, one is now oriented in the [0001] direction and this bond is the one which is mostly connected to the piezoelectric polarization of wurtzite crystals. The [0001] direction or c-axis is not the only polar direction in a wurtzite crystal (polarization in the (0001) plane can also be nonzero), but it is the only polar direction for quantum wells that have the c-axis as the growth direction.

8.4.3 Classic Description of Strain-Induced Polarization

In the previous section, we have just outlined how a piezoelectric polarization field with nonzero values exists along the [111] polar direction, which is normally at an angle from the conventional growth axis (001) in zincblende structures. The polarization is present instead along the c-axis direction in an [0001] epitaxially grown coherently strained wurtzite III-N quantum well. For both of these cases, the linear strain can be easily predicted, provided one has knowledge of the elastic constants of the material.

In the remainder of this section, we will make use of two very widely diffused definitions, particularly among experimentalists. The first is the use of $\boldsymbol{\varepsilon}_{\perp}$ to identify the strain component along the growth direction of the crystal. The second is to use $\boldsymbol{\varepsilon}_{\parallel}$ to describe strain in the growth plane, assuming that in such a plane (orthogonal to the growth direction) the strain is perfectly isotropic. The advantage of this notation is that from the point of view of experimental measurements, it is independent of the crystal orientation or even crystal symmetry. We see that this is not the case when one has to determine the strain theoretically, from the elastic properties of the material.

8.4.3.1 Strain-Induced Polarization in Zincblende Crystals

In the case of zincblende, the strain along the growth direction is given by

$$\boldsymbol{\varepsilon}_{\perp} = -2\frac{\mathbf{c}_{12}}{\mathbf{c}_{11}}\boldsymbol{\varepsilon}_{\parallel}. \tag{8.10}$$

This expression is only valid for material with the direction ⊥ aligned with the [001] crystal direction (and || aligned with either of the [100] or [010] crystal directions, assuming that the applied strain is the same for both directions). In the other common case of material with the direction ⊥ aligned with the [111] crystal direction, which is the only one where piezoelectric polarization is present, then the correct expression is

$$\varepsilon_{\perp} = -2\frac{c_{11} + c_{12} + 4c_{44}}{c_{11} + 2c_{12} + c_{44}}\varepsilon_{\|}. \tag{8.11}$$

In this case, it is easy to see how the strain produced by $\varepsilon_{\|}$ is now equivalent to that produced by the off-diagonal components of the strain tensor, which are the only ones that couple with the piezoelectric coefficients of Equation 8.7:

$$P^{zb}_{pz,1} = e_{14}\varepsilon_4 \tag{8.12}$$

$$P^{zb}_{pz,2} = e_{25}\varepsilon_5 \tag{8.13}$$

$$P^{zb}_{pz,3} = e_{36}\varepsilon_6. \tag{8.14}$$

The quantity that links the parallel and perpendicular strain is usually referred to as the Poisson ratio (μ):

$$\varepsilon_{\perp} = -\frac{\varepsilon_{\|}}{\mu}. \tag{8.15}$$

While the expressions given indicate that μ is a constant that only depends on the crystal direction, in reality it is a nonlinear quantity in the strain. We show in later sections that most of the quantities discussed in this chapter are nonlinear in the strain.

8.4.3.2 Strain-Induced Polarization in Wurtzite Crystals

In wurtzite crystals, a free-surface boundary condition for the surface charge ($\sigma_{zz} \equiv \sigma_{33} = 0$) is expected to be present as the polarization is directed in the [0001] or *z*-direction. In this case, we can still use the formalism of Equations 8.10 and 8.15, but the expression for the Poisson ratio (μ) is different compared to the zincblende crystal phase:

$$\varepsilon_{\perp} = -2\frac{c_{13}}{c_{33}}\varepsilon_{\|}. \tag{8.16}$$

This expression is only valid for materials with the direction ⊥ aligned with the [0001] crystal direction (and || aligned with either of the [10$\bar{1}$0] direction and the other five vectors obtained by cyclic permutation of the first three indexes, assuming that the applied strain is the same for all directions). This is the case for the majority of III-N semiconductor optoelectronic devices grown on either sapphire or silicon substrates. Given these definitions, the polarization along, e.g., the axis, originating from a simple tetragonal distortion of the wurtzite crystal, can be expressed as

$$P_{pz,3} = 2\left(e_{31} - \frac{c_{13}}{c_{33}}e_{33}\right)\varepsilon_1 = 2d_{31}\left(c_{11} + c_{12} - 2\frac{c_{13}^2}{c_{33}}e_{33}\right)\varepsilon_1, \tag{8.17}$$

where for the last passage we also used the piezoelectric moduli instead of the piezoelectric coefficient.

For completeness, it is also worth mentioning that an additional term can be added to Equation 8.17, which is linked to the piezoelectric coefficient $e_{15} = e_{24}$. This is linked to shear strain, whereas the normally rectangular facets orthogonal to either the [10$\bar{1}$0] direction or all the other five vectors obtained by cyclic permutation of the first three indexes become more diamond like. This can be important in 3D nanostructures such as quantum dots (Tomić et al., 2015) or flexible nanowires, but not in conventional two-dimensional (2D) structures like the quantum wells typically used for realizing LEDs.

For recommended values of the lattice and elastic parameters to be used for III-N semiconductors in Equations 8.10 through 8.17, we direct the reader to the widely used compendiums provided by Vurgaftman and Meyer (2003) and Romanov et al. (2006), although amendments to the recommended sets are routinely proposed (see, e.g., Pedesseau et al., 2012). A comprehensive discussion of the recommended values of the piezoelectric coefficients will be given in Section 8.5.1.

8.4.4 Spontaneous Polarization in Wurtzite Crystals

Though epitaxially grown wurtzite semiconductors tend to have a sizeable strain-induced piezoelectric polarization, an additional feature not present in zincblende semiconductors is spontaneous polarization, i.e., the polarization that exists even in the absence of any strain, in both III-N and II-VI semiconductors. Spontaneous polarization (like the strain-induced one) cannot be measured directly. It can only be inferred from experimental data of electro-optical properties and their change as a result of strain. This meant that for a long time the only available values were calculated ones. This has changed recently, and later we discuss the recent values of spontaneous polarization proposed in experimental measurements.

Spontaneous polarization develops from the atomic layout in the bulk materials. In the wurtzite structure, the *c*-axis marks the overlap of the neighboring dual tetrahedrons in contrast to the zincblende crystals (Figure 8.1e through h). This arrangement gives rise to closer second-nearest neighbors in the wurtzite compared to the zincblende phase. The modified interatomic forces lead also to a marginal change of the interatomic separation (embodied in the so-called *u* parameter) between the first-nearest neighbors (ionic contribution) together with a rearrangement of the electronic clouds (electronic contribution). The combination of these contributions is that all bonds that are lined up along the [0001] direction become stretched, which results in a nonzero dipole moment along the [0001] crystal direction being present in the absence of any strain. The result of this atomic scale phenomenon is the presence of a macroscopic spontaneous polarization ($P_{3,\mathrm{sp}}$ or simply P_{sp}).

8.5 Piezoelectric Coefficients of Wurtzite III-N Semiconductors

The wurtzite III-N structure demonstrates a peculiar property in having piezoelectric coefficients similar to those of group II–VI materials and quite strikingly opposite to the other group III–V materials. The III-N materials also differ from the other III–V compounds with a larger ionic charge and the internal-strain ionic contribution becomes larger than the clamped-ion term. The fundamental contribution to the understanding of piezoelectric fields in III-N is the list of all piezoelectric coefficients and spontaneous polarization terms provided by Bernardini et al. (1997). Already cited 1500 times according to the APS database, it is impossible to reference each and every paper that has made use of the coefficients. We wish, however, to discuss in some detail the validity of the linear model in view of outlining the now generally agreed need for including nonlinear effects.

8.5.1 Proper versus Improper Coefficients

Before discussing the calculated and measured piezoelectric coefficients it is worth making a distinction between proper and improper piezoelectric coefficients. A very clear explanation is given in Caro et al. (2015b) and a more detailed discussion is available from Vanderbilt (2000) and references therein. The "improper" coefficients link the electric polarization vector and strain through Equation 8.1. An alternative yet equivalent expression is

$$e_{\mathrm{i}\alpha} = \frac{\partial P_{\mathrm{i}}}{\partial \varepsilon_{\alpha}}. \tag{8.18}$$

Such an expression is obviously intuitive as the coefficients express the change in the polarization vector components as a function of the change in one of the strain components.

The "proper" coefficients instead are conceptually more complex as they link the adiabatic (no change in heat) change in the components of the current density vector J in response to a "slow" deformation:

$$e^{\mathrm{p}}_{\mathrm{i}\alpha} = \frac{\partial J_{\mathrm{i}}}{\partial \dot{\varepsilon}_{\alpha}} \tag{8.19}$$

$$\dot{\varepsilon}_{\alpha} = \frac{d\varepsilon_{\alpha}}{dt}. \tag{8.20}$$

"Proper" and "improper" coefficients are generally different and there is a set of transformations which allows one to obtain one set from another. For wurtzite this set is

$$e^{\mathrm{p}}_{15} = e_{15} - \frac{1}{2}P_{3,\mathrm{sp}} \tag{8.21}$$

$$e^{\mathrm{p}}_{31} = e_{31} + P_{3,\mathrm{sp}} \tag{8.22}$$

$$e^{\mathrm{p}}_{33} = e_{33}. \tag{8.23}$$

The "improper" coefficient is the correct one to use when calculating polarization charges, e.g., interfacial charge accumulation in quantum wells (Bernardini et al., 2001). The "proper" coefficient instead is the quantity that can be readily compared to experiments where flowing currents are measured. For systems with small spontaneous polarization the corrections can be negligible, but that is often not the case. In the following sections, for piezoelectric coefficients we always intend improper rather than proper coefficients.

8.5.2 Linear Piezoelectric Coefficients of III-N Semiconductors

The first report on the magnitude of the piezoelectric coefficients of III-N was by Bernardini et al. (1997), followed a few years later by a second report (Bernardini et al., 2001) with updated values. The coefficients (Table 8.1) were calculated using *ab initio* density functional theory and the Berry-phase approach (Vanderbilt, 2000).

TABLE 8.1 Calculated Piezoelectric Coefficients (~ in Units of C/m^2) for Wurtzite III-N

Coefficient	GaN	AlN	InN
P_{sp} (C/m^2)	−0.034 (−0.029)	−0.090 (−0.081)	−0.042 (−0.032)
e^{p}_{31} (C/m^2)	−0.37	−0.62	−0.45
e_{31} (C/m^2)	−0.34 (−0.49)	−0.53 (−0.60)	−0.41 (−0.57)
e_{33} (C/m^2)	0.67 (0.73)	1.50 (1.46)	0.81 (0.97)
e_{15} (C/m^2)	−0.57 Pal et al. (2011a), −0.38 Shimada et al. (1998), −0.29 Pedesseau et al. (2012)	−0.6 Pal et al. (2011a), −0.48 Tsubouchi and Mikoshiba (1985), −0.35 Pedesseau et al. (2012)	−0.65 Pal et al. (2011a), −0.44 Shimada et al. (1998)

Source: Reproduced with permission from Bernardini F et al., *Physical Review B,* 56, R10024–R10027, 1997. Copyright 1997 by the **American Physical Society;** Bernardini F et al., *Physical Review B,* 63, 193201, 2001. Copyright 2001 by the **American Physical Society**.

The superscript p indicates the proper rather than improper coefficient, which for e_{33} has always the same value (Equation 8.23). To complete the table we have also added the values of e_{15} from Pal et al. (2011a) who pointed out the erroneous sign given in Vurgaftman and Meyer (2003) together with proposed values from Shimada et al. (1998), Tsubouchi and Mikoshiba (1985), and Pedesseau et al. (2012). Where available, values in brackets are from Bernardini et al. (1997).

Only the coefficients coupling with strains along the *c*-plane or along the *c*-direction were calculated, i.e., no value for e_{15} was calculated. Values of the spontaneous polarization were given instead.

To complete Table 8.1 we have also added the values of e_{15} from Pal et al. (2011a) who pointed out the erroneous sign given in Vurgaftman and Meyer (2003) together with proposed values from Shimada et al. (1998), Tsubouchi and Mikoshiba (1985), and Pedesseau et al. (2012). It is evident that there is a large spread of proposed values for e_{15} and at the time this writing, it is still unsure which should be the recommended one.

As many have pointed out, the piezoelectric coefficients are one order of magnitude larger than the typical III–V semiconductor. The largest P_{sp}, that of AlN, is only about three to five times smaller than in typical ferroelectric perovskites. Where experimental data were available, the calculated values were in agreement within 5%.

It is worth pointing out that the results were obtained for the binary III–N and device modeling required a further step. In fact, most experimental data are typically obtained for heterostructures and quantum wells or superlattices. In device modeling, the general wisdom was for many years that the properties of the alloys could be obtained by linear interpolation of the binary values, what is commonly known as Vegard's law (Vegard, 1921). Over the years, anecdotal evidence suggested that the use in itself of the linear piezoelectric coefficients of Bernardini et al. (1997), within any linear or parabolic interpolation scheme for alloys, can lead to piezoelectric fields which are substantially larger than those needed to fit experimental data. A rule of thumb employed by many in the device modeling community has, hence, been to scale down the calculated fields by multiplying by an empirical factor found to be roughly 0.7. However, it could well be that the coefficients or model are not to blame at all and that the correction applied to the piezoelectric field simply corrects other factors incorrectly or is not explicitly taken into account in the models. We discuss some of these concepts in Section 8.8.

8.5.3 Nonlinearities in III-N Semiconductors

In this section, we outline the various nonlinear models of piezoelectricity in III-N semiconductors. Such nonlinearities in the polarization were first reported for III-N by Shimada et al. (1998), expressing theoretically and calculating the dependence of the polarization on deviatoric strain, but without expressly providing a set of nonlinear coefficients. These have ever since been the subject of intense debate also in the context of zincblende semiconductors. It is still not clear whether a conclusive and universally accepted model has been achieved yet. The only thing that can be agreed upon is that nonlinearities are large and in many situations can dominate the magnitude and sign of the polarization.

8.5.3.1 Nonlinearities under Biaxial Strain

Ambacher et al. (2002) used *ab initio* calculations and calculated many polarization properties, while also reporting nonlinearities of the bulk polarization of the binary nitrides subject to biaxial strain ε, which is the difference between $\boldsymbol{\varepsilon}_\perp$ and $\boldsymbol{\varepsilon}_\parallel$. The quadratic dependence for [0001] crystals is given as

$$\begin{aligned}
P_{\mathrm{pz,AlN}}(\varepsilon < 0) &= (-1.808\varepsilon + 5.624\varepsilon^2)\ \mathrm{C/m^2} \\
P_{\mathrm{pz,AlN}}(\varepsilon > 0) &= (-1.808\varepsilon - 7.888\varepsilon^2)\ \mathrm{C/m^2} \\
P_{\mathrm{pz,GaN}} &= (-0.918\varepsilon + 9.541\varepsilon^2)\ \mathrm{C/m^2} \\
P_{\mathrm{pz,\,InN}} &= (-1.373\varepsilon + 7.559\varepsilon^2)\ \mathrm{C/m^2}.
\end{aligned} \tag{8.24}$$

In 2002, Equation 8.24 constituted the first report in a usable form of nonlinear piezoelectric polarization in III–V semiconductors. However, the reported nonlinear equations are limited to biaxial strain combinations and are not a full second-order piezoelectric tensor treatment. Furthermore, e_{15} is not included as the equations are explicitly designed for *c*-axis-grown quantum wells where typically e_{15} would not contribute to the total polarization.

8.5.4 Full Second-Order Piezoelectric Tensor Formalism

Before discussing the reported second-order parameters found in the literature, we need to describe here the second-order piezoelectric tensor in some detail. While the 18 components of the first-order strain tensor were given in Equation 8.5, the 18 × 6 = 108 components of the second-order tensor couple to two strain components to generate a second-order contribution to the total polarization

$$P_i'' = \mathrm{B}_{\mathrm{ijklm}}\varepsilon_{\mathrm{jk}}\varepsilon_{lm}. \tag{8.25}$$

The tensor $\mathrm{B}_{\mathrm{ijklm}}$ given in Voigt notation which compresses the indexes jk and lm into the reduced ones α and β, is given by

$$\mathrm{B}_{\mathrm{i}\alpha\beta} = \begin{pmatrix} \mathrm{B}_{11\beta} & \mathrm{B}_{12\beta} & \mathrm{B}_{13\beta} & \mathrm{B}_{14\beta} & \mathrm{B}_{15\beta} & \mathrm{B}_{16\beta} \\ \mathrm{B}_{21\beta} & \mathrm{B}_{22\beta} & \mathrm{B}_{23\beta} & \mathrm{B}_{24\beta} & \mathrm{B}_{25\beta} & \mathrm{B}_{26\beta} \\ \mathrm{B}_{31\beta} & \mathrm{B}_{32\beta} & \mathrm{B}_{33\beta} & \mathrm{B}_{34\beta} & \mathrm{B}_{35\beta} & \mathrm{B}_{36\beta} \end{pmatrix}, \tag{8.26}$$

where $\beta = 1 \dots 6$. Just like in the case of first order, symmetry arguments can be used to reduce the number of nonunique components. It is worth going into some detail to explain how this is done for a wurtzite crystal. We want to stress that in Equation 8.26 the second-order tensor components are defined using Voigt notation, using three subscripts, whereas the unreduced representation would require five subscripts, e.g., $\mathrm{B}_{\mathrm{ijklm}}$. Equation 8.25 too should therefore be expressed in Voigt notation:

$$P_{\mathrm{i}}'' = \mathrm{B}_{\mathrm{i}\alpha\beta}\varepsilon_\alpha\varepsilon_\beta. \tag{8.27}$$

The only second-order piezoelectric coefficients that are nonzero are those fulfilling the crystal symmetry which for hexagonal crystals such as wurtzite are those invariant under a rotation of $\pi/3$ about the *z*-direction (*c*-axis). Reverting to Cartesian coordinates and initially avoiding Voigt notation, we introduce the complex variables

$$\xi = x + \Im y; \quad \eta = x - \Im y. \tag{8.28}$$

These make rotational invariance more strikingly obvious, as under a rotation of $\pi/3$ the following transformations hold:

$$\xi \to \xi e^{\Im\pi/3}; \quad \eta \to \eta e^{-\Im\pi/3}. \tag{8.29}$$

And if, e.g., in the sequence of subscripts ijklm any number of subscripts is transformed according to Equation 8.29, then the associated coefficient would be invariant and nonzero if the transformation yields overall an invariant contribution to the free energy. It is easily verified that this is fulfilled for the following 11 terms: B_{zzzzz}, $\mathrm{B}_{z\xi\eta zz}$, $\mathrm{B}_{zz\xi z\eta}$, $\mathrm{B}_{z\xi\eta\xi\eta}$, $\mathrm{B}_{z\xi\xi\eta\eta}$, $\mathrm{B}_{\eta zzz\xi}$, $\mathrm{B}_{\eta z\eta\xi\xi}$, $\mathrm{B}_{\eta z\xi\eta\xi}$, $\mathrm{B}_{\xi zzz\eta}$, $\mathrm{B}_{\xi z\xi\eta\eta}$, $\mathrm{B}_{\xi z\eta\xi\eta}$. We now show that not all of these 11 terms need to be independently determined. Using the relations

$$\begin{aligned} \xi^2 &= \left(x + \Im y\right)^2 = x^2 - y^2 + 2\Im xy \\ \eta^2 &= \left(x - \Im y\right)^2 = x^2 - y^2 - 2\Im xy \\ \xi\eta &= \left(x + \Im y\right)\left(x - \Im y\right) = x^2 + y^2 \end{aligned} \tag{8.30}$$

the associated strain components can then be assumed to transform as

$$\begin{aligned} \varepsilon_{\xi\xi} &= \varepsilon_{\mathrm{xx}} - \varepsilon_{\mathrm{yy}} + 2\Im\varepsilon_{\mathrm{xy}} \quad \text{Using Method} \\ \varepsilon_{\eta\eta} &= \varepsilon_{\mathrm{xx}} - \varepsilon_{\mathrm{yy}} - 2\Im\varepsilon_{\mathrm{xy}} \\ \varepsilon_{\xi\eta} &= \varepsilon_{\mathrm{xx}} + \varepsilon_{\mathrm{yy}}, \end{aligned} \tag{8.31}$$

while

$$\begin{aligned} \varepsilon_{z\xi} &= \varepsilon_{zx} + \Im\varepsilon_{zy} \\ \varepsilon_{z\eta} &= \varepsilon_{zx} - \Im\varepsilon_{zy}. \end{aligned} \tag{8.32}$$

This now allows us to write the second-order polarization in the z-direction (c-axis), which is invariant for a rotation of $\pi/3$ around the z-direction, as

$$\begin{aligned} P_z'' = {} & B_{zzzzz}\varepsilon_{zz}^2 + B_{z\xi\eta zz}\left(\varepsilon_{\mathrm{xx}} + \varepsilon_{\mathrm{yy}}\right)\varepsilon_{zz} + B_{zz\xi z\eta}\left(\varepsilon_{zx} + \Im\varepsilon_{zy}\right)\left(\varepsilon_{zx} - \Im\varepsilon_{zy}\right) + B_{z\xi\eta\xi\eta}\left(\varepsilon_{\mathrm{xx}} + \varepsilon_{\mathrm{yy}}\right)^2 \\ & + B_{z\xi\xi\eta\eta}\left(\varepsilon_{\mathrm{xx}} - \varepsilon_{\mathrm{yy}} + 2\Im\varepsilon_{\mathrm{xy}}\right)\left(\varepsilon_{\mathrm{xx}} - \varepsilon_{\mathrm{yy}} - 2\Im\varepsilon_{\mathrm{xy}}\right). \end{aligned} \tag{8.33}$$

In order to obtain the c-plane components, we make use of Equations 8.31 and 8.32 and start from

$$P_\eta'' = B_{\eta zzz\xi}\varepsilon_{zz}\left(\varepsilon_{zx} + \Im\varepsilon_{zy}\right) + B_{\eta z\eta\xi\xi}\left(\varepsilon_{zx} - \Im\varepsilon_{zy}\right)\left(\varepsilon_{\mathrm{xx}} - \varepsilon_{\mathrm{yy}} + 2\Im\varepsilon_{\mathrm{xy}}\right) + B_{\eta z\xi\eta\xi}\left(\varepsilon_{zx} + \Im\varepsilon_{zy}\right)\left(\varepsilon_{\mathrm{xx}} + \varepsilon_{\mathrm{yy}}\right) \tag{8.34}$$

and

$$P_\xi'' = B_{\xi zzz\eta}\varepsilon_{zz}\left(\varepsilon_{zx} - \Im\varepsilon_{zy}\right) + B_{\xi z\xi\eta\eta}\left(\varepsilon_{zx} + \Im\varepsilon_{zy}\right)\left(\varepsilon_{\mathrm{xx}} - \varepsilon_{\mathrm{yy}} - 2\Im\varepsilon_{\mathrm{xy}}\right) + B_{\xi z\eta\xi\eta}\left(\varepsilon_{zx} - \Im\varepsilon_{zy}\right)\left(\varepsilon_{\mathrm{xx}} + \varepsilon_{\mathrm{yy}}\right). \tag{8.35}$$

The c-plane components of the polarization P_ξ'' and P_η'' obey the same transformation rules as ξ and η, respectively, for a rotation of $\pi/3$ around the z-direction and using the complex Cartesian coordinates notation of Equation 8.28 we can write

$$\begin{aligned} P_\xi'' &= P_{\mathrm{x}}'' + \Im P_{\mathrm{y}}'' \\ P_\eta'' &= P_{\mathrm{x}}'' - \Im P_{\mathrm{y}}''. \end{aligned} \tag{8.36}$$

We can now rewrite Equations 8.33 through 8.35 as

$$\begin{aligned} 2P_{\mathrm{x}}'' = {} & \left(B_{\eta z\xi\eta\xi} + B_{\xi z\eta\xi\eta} + B_{\eta z\eta\xi\xi} + B_{\xi z\xi\eta\eta}\right)\varepsilon_{zx}\varepsilon_{\mathrm{xx}} \\ & + \left(B_{\eta z\xi\eta\xi} + B_{\xi z\eta\xi\eta} - B_{\eta z\eta\xi\xi} - B_{\xi z\xi\eta\eta}\right)\varepsilon_{zx}\varepsilon_{\mathrm{yy}} \\ & + \left(B_{\eta zzz\xi} + B_{\xi zzz\eta}\right)\varepsilon_{zz}\varepsilon_{zx} + 2\left(B_{\eta z\eta\xi\xi} + B_{\xi z\xi\eta\eta}\right)\varepsilon_{zy}\varepsilon_{\mathrm{xy}} \end{aligned} \tag{8.37}$$

$$\begin{aligned} 2P_{\mathrm{y}}'' = {} & -\left(B_{\eta z\xi\eta\xi} + B_{\xi z\eta\xi\eta} + B_{\eta z\eta\xi\xi} + B_{\xi z\xi\eta\eta}\right)\varepsilon_{zy}\varepsilon_{\mathrm{yy}} \\ & + \left(B_{\eta z\xi\eta\xi} + B_{\xi z\eta\xi\eta} - B_{\eta z\eta\xi\xi} - B_{\xi z\xi\eta\eta}\right)\varepsilon_{zy}\varepsilon_{\mathrm{xx}} \\ & - \left(B_{\eta zzz\xi} + B_{\xi zzz\eta}\right)\varepsilon_{zz}\varepsilon_{zy} - 2\left(B_{\eta z\eta\xi\xi} + B_{\xi z\xi\eta\eta}\right)\varepsilon_{\mathrm{xy}}\varepsilon_{zx} \end{aligned} \tag{8.38}$$

$$\begin{aligned} P_z'' = {} & \left(B_{z\xi\eta\xi\eta} + B_{z\xi\xi\eta\eta}\right)\left(\varepsilon_{\mathrm{xx}}^2 + \varepsilon_{\mathrm{yy}}^2\right) + B_{zzzzz}\varepsilon_{zz}^2 + B_{z\xi\eta zz}\left(\varepsilon_{\mathrm{xx}} + \varepsilon_{\mathrm{yy}}\right)\varepsilon_{zz} \\ & + 2\left(B_{z\xi\eta\xi\eta} - B_{z\xi\xi\eta\eta}\right)\varepsilon_{\mathrm{xx}}\varepsilon_{\mathrm{yy}} + B_{zz\xi z\eta}\left(\varepsilon_{zx}^2 + \varepsilon_{zy}^2\right) + B_{z\xi\xi\eta\eta}\varepsilon_{\mathrm{xy}}^2. \end{aligned} \tag{8.39}$$

Note that the polarization terms on the c-plane (x- and y-directions) can be described by only three independent parameters. Furthermore, five further independent parameters are needed for the c-axis polarization, making a total of eight independent second-order piezoelectric coefficients.

It is worth stressing that after we have transformed the second-order piezoelectric coefficients using subscripts ξ and η, there is no longer a direct correspondence with the coefficients $\mathrm{B}_{\mathrm{ijklm}}$. Rather, it is combinations of the transformed coefficients which equate directly to the $\mathrm{B}_{\mathrm{ijklm}}$, as it is obvious from Equations 8.37 through 8.39, with the correct choice of indexes being determined by which strain components the $\mathrm{B}_{\mathrm{ijklm}}$ coefficients couple to. As an example, in Equation 8.39, the combination of transformed coefficients $\left(B_{z\xi\eta\xi\eta} + B_{z\xi\xi\eta\eta}\right)$ couple to the strain combination $\left(\varepsilon_{\mathrm{xx}}^2 + \varepsilon_{\mathrm{yy}}^2\right)$. Therefore, the correct choice would be $B_{311} = B_{322} = \left(B_{z\xi\eta\xi\eta} + B_{z\xi\xi\eta\eta}\right)$.

Once this is understood, a very useful strategy for visualizing the entire second-order tensor is first, to compact all coefficients $\mathrm{B}_{\mathrm{ijklm}}$ using Voigt notation ($\mathrm{B}_{i\alpha\beta}$), then order them by i along the rows and by αβ (using the invariance between exchanging α and β to avoid repetitions) along the columns of a 3 × 21 matrix, given for both zincblende and wurtzite in Table 8.2, following the formalism of Grimmer (2007).

The resulting matrix contains some very interesting results. Intuitively, one could think that to have a nonzero second-order coefficient, the relevant first order should also be nonzero. That turns out to be incorrect. One example is wurtzite B_{366}. The fact that e_{36} is zero does not make B_{366} automatically zero. In fact, e_{36} is zero not because the dipole symmetry at the atomic level is not broken under strain in the xy-plane $\varepsilon_{\mathrm{xy}}$ (it actually is!) but because symmetry dictates that the polarization along the c-axis must be invariant for positive and negative (shear) strain in the xy-plane. Such condition can only be achieved by the even orders in the polarization, hence it is second (and, e.g., fourth) order that is nonzero while first (and, e.g., third) order are vanishing.

In wurtzite, all the second-order coefficients associated with shear strains obey the invariance rule consistently and the reverse is also true. Take, e.g., the first-order coefficient e_{15}. This coefficient is nonzero because the polarization in the x-direction is not invariant for strain in the yz-plane. In fact, if strain is inverted, the polarization direction is also inverted. On the other hand, second-order coefficients like, e.g., B_{155} are zero because they are instead invariant coefficients that must induce exactly the same polarization for positive and negative values of the strain, a condition that is only fulfilled if the coefficient itself is zero. This is also the case for the zincblende crystal coefficients B_{144}, B_{255}, and B_{366}, which are all vanishing despite e_{14}, e_{25}, and e_{36} being nonzero.

TABLE 8.2 Second-Order Piezoelectric Coefficients $\mathrm{B}_{\mathrm{i}\alpha\beta}$ for the Cubic Zincblende (Point Group $F\bar{4}3m$) and Wurtzite Crystal (Point Group [6 mm])

	11	22	33	44	55	66	12	13	23	45	36	26	16	25	15	35	46	14	24	34	56
Zincblende Crystals																					
1	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	λ	μ	μ	ν
2	–	–	–	–	–	–	–	–	–	–	–	–	–	λ	μ	μ	ν	–	–	–	–
3	–	–	–	–	–	–	–	–	–	ν	λ	μ	μ	–	–	–	–	–	–	–	–
Wurtzite Crystals																					
1	–	–	–	–	–	–	–	–	–	–	–	–	–	$2b$	$2a$	c	$a–b$	–	–	–	–
2	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	$2b$	$2a$	c	$a–b$
3	$2d$	$2d$	g	h	h	$d–e$	$2e$	f	f	–	–	–	–	–	–	–	–	–	–	–	–

Source: Reproduced with permission from Prodhomme P.-Y. et al., *Physical Review B,* 88, 121304, 2013. Copyright [2013] by the **American Physical Society**.

There are three (λ, μ, and μ) and eight (a–h) unique coefficients for the zincblende and wurtzite phases, respectively. A dash indicates vanishing coefficients. Rows and columns correspond to the subscript iαβ in $B_{i\alpha\beta}$.

It has to be mentioned that the majority of second-order coefficients typically couple with vanishing strains in the case of wurtzite *c*-axis grown quantum wells. Therefore, for wurtzite crystals the only coefficients of interest are often $B_{311} = B_{322}$, B_{333}, B_{312} and $B_{313} = B_{323}$. In both the zincblende and wurtzite crystals, first- or second-order coefficients, which couple to at least one shear strain component, only become relevant in nanostructures such as nanowires or quantum dots (Bimberg et al., 1998; Migliorato et al., 2002, 2004, 2005; Bester et al., 2006b; Tartakovskii et al., 2006; Ediger et al., 2007; Ding et al., 2010; Tomić et al., 2015).

8.5.5 Calculated Second-Order Piezoelectric Coefficients

After the initial report of nonlinearities in III-N semiconductors under biaxial strain Ambacher et al. (2002), many more reports appeared in the literature. We cover here a review of calculation approaches for both zincbende and wurtzite crystals but provide a compilation of values only for the III-N semiconductors.

8.5.5.1 Numerical Approaches Used for III–V and II–VI Semiconductors

A theoretical and numerical framework for evaluating the full second-order piezoelectric coefficients in III–V zincblende semiconductor materials was first proposed by Bester et al. (2006a). The method used was the linear response technique within density functional perturbation theory (DFPT). The motivation for this theoretical development was mostly that piezoelectricity in the InGaAs system appeared to be of some importance in determining the electronic properties (particularly in relation to symmetry) of self-assembled quantum dots (Bimberg et al., 1998; Cullis et al., 2002, 2005). Such nanostructures were predicted to exhibit sizeable shear strains (and hence piezoelectricity) even if the material was grown along the [001] crystal direction, where instead, for the same direction of growth, quantum wells exhibited no polarization at all (Migliorato et al., 2002, 2004, 2005; Tartakovskii et al., 2006). The nonlinear model of piezoelectricity appeared to be much more reliable than the linear model in predicting the correct magnitude and size of the resulting piezoelectric fields (Bester et al., 2006b; Ediger et al., 2007; Ding et al., 2010).

An alternative approach to estimating piezoelectric coefficients, based on a method initially proposed by Harrison (1989), was used for uniaxially strained crystals, tetragonal distortion (the strain present during pseudomorphic growth on (001) substrates), and also to determine the magnitude of nonlinearities up to third order in the diagonal terms of the strain tensor (Migliorato et al., 2006, 2014; Garg et al., 2009; Tse et al., 2013). While the only complete set of second-order coefficients is provided in Beya-Wakata et al. (2011) the validity of using the local density approximation within the Berry phase method for calculating polarization properties of semiconductors has recently been called into question. The use of a hybrid functional for the exchange correlation interaction in DFPT appears to yield results closer to experimental values (Caro et al., 2015a). Harrison's model was also recently employed to predict the nonlinear piezoelectric properties of ZnO (Al-Zahrani et al., 2013) and some wurtzite III-As and III-P semiconductors (Al-Zahrani et al., 2015).

8.5.5.2 Numerical Approaches Used for Wurtzite III-N Semiconductors

For III-N semiconductors, reports in 2011 from Pal et al. (2011a, b, c) provided a more general set of second-order coefficients (Table 8.3) compared to the case of biaxial strain of Equation 8.24, based on Harrison's method. The coefficients, numerical values relate to the following second-order equation:

$$\mathbf{P}_{\mathbf{Tot,\,3}} = \mathbf{P}_{\mathbf{3,sp}} + \mathbf{e}_{\mathbf{33}}\boldsymbol{\varepsilon}_{\perp} + 2\mathbf{e}_{\mathbf{31}}\boldsymbol{\varepsilon}_{\parallel} + \tilde{\mathbf{B}}_{\mathbf{311}}\boldsymbol{\varepsilon}_{\parallel}^{2} + \tilde{\mathbf{B}}_{\mathbf{333}}\boldsymbol{\varepsilon}_{\perp}^{2} + \tilde{\mathbf{B}}_{\mathbf{313}}\boldsymbol{\varepsilon}_{\parallel}\boldsymbol{\varepsilon}_{\perp} \tag{8.40}$$

Equation 8.40, together with the coefficients of Table 8.3, predicts large deviations from the linear model (Figure 8.2) and appears to have had some success at explaining trends in the internal quantum efficiency (IQE) in blue and green LEDs under external pressure. In particular, it was claimed that the linear model would incorrectly predict an increase in IQE under compressive pressure (associated with a decrease of

TABLE 8.3 Linear and Partial Nonlinear Piezoelectric Coefficients for III-N Semiconductors Calculated Using Harrison's Method

Coefficient (C/m^2)	GaN	AlN	InN
e_{31}	−0.55	−0.60	−0.55
e_{33}	1.05	1.47	1.07
e_{15}	−0.57	−0.60	−0.65
$\tilde{B}_{311}$	6.185	5.850	5.151
$\tilde{B}_{333}$	−8.090	−10.75	−6.680
$\tilde{B}_{313}$	1.543	4.533	1.280

Source: Adapted with permission from Pal J et al., *Physical Review B,* 84, 2011a. Copyright [2011] by the **American Physical Society**.

Please note in the original paper the $\tilde{B}_{313}$ was erroneously labeled as $\tilde{B}_{133}$.

$\mathbf{P}_{\mathbf{Tot},3}$), when both experiment and the second-order model instead agreed in registering a decrease in IQE, due to an increase of $\mathbf{P}_{\mathbf{Tot},3}$ (Crutchley et al., 2013; Pal et al., 2013).

A word of warning is necessary here. The coefficients of Table 8.3 do not directly equate to the $B_{i\alpha\beta}$ coefficients of Table 8.2 (and were obviously not intended to!), and we have used the symbol (~) as a cap for $\tilde{B}_{i\alpha\beta}$ to emphasize this (in the original paper the notation $e_{i\alpha\beta}$ was used for the same purpose). This is because Equation 8.40 uses a simplified expression valid for the case of $\varepsilon_{||} = \varepsilon_1 = \varepsilon_2$. The $B_{i\alpha\beta}$ and $\tilde{B}_{i\alpha\beta}$ coefficients are, however, easily linked by

$$\tilde{B}_{311} = B_{311} + 2B_{312} + B_{322} = 4B_{311} \tag{8.41}$$

$$\tilde{B}_{313} = 2B_{313} + 2B_{323} = 4B_{313} \tag{8.42}$$

$$\tilde{B}_{333} = B_{333}. \tag{8.43}$$

A complete set of second-order coefficients for AlN and GaN (Table 8.4) was provided by Pedesseau et al. (2012) who used perturbation density functional theory within the linear response method. The $B_{i\alpha\beta}$ coefficients are the same as those of Table 8.2. The authors reported that compared to the linear model, second-order piezoelectricity leads to a substantially increased prediction of the electric field in pseudomorphic AlGaN/GaN quantum wells, by up to 35% (Figure 8.3).

The full second-order tensor for III-N and ZnO was later (2013) revisited by Prodhomme et al. (2013), using a very similar method to the one used by Pedesseau et al. (2012). Coefficients are given (Table 8.5) in a slightly different style than Table 8.4 but consistent with Table 8.2. Despite the supposed equivalency of the methods used in the calculations, there are many similarities between the values of Tables 8.4 and 8.5, but also many differences in the proposed values. It is outside the scope of this chapter to discuss the merits or demerits of the different proposed coefficients, and we limit our analysis to acknowledging that such differences exist. Furthermore, on close scrutiny, the differences are not significantly large and are unlikely to result in radically different total polarizations.

8.5.6 Calculated versus Experimental Values of the Spontaneous Polarization

Different calculated values for the P_{sp} terms for III-N semiconductors are listed in Table 8.6. It is self-evident that there is a divergence between calculated values, and also between calculated and measured values. This is because they are both difficult calculations to undertake and troublesome experimental determinations to perform.

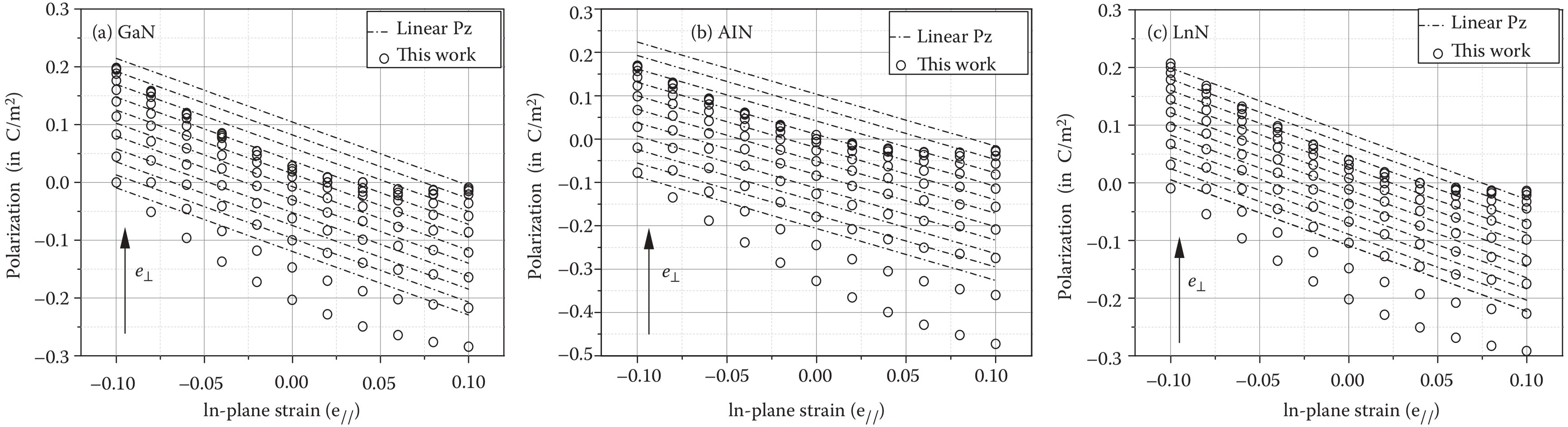


FIGURE 8.2 Comparison of the total polarization as a function of perpendicular and parallel strain calculated in this work (circles) and using the linear model with parameters from Bernardini et al. (2001) (dashed lines). The perpendicular strain varies from −0.1 to 0.1 in steps of 0.02. (Adapted with permission from Pal J et al., *Physical Review B*, 84, 2011a. Copyright [2011] by the **American Physical Society**.)

TABLE 8.4 Linear and Complete Nonlinear Piezoelectric Coefficients for III-N Semiconductors

Coefficient	GaN	AlN
e_{31} (C/m^2)	−0.41	− 0.67
e_{33} (C/m^2)	0.76	1.67
e_{15} (C/m^2)	−0.29	− 0.35
B_{311} (C/m^2)	6.26	3.77
B_{312} (C/m^2)	2.37	4.12
B_{313} (C/m^2)	−1.47	− 8.64
B_{333} (C/m^2)	−23.43	−25.60
B_{115} (C/m^2)	6.48	8.69
B_{125} (C/m^2)	5.09	5.19
B_{135} (C/m^2)	7.02	3.20
B_{344} (C/m^2)	1.41	2.88

Source: Reproduced with permission from Pedesseau L et al., *Applied Physics Letters* 100:031903. Copyright [2012], American Institute of Physics Publishing LLC.

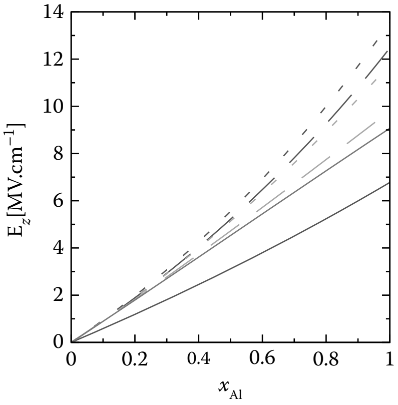


FIGURE 8.3 Variation of the electric field E_z as a function of Al mole fraction for various models. The results of the uncoupled and coupled linear models are represented by straight (first [bottom] curve) and dashed (fifth curve) lines, the coupling strongly enhancing E_z. The linear semi-coupled “standard” model (From Jogai B et al., *Journal of Applied Physics* 94, 3984, 2003) is represented by a dotted line (sixth curve). The straight line (second curve) includes all the nonlinear contributions. The electrostrictive and the nonlinear piezoelectric ones are represented by dashed (third curve) and dotted (fourth curve) lines. (Reproduced with permission from Pedesseau L et al., *Applied Physics Letters*, 100,031903, 2012. Copyright [2012], American Institute of Physics Publishing LLC.)

The values from Pal et al. (2011a) are calculated from the ionic contribution only and neglect entirely the electronic contribution, which leads to a large underestimation compared to that from, e.g., Bernardini et al. (1997, 2001), between −0.020 and −0.040 C/m^2 for the various III-N materials. Despite this, the predicted spontaneous polarization of AlN of −0.051 C/m^2 is the theoretical value that is the closest to the only reported measured one of −0.040 C/m^2 (Park and Chuang, 2000). For GaN there is a spread of calculated values but there are at least two reported experimental values (Yan et al., 2009; Lähnemann et al., 2012) which remarkably agree with each other and are also not too different from the values of both

TABLE 8.5 Piezoelectric Coefficients in C/m^2

	$2a$	$\mathbf{2b}$	$\mathbf{C}$	$2d$	$2e$	f	g	h	e_{15}	e_{31}	e_{33}	$\mathbf{P_{sp}}$
AlN	4.4	2.4	−0.1	3.0	3.0	3.8	−26.0	3.2	−0.35	−0.67	1.67	−0.095
GaN	3.8	2.3	2.7	6.2	3.3	0.4	−21.4	0.4	−0.31	−0.44	0.75	−0.027
InN	4.5[±0.4]	2.8[±0.2]	1.6[±0.4]	4.8	3.7	0.5[±0.1]	−18.6	0.5	−0.43	−0.59	1.14	−0.042

Source: Reproduced with permission from Prodhomme P-Y et al., *Physical Review B,* 88, 121304, 2013. Copyright [2013] by the **American Physical Society**.

Error bars for the values obtained for InN are due to the poor description of this material within DFT-LDA.

TABLE 8.6 Calculated Values of the Spontaneous Polarization (P_{sp}) in C/m^2 According to Different Reports

	Bernardini et al. (2001)	Pal et al. (2011a)	Pedesseau et al. (2012)	Prodhomme et al. (2013)	Experimental
GaN	−0.034	−0.007	0.014	−0.027	−70142 ± 0.007 Lähnemann et al. (2012) −0.022 Yan et al. (2009)
AlN	−0.090	−0.051	−0.075	−0.095	−0.040 Park and Chuang (2000)
InN	−0.042	−0.012	N/A	−0.035	N/A

The last column presents some recently reported experimental values.

Bernardini et al. (1997, 2001) and Prodhomme et al. (2013). To date, no experimental values are available for InN.

The experimental technique used by Lähnemann et al. (2012) is rather elegant. They performed optical measurements of polymorphic structures, i.e., mixed zincblende and wurtzite phases within the same material. The formation of such polymorphs arises simply from occasional stacking faults in otherwise perfect wurtzite crystals, which create zincblende inclusions typically of 1–3 atomic layers. Such naturally forming inclusions, due to a difference in the band offsets of the zincblende and wurtzite phases, can confine electrons within the zincblende regions, allowing them to be optically probed. The general strategy of experimentally measuring the electrooptical behavior of polymorphic structures is a very notable and welcome step forward, one likely in the near future to provide more quantitative data to further our understanding of piezoelectricity in semiconductors.

8.5.7 Continuity Equation in Superlattices and Quantum Well Structures

For completeness, it is worth mentioning how the coefficients outlined in the previous sections can be used to evaluate the electric field in quantum well structures or superlattices.

The conventional equations are obtained from the condition of continuity of the electrostatic potential at the interfaces, often referred to as the superlattice equations, where the subscripts w and b refer to the quantum well and barrier region, respectively, where the barrier region is the area between the wells in a periodic superlattice:

$$E_{3,\mathrm{w}} = \frac{L_\mathrm{b}\left(P_{\mathrm{sp},3,\mathrm{b}} + P_{\mathrm{pz},3,\mathrm{b}} - P_{\mathrm{sp},3,\mathrm{w}} - P_{\mathrm{pz},3,\mathrm{w}}\right)}{L_\mathrm{b}\lambda_\mathrm{w} + L_\mathrm{w}\lambda_\mathrm{b}} \tag{8.44}$$

$$E_{3,\mathrm{b}} = -E_{3,\mathrm{w}}\frac{L_\mathrm{w}}{L_\mathrm{b}}. \tag{8.45}$$

The subscript $3 = z$ is the c-axis direction, L indicates the width of the well or barrier region and, to avoid confusion with the symbol used for the strain components, λ indicates the relative dielectric constant $\varepsilon_r\varepsilon_0$

of the well or barrier region. Note that if the barrier or well material is lattice matched to the substrate then $P_{\text{pz},3,\text{b}} = 0$ or $P_{\text{pz},3,\text{w}} = 0$ (because the strain is vanishing). On the other hand, even in such a case, Equations 8.45 and 8.46 correctly indicate that both $E_{3,\text{w}}$ and $E_{3,\text{b}}$ are simultaneously nonzero. This is because a nonzero field in one region needs to be compensated by one in the other region, to maintain continuity of the displacement vector D.

Equations 8.44 and 8.45 describe superlattices but can easily be turned into the expression for a single isolated quantum well, by imposing the condition $L_\text{b} \gg L_\text{w}$. In this case, in fact, Equation 8.45 simply gives $E_{3,\text{b}} = 0$, while Equation 8.41 becomes

$$E_{3,\text{w}} = \frac{\left(P_{\text{sp},3,\text{b}} + P_{\text{pz},3,\text{b}} - P_{\text{sp},3,\text{w}} - P_{\text{pz},3,\text{w}}\right)}{\lambda_\text{w} + \left[(L_\text{w}/L_\text{b})\lambda_\text{b}\right]} \approx \frac{\left(P_{\text{sp},3,b} + P_{\text{pz},3,b} - P_{\text{sp},3,\text{w}} - P_{\text{pz},3,w}\right)}{\lambda_\text{w}}, \tag{8.46}$$

which can assume an even simpler form in the case where the barrier material is lattice matched to the substrate and hence $P_{\text{pz},3,\text{b}} = 0$.

8.5.8 Schemes for Dealing with Alloys

It should be noted that barrier and well regions in Equations 8.44 through 8.46, or any nanostructure for the matter, are not necessarily formed by binary III-N, and, in fact, in the majority of cases of interest at least one will be constituted by an alloy. In the case of alloys a suitable interpolation scheme needs to be employed for all the different quantities involved. Vegard's law (linear interpolation) is naturally the first scheme to consider. The validity of Vegard's law was tested for III-N semiconductor quantum wells by Ambacher et al. (2002) and found to be insufficient in many situations, which led to proposing a quadratic dependence on composition. For the sake of completeness we reproduce the proposed equations as they form the basis for device modeling of structures such as those used for fabrication of LEDs. We start with those related to the structural properties.

The lattice constants $a(x)$ and $c(x)$ of wurtzite group III nitride alloys $\text{A}_\text{x}\text{B}_{1-x}\text{N}$ ($\text{Al}_\text{x}\text{Ga}_{1-x}\text{N}$, $\text{In}_\text{x}\text{Ga}_{1-x}\text{N}$, and $\text{Al}_\text{x}\text{In}_{1-x}\text{N}$), can be assumed to follow Vegard's law:

$$a_{\text{AlGaN}}(x) = (3.1986 - 0.0891x)\text{Å} \tag{8.47}$$
$$a_{\text{InGaN}}(x) = (3.1986 + 0.3862x)\text{Å}$$
$$a_{\text{AlInN}}(x) = (3.5848 - 0.4753x)\text{Å} \quad \text{Lasers}$$
$$c_{\text{AlGaN}}(x) = (5.2262 - 0.2323x)\text{Å} \tag{8.48}$$
$$c_{\text{InGaN}}(x) = (5.2262 + 0.5740x)\text{Å}$$
$$c_{\text{AlInN}}(x) = (5.8002 - 0.8063x)\text{Å}.$$

These results were found to be within 2% of experimental data obtained through a combination of high-resolution x-ray diffraction, Rutherford, and elastic recoil detection analysis. In the same work authors also recorded various types of nonlinear behavior. First, the u parameter (which expresses the deviation of the unit cell from the ideal structure) appears to have a quadratic dependence and hence does not follow Vegard's law:

$$u_{\text{AlGaN}}(x) = (0.3819x + 0.3772(1-x) - 0.0032x(1-x)) \tag{8.49}$$
$$u_{\text{InGaN}}(x) = (0.3793x + 0.3772(1-x) - 0.0057x(1-x))$$
$$u_{\text{AlInN}}(x) = (0.3819x + 0.3793(1-x) - 0.0086x(1-x)).$$

Such quadratic and hence nonlinear behavior is unsurprisingly found also in the polarization properties. The spontaneous polarization terms for random, ternary alloys with wurtzite crystal structure are fitted by high accuracy through quadratic equations (Figure 8.4), also following earlier work by Bernardini and Fiorentini (2002), and are thus given by

$$P_{\text{sp, AlGaN}}(x) = (-0.090x - 0.034(1-x) + 0.021x(1-x))\ \text{C/m}^2 \tag{8.50}$$

$$P_{\text{sp,InGaN}}(x) = (-0.042x - 0.034(1-x) + 0.037x(1-x))\ \text{C/m}^2$$

$$P_{\text{sp,AlInN}}(x) = (-0.090x - 0.042(1-x) + 0.070x(1-x))\ \text{C/m}^2,$$

while the strain-induced polarization, which obviously depends on the alloy composition but also on the substrate, is given by

$$P_{\text{pz, AlGaN/InN}}(x) = [-0.28x - 0.113(1-x) + 0.042x(1-x)]\text{C/m}^2 \tag{8.51}$$

$$P_{\text{pz,AlGaN/GaN}}(x) = [-0.0525x + 0.0282x(1-x)]\text{C/m}^2$$

$$P_{\text{pz,AlGaN/AlN}}(x) = [0.026\,(1-x) + 0.0248x\,(1-x)]\text{C/m}^2$$

$$P_{\text{pz, InGaN/InN}}(x) = [-0.113(1-x) - 0.0276x(1-x)]\text{C/m}^2 \tag{8.52}$$

$$P_{\text{pz,InGaN/GaN}}(x) = [0.148x - 0.0424x(1-x)]\text{C/m}^2$$

$$P_{\text{pz,InGaN/AlN}}(x) = [0.182x + 0.026(1-x) - 0.0456x(1-x)]\text{C/m}^2$$

$$P_{\text{pz,AlInN/InN}}(x) = [-0.28x + 0.104x(1-x)]\text{C/m}^2 \tag{8.53}$$

$$P_{\text{pz,AlInN/GaN}}(x) = [-0.0525x + 0.148(1-x) + 0.0938x(1-x)]\text{C/m}^2$$

$$P_{\text{pz,AlInN/AlN}}(x) = [0.182(1-x) + 0.092x(1-x)]\text{C/m}^2.$$

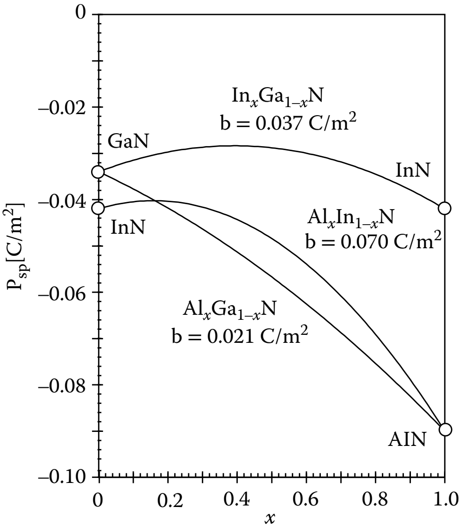


FIGURE 8.4 Predicted spontaneous polarization (P_{sp}) for random, ternary alloys with wurtzite crystal structure. The dependence of P_{sp} on x can be approximated to a high degree of accuracy by quadratic equations (Equation 8.47). The nonlinearity can be described by positive bowing parameters increasing from AlGaN to InGaN and AlInN. (Reproduced with permission Ambacher O et al., *Journal of Physics: Condensed Matter*, 2002 © IOP Publishing. All rights reserved.)

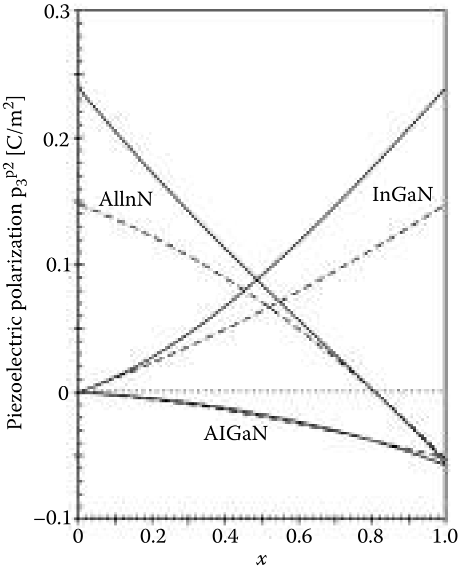


FIGURE 8.5 The piezoelectric polarization of random ternary alloys pseudomorphically grown on relaxed GaN buffer layers calculated by (1) Equation 8.17 and using linear interpolations of the physical properties for the relevant binary compounds (dashed curves); (2) taking into account the nonlinearity in the piezoelectric polarization in terms of strain (Equation 8.24) (solid curves). For alloys under high biaxial strain (layers with high In concentration), the piezoelectric polarization is underestimated by approach (1). (Reproduced with permission Ambacher O et al., *Journal of Physics: Condensed Matter*, 2002 © IOP Publishing. All rights reserved.)

The piezoelectric polarization of a random ternary III-N alloy, pseudomorphically grown on relaxed GaN buffer layers is given in Figure 8.5, where it is possible to compare the polarization calculated by (1) Equation 8.17 and that obtained using Vegard's law (dashed curves) or (2) taking into account the nonlinearity in the piezoelectric polarization in terms of strain (Equation 8.24) (solid curves). In general, the polarization using Vegard's law is always underestimated.

Pal et al. (2011a, c) also proposed an interpolation scheme (Table 8.7) for the piezoelectric properties of the ternary nitride alloys ($Al_xGa_{1-x}N$, $In_xGa_{1-x}N$), which can also be applied to any nanostructure, consistent with their proposed first- and second-order coefficients and values of the spontaneous polarization. A parabolic equation was proposed.

Finally, Prodhomme et al. (2013) also provided a further mixing scheme for alloys, only for the linear coefficients (second-order coefficients are assumed to follow Vegard's law instead). The corresponding bowing parameters are listed in Table 8.8, and the effect of using this mixing scheme can be observed in Figure 8.6. The mixing scheme is quantitatively within 10% of that of Ambacher et al. (2002).

8.6 Electromechanical Coupling in Wurtzite Semiconductors

It has recently been reported by Suski et al. (2012) that when attempting to model variations of optical output in III-N quantum wells, the combination of linear elasticity with linear piezoelectricity appears to be lacking the correct framework to provide predictions that compare well with available experimental data (Figure 8.7). Linear elasticity should be replaced by the more comprehensive nonlinear elasticity (Łepkowski and Majewski, 2006), where electromechanical coupling is included.

However, this is still not sufficient. It was noted that the slope of dE_{PL}/dP, i.e., the change in peak energy (as measured from photoluminescence or calculated from **k·p** theory) for a change in applied

TABLE 8.7 Quadratic Dependence on Alloy Composition for the Ternary Nitride Alloys ($Al_x\, Ga_{1-x}\, N$, $In_x\, Ga_{1-x}\, N$)

	Y	***A***	***B***	***C***
$\mathbf{Al_xGa_{1-x}N}$	$\mathbf{P_{sp}}$	−0.025	−0.019	−0.007
	$\mathbf{e_{31}}$	0.064	−0.114	−0.550
	$\mathbf{e_{33}}$	0.141	0.279	1.050
	$\mathbf{\tilde{B}_{311}}$	0.674	−1.000	6.185
	$\mathbf{\tilde{B}_{333}}$	1.055	−3.715	−8.090
	$\mathbf{\tilde{B}_{313}}$	0.340	2.650	1.543
$\mathbf{In_xGa_{1-x}N}$	$\mathbf{P_{sp}}$	−0.001	−0.005	−0.007
	$\mathbf{e_{31}}$	−0.368	0.368	−0.550
	$\mathbf{e_{33}}$	0.119	−0.099	1.050
	$\mathbf{\tilde{B}_{311}}$	0.635	−1.669	6.185
	$\mathbf{\tilde{B}_{333}}$	1.182	0.228	−8.090
	$\mathbf{\tilde{B}_{313}}$	0.226	−0.489	1.543

Source: Adapted with permission from Pal J et al., *Physical Review B,* 84, 2011a; Pal J et al., *Physical Review B*, 84, 159902, 085211, 2011c. Copyright [2011] by the **American Physical Society**.

The parameters are for the equation $Y = Ax^2 + Bx + C$.

TABLE 8.8 Bowing Coefficients (MV/cm^2) for the Nonlinear Dependence on Ternary Composition for the Linear Piezoelectric Coefficients and Lattice Mismatch (Fifth Column) and Piezoelectric Coefficients Difference between Binaries (Sixth Column)

	$\mathbf{b_{15}}$	$\mathbf{b_{15}}$	$\mathbf{b_{15}}$	$\mathbf{\Delta a / \langle a \rangle}$	$\mathbf{\Delta e_{33} / \langle e_{33} \rangle}$
GaN/AlN	−0.032	0.106	−0.671	2.23%	75%
GaN/InN	−0.275	0.222	−0.550	10.4%	41%
InN/AlN	−0.305	0.540	−2.178	12.65%	38%

Source: Reproduced with permission from Prodhomme P-Y et al., *Physical Review B*, 88, 121304, 2013. Copyright [2013] by the **American Physical Society**.

pressure, as a function of the unstrained peak energy $E_{PL,P=0}$, could only match experimental data by combining both nonlinear elasticity and nonlinear piezoelectricity simultaneously in the model. The nonlinear piezoelectric parameters of Pal et al. (2011a, c) were used for the demonstration (Figure 8.7). Given the importance of this development we will discuss the nonlinear elasticity model in some detail, first with linear piezoelectric coefficients, then with nonlinear ones.

8.6.1 Electromechanical Coupling with Linear Piezoelectricity

The starting point is the definition of the density of enthalpy in a material with strain-induced polarization and under external hydrostatic pressure, given by

$$H = U - E \cdot D + PdV, \tag{8.54}$$

where P is the external (hydrostatic) pressure, $dV = \mathrm{Tr}(\varepsilon_{\mathrm{ij}})$ is the normalized infinitesimal volume change, E is the electric field, and D is the displacement vector such that

$$E_{\mathrm{i}} \cdot D_{\mathrm{i}} = E_{\mathrm{i}} \cdot \left(e_{\mathrm{ijk}} \varepsilon_{\mathrm{jk}} + \lambda_{\mathrm{ij}} E_{\mathrm{j}} + P_{sp,i} \right) = e_{\mathrm{ijk}} \varepsilon_{\mathrm{jk}} E_{\mathrm{i}} + \lambda_{\mathrm{ij}} E_{\mathrm{i}} E_{\mathrm{j}} + E_{\mathrm{i}} \cdot P_{sp,i}, \tag{8.55}$$

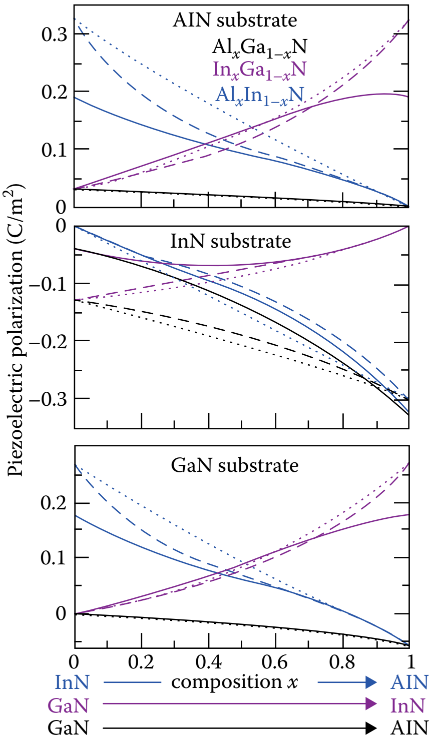


FIGURE 8.6 Piezoelectric polarization with respect to the ternary compositions for different compounds on different (0001) substrates. The curves are Ppz (dotted line), Ppz with bowing (dashed line), and Ppz with bowing + second-order Ppz (full line). (Reproduced with permission from Prodhomme P-Y et al., *Physical Review B*, 88, 121304, 2013. Copyright [2013] by the **American Physical Society**.)

where λ_{ij} is the electric permittivity tensor and the usual symbols are used for the piezoelectric coefficients and spontaneous polarization. The tensor U in Equation 8.54 is instead the elastic energy (per unit volume) defined as

$$U = \frac{1}{2}\sum_{\mathrm{ijkl}}\left(c_{\mathrm{ijkl}}\varepsilon_{\mathrm{ij}}\varepsilon_{\mathrm{kl}} + \lambda_{\mathrm{ij}}E_{\mathrm{i}}E_{\mathrm{j}}\right), \tag{8.56}$$

where c_{ijkl} are the components of the elastic tensor. The stress tensor components instead are obtained as derivatives of the enthalpy with respect to the strain components, which yields

$$\begin{aligned}\sigma_{ij} &= \frac{\mathrm{d}H}{\mathrm{d}\varepsilon_{\mathrm{ij}}} = \\ &= \frac{\mathrm{d}}{\mathrm{d}\varepsilon_{\mathrm{ij}}}\left[\sum_{\mathrm{ijkl}}\left(\frac{1}{2}c_{\mathrm{ijkl}}\varepsilon_{\mathrm{ij}}\varepsilon_{\mathrm{kl}} - \frac{1}{2}\lambda_{\mathrm{ij}}E_{\mathrm{i}}E_{\mathrm{j}} - e_{\mathrm{ijk}}\varepsilon_{\mathrm{jk}}E_{\mathrm{i}} - E_{\mathrm{i}}\cdot P_{sp,i} + PTr(\varepsilon_{\mathrm{ij}})\right)\right] \\ &= \sum_{\mathrm{kl}}\left(c_{\mathrm{ijkl}}\varepsilon_{\mathrm{kl}} - e_{\mathrm{kij}}E_{\mathrm{k}}\right) + P\delta_{\mathrm{ij}},\end{aligned} \tag{8.57}$$

where δ_{ij} is the conventional Kronecker delta. We now need to consider the symmetry of the wurtzite crystal and explicitly take into account that there are only five independent elastic constants and three

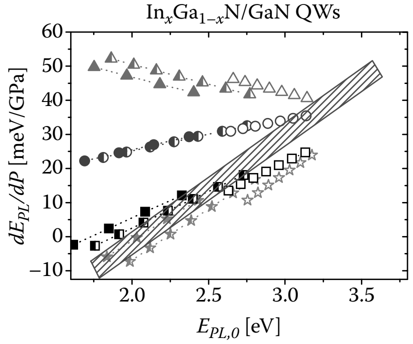


FIGURE 8.7 Comparison between the experimentally determined dependence of $\frac{dE_{PL}}{dP} vsE_{PL,P=0}$ and the results of k·p calculations performed for $In_xGa_{1-x}N/GaN$ quantum wells with $x = 0.1$ (empty symbols), 0.2 (half-filled symbols), and 0.3 (solid symbols). Triangles, circles, squares, and stars correspond to theoretical results obtained using (1) linear piezoelectricity following (From Bernardini F et al., *Physical Review B*, 56, R10024–R10027, 1997; Bernardini F et al., *Physical Review B*, 63, 193201, 2001) and linear elasticity, (2) second-order piezoelectricity following (From Pal J et al., *Physical Review B*, 84, 2011a; Pal J et al., *Physical Review B*, 84, 159902, 085211, 2011c) and linear elasticity, (3) second-order piezoelectricity (Pal et al., 2011a, c) and nonlinear elasticity (From Łepkowski SP, Majewski JA, *Physical Review B*, 74, 8, 2006), (4) nonlinear piezoelectricity following (From Shimada K et al., *Japanese Journal of Applied Physics*, 37, L1421, 1998) and nonlinear elasticity (Łepkowski and Majewski, 2006), respectively. The hatched rectangles represent available experimental results. (Reproduced with permission from Suski T et al., *Journal of Applied Physics*, 112, 053509, 2012. Copyright 2012 by the **American Physical Society**.)

independent piezoelectric coefficients, as discussed in relation to Equation 8.2. Using Voigt notation the result is

$$\begin{bmatrix}\sigma_1\\\sigma_2\\\sigma_3\\\sigma_4\\\sigma_5\\\sigma_6\end{bmatrix} = \Lambda_i - \Pi_i' + P_i = \begin{bmatrix}c_{11}\varepsilon_1 + c_{12}\varepsilon_2 + c_{13}\varepsilon_3\\ c_{12}\varepsilon_1 + c_{11}\varepsilon_2 + c_{13}\varepsilon_3\\ c_{13}\left(\varepsilon_1 + \varepsilon_2\right) + c_{33}\varepsilon_3\\ 2c_{44}\varepsilon_4\\ 2c_{44}\varepsilon_5\\ \left(c_{11} - c_{12}\right)\varepsilon_6\end{bmatrix} - \begin{bmatrix}e_{31}E_3\\ e_{31}E_3\\ e_{33}E_3\\ e_{15}E_2\\ e_{15}E_1\\ 0\end{bmatrix} + \begin{bmatrix}P\\P\\P\\0\\0\\0\end{bmatrix}, \tag{8.58}$$

having indicated with Λ_i, Π_i', and P_i the elastic, piezoelectric, and external pressure contribution, respectively.

If we assume that any stress component along the *c*-axis (*z*-direction) is equal to zero, which is reasonable, since, typically in that direction, the material is free to relax to minimize its enthalpy, then from the third row of Equation 8.55 we can write

$$\varepsilon_3 = \frac{-c_{13}\left(\varepsilon_1 + \varepsilon_2\right)}{c_{33}} - \frac{P}{c_{33}} + \frac{e_{33}E_3}{c_{33}}. \tag{8.59}$$

The first term of Equation 8.59 is the expected linear expression for linking parallel and perpendicular strain of Equation 8.16, the second term expresses the external pressure, while the third term, given by

$$\Omega = \frac{e_{33}E_3}{c_{33}}, \tag{8.60}$$

expresses the electromechanical coupling. If we now incorporate Equation 8.52 into the continuity equations (Equation 8.59), where the assumption of isotropic relative dielectric constant is made, and solve for E_z in the well region, we obtain that

$$E_{3,\mathrm{w}} = \frac{L_\mathrm{b}\left[P_{\mathrm{sp},3,\mathrm{b}} + \left(e_{31,\mathrm{b}} - \frac{c_{13,\mathrm{b}}}{c_{33,\mathrm{b}}} e_{33,\mathrm{b}}\right)\left(\varepsilon_{1,\mathrm{b}} + \varepsilon_{2,\mathrm{b}}\right) - P_{\mathrm{sp},3,\mathrm{w}} - \left(e_{31,\mathrm{w}} - \frac{c_{13,\mathrm{w}}}{c_{33,\mathrm{w}}} e_{33,\mathrm{w}}\right)\left(\varepsilon_{1,\mathrm{w}} + \varepsilon_{2,\mathrm{w}}\right)\right]}{L_\mathrm{b}\left(\lambda_\mathrm{w} + \frac{e_{33,\mathrm{w}}^2}{c_{33,\mathrm{w}}}\right) + L_\mathrm{w}\left(\lambda_\mathrm{b} + \frac{e_{33,\mathrm{b}}^2}{c_{33,\mathrm{b}}}\right)}$$

$$+ P\frac{L_\mathrm{b}\left(\frac{e_{33,\mathrm{w}}}{c_{33,\mathrm{w}}} - \frac{e_{33,\mathrm{b}}}{c_{33,\mathrm{b}}}\right)}{L_\mathrm{b}\left(\lambda_\mathrm{w} + \frac{e_{33,\mathrm{w}}^2}{c_{33,\mathrm{w}}}\right) + L_\mathrm{w}\left(\lambda_\mathrm{b} + \frac{e_{33,\mathrm{b}}^2}{c_{33,\mathrm{b}}}\right)} = E_{3,\mathrm{w}}^{*} + PE_{3,\mathrm{w}}^{**}, \tag{8.61}$$

The field in the barrier is easily obtained from Equation 8.45. It is possible to extract the strain ε_3 from Equation 8.61, which shows how the strain is changed compared to Equation 8.59

$$\varepsilon_{3,\mathrm{w}} = \frac{-c_{13,\mathrm{w}}\left(\varepsilon_{1,\mathrm{w}} + \varepsilon_{2,\mathrm{w}}\right)}{c_{33,\mathrm{w}}} + \frac{e_{33,\mathrm{w}} E_{3,\mathrm{w}}^{*}}{c_{33,\mathrm{w}}} - \frac{P}{c_{33,\mathrm{w}}}\left(e_{33,\mathrm{w}} E_{3,\mathrm{w}}^{**} - 1\right). \tag{8.62}$$

While Equations 8.61 and 8.62 entirely define the problem, it is worth noting that such formalism simplifies considerably if we assume that L_b is much larger than L_w. Assuming also $\varepsilon_1 = \varepsilon_2$, we obtain the isolated quantum well configuration, which can be written as

$$E_{3,\mathrm{w}}^{*} = \frac{1}{\kappa_\mathrm{w}}\left[P_{\mathrm{sp},3,\mathrm{b}} + 2\left(e_{31,\mathrm{b}} - \frac{c_{13,\mathrm{b}}}{c_{33,\mathrm{b}}} e_{33,\mathrm{b}}\right)\varepsilon_{1,\mathrm{b}} - P_{\mathrm{sp},3,\mathrm{w}} - 2\left(e_{31,\mathrm{w}} - \frac{c_{13,\mathrm{w}}}{c_{33,\mathrm{w}}} e_{33,\mathrm{w}}\right)\varepsilon_{1,\mathrm{w}}\right] \tag{8.63}$$

$$E_{3,\mathrm{w}}^{**} = \frac{1}{\kappa_\mathrm{w}}\left(\frac{e_{33,\mathrm{w}}}{c_{33,\mathrm{w}}} - \frac{e_{33,\mathrm{b}}}{c_{33,\mathrm{b}}}\right), \tag{8.64}$$

having defined

$$\kappa_\mathrm{w} = \lambda_\mathrm{w} + \frac{e_{33,\mathrm{w}}^2}{c_{33,\mathrm{w}}}. \tag{8.65}$$

8.6.2 Electromechanical Coupling with Nonlinear Piezoelectricity

Including nonlinear piezoelectricity (to second order) in the electromechanical coupling is fairly straightforward. This is not given explicitly elsewhere and is original to this chapter.

This first step is to add to Equation 8.55 the second-order piezoelectric polarization, given by

$$E_\mathrm{i} \cdot D_\mathrm{i}'' = B_{\mathrm{i}\alpha\beta}\varepsilon_\alpha\varepsilon_\beta E_\mathrm{i}, \tag{8.66}$$

which adds a further tensor to the expression of the stress tensor of Equation 8.58

$$\sigma_\mathrm{i} = \Lambda_\mathrm{i} - \Pi_\mathrm{i}' + P_\mathrm{i} - \Pi_\mathrm{i}'', \tag{8.67}$$

with Π_i'' given by

$$\Pi_i'' = \begin{bmatrix} B_{115}\varepsilon_5 E_1 + B_{125}\varepsilon_4 E_2 + \left(2B_{311}\varepsilon_1 + B_{312}\varepsilon_2 + B_{313}\varepsilon_3\right) E_3 \\ B_{125}\varepsilon_5 E_1 + B_{115}\varepsilon_4 E_2 + \left(B_{312}\varepsilon_1 + 2B_{311}\varepsilon_2 + B_{313}\varepsilon_3\right) E_3 \\ B_{135}\left(\varepsilon_5 E_1 + \varepsilon_4 E_2\right) + \left(B_{313}\varepsilon_1 + B_{313}\varepsilon_2 + 2B_{333}\varepsilon_3\right) E_3 \\ B_{146}\varepsilon_6 E_1 + \left(B_{125}\varepsilon_1 + B_{115}\varepsilon_2 + B_{135}\varepsilon_3\right) E_2 + 2B_{344}\varepsilon_4 E_3 \\ \left(B_{115}\varepsilon_1 + B_{125}\varepsilon_2 + B_{135}\varepsilon_3\right) E_1 + B_{146}\varepsilon_6 E_2 + 2B_{344}\varepsilon_5 E_3 \\ B_{146}\left(\varepsilon_4 E_1 + \varepsilon_5 E_2\right) + 2B_{366}\varepsilon_6 E_3 \end{bmatrix}. \tag{8.68}$$

In the absence of shear strain (or if these are sufficiently small to be neglected), and following the same steps of the previous section, the addition of second-order piezoelectric terms adds a further electromechanical coupling to Equations 8.59 and 8.60, leading to

$$\Omega = \frac{e_{33}E_3}{c_{33}} + \left(\frac{B_{313}\left(\varepsilon_1 + \varepsilon_2\right)}{c_{33}} + \frac{2B_{333}\varepsilon_3}{c_{33}}\right) E_3. \tag{8.69}$$

And Equations 8.58 through 8.62 are easily modified if you replace each first-order coefficient e_{33} with

$$\tilde{e}_{33} = e_{33} + B_{313}\left(\varepsilon_1 + \varepsilon_2\right) + 2B_{333}\varepsilon_3 \tag{8.70}$$

leading to

$$\begin{aligned} E_{3,\mathrm{w}} &= \frac{L_\mathrm{b}\left[P_{\mathrm{sp},3,\mathrm{b}} + \left(e_{31,\mathrm{b}} - \frac{c_{13,\mathrm{b}}}{c_{33,\mathrm{b}}}\tilde{e}_{33,\mathrm{b}}\right)\left(\varepsilon_{1,\mathrm{b}} + \varepsilon_{2,\mathrm{b}}\right) - P_{\mathrm{sp},3,\mathrm{w}} - \left(e_{31,\mathrm{w}} - \frac{c_{13,\mathrm{w}}}{c_{33,\mathrm{w}}}\tilde{e}_{33,\mathrm{w}}\right)\left(\varepsilon_{1,\mathrm{w}} + \varepsilon_{2,\mathrm{w}}\right)\right]}{L_\mathrm{b}\left(\lambda_\mathrm{w} + \frac{\tilde{e}^2_{33,\mathrm{w}}}{c_{33,\mathrm{w}}}\right) + L_\mathrm{w}\left(\lambda_\mathrm{b} + \frac{\tilde{e}^2_{33,\mathrm{b}}}{c_{33,\mathrm{b}}}\right)} \\ &\quad + P\frac{L_\mathrm{b}\left(\frac{\tilde{e}_{33,\mathrm{w}}}{c_{33,\mathrm{w}}} - \frac{\tilde{e}_{33,\mathrm{b}}}{c_{33,\mathrm{b}}}\right)}{L_\mathrm{b}\left(\lambda_\mathrm{w} + \frac{\tilde{e}^2_{33,\mathrm{w}}}{c_{33,\mathrm{w}}}\right) + L_\mathrm{w}\left(\lambda_\mathrm{b} + \frac{\tilde{e}^2_{33,\mathrm{b}}}{c_{33,\mathrm{b}}}\right)} = E^*_{3,\mathrm{w}} + PE^{**}_{3,\mathrm{w}} \end{aligned} \tag{8.71}$$

$$\begin{aligned} \varepsilon_{3,\mathrm{w}} &= \frac{-c_{13,\mathrm{w}}\left(\varepsilon_{1,\mathrm{w}} + \varepsilon_{2,\mathrm{w}}\right)}{c_{33,\mathrm{w}}} + \frac{\left(e_{33,\mathrm{w}} + B_{313,\mathrm{w}}\left(\varepsilon_{1,\mathrm{w}} + \varepsilon_{2,\mathrm{w}}\right) + 2B_{333,\mathrm{w}}\varepsilon_{3,\mathrm{w}}\right)E^*_{3,\mathrm{w}}}{c_{33,w}} \\ &\quad - \frac{P}{c_{33,\mathrm{w}}}\left(\left(e_{33,\mathrm{w}} + B_{313,\mathrm{w}}\left(\varepsilon_{1,\mathrm{w}} + \varepsilon_{2,\mathrm{w}}\right) + 2B_{333,\mathrm{w}}\varepsilon_{3,\mathrm{w}}\right)E^{**}_{3,\mathrm{w}} - 1\right). \end{aligned} \tag{8.72}$$

For the isolated quantum well configuration, where $\varepsilon_1 = \varepsilon_2$, we finally obtain

$$\begin{aligned} \tilde{\kappa}_w E^*_{3,\mathrm{w}} &= P_{\mathrm{sp},3,\mathrm{b}} + 2\left(e_{31,\mathrm{b}} - \frac{c_{13,\mathrm{b}}}{c_{33,\mathrm{b}}}\left(e_{33,\mathrm{b}} + 2B_{313,\mathrm{b}}\varepsilon_{1,\mathrm{w}} + 2B_{333,\mathrm{b}}\varepsilon_{3,\mathrm{b}}\right)\right)\varepsilon_{1,\mathrm{b}} \\ &\quad - P_{\mathrm{sp},3,\mathrm{w}} - 2\left(e_{31,\mathrm{w}} - \frac{c_{13,\mathrm{w}}}{c_{33,\mathrm{w}}}\left(e_{33,\mathrm{w}} + 2B_{313,\mathrm{w}}\varepsilon_{1,\mathrm{w}} + 2B_{333,\mathrm{w}}\varepsilon_{3,\mathrm{w}}\right)\right)\varepsilon_{1,\mathrm{w}} \end{aligned} \tag{8.73}$$

$$\tilde{\kappa}_{\mathrm{w}} E^{**}_{3,\mathrm{w}} = \frac{\left(e_{33,\mathrm{w}} + 2B_{313,\mathrm{w}}\varepsilon_{1,\mathrm{w}} + 2B_{333,\mathrm{w}}\varepsilon_{3,\mathrm{w}}\right)}{c_{33,\mathrm{w}}} - \frac{\left(e_{33,\mathrm{b}} + 2B_{313,\mathrm{b}}\varepsilon_{1,\mathrm{w}} + 2B_{333,\mathrm{b}}\varepsilon_{3,\mathrm{b}}\right)}{c_{33,\mathrm{b}}}, \tag{8.74}$$

where $\tilde{\kappa}_{\mathrm{w}}$ is now given by

$$\tilde{\kappa}_{\mathrm{w}} = \lambda_{\mathrm{w}} + \frac{\left(e_{33,\mathrm{w}} + 2B_{313,\mathrm{w}}\varepsilon_{1,\mathrm{w}} + 2B_{333,\mathrm{w}}\varepsilon_{3,\mathrm{w}}\right)^2}{c_{33,\mathrm{w}}}. \tag{8.75}$$

Attempting the substitutions makes it rather complicated to obtain an analytic solution, as both $E^{*}_{3,\mathrm{w}}$ and $E^{**}_{3,\mathrm{w}}$ now depend on ε_3. Therefore, a numerical solution is the most efficient if not the only realistically possible approach.

8.7 Nonpolar Orientations in Wurtzite Semiconductors

III-N–based heterostructures grown along the polar *c*-axis exhibit large piezoelectric field that can localize electrons at interfaces and create reduced efficiency through Coulomb repulsion. There has thus been substantial interest in growing structures on semi- and nonpolar substrates. If the growth direction of the heterostructure forms a nonzero angle with the polar *c*-axis, even a complete elimination of the electrostatic built-in field can occur (Romanov et al., 2006).

With this motivation, Schulz and Marquardt (2015) have calculated a very useful set of equations to evaluate the polarization of III-N for any substrate orientation at an angle θ from the *c*-axis.

$$\begin{aligned} P'_1 &= -P_{\mathrm{sp}} \sin\theta - 2e_{15}\cos\theta \cdot \left[\left(\varepsilon'_1 - \varepsilon'_3\right)\cos\theta\sin\theta - \varepsilon'_5\cos 2\theta\right] \\ &- e_{31}\sin\theta \cdot \left[\varepsilon'_2 + \varepsilon'_1\cos^2\theta + \varepsilon'_5\sin 2\theta + \varepsilon'_3\sin^2\theta\right] \\ &- e_{33}\sin\theta \cdot \left[\varepsilon'_3\cos^2\theta - \varepsilon'_5\sin 2\theta + \varepsilon'_3\sin^2\theta\right] \end{aligned} \tag{8.76}$$

$$P'_2 = 2e_{15} \cdot \left[\varepsilon'_4\cos\theta - \varepsilon'_6\sin\theta\right] \tag{8.77}$$

$$\begin{aligned} P'_3 &= P_{\mathrm{sp}}\cos\theta - 2e_{15}\sin\theta\left[\left(\varepsilon'_2 - \varepsilon'_3\right)\cos\theta\sin\theta - \varepsilon'_5\cos 2\theta\right] \\ &+ e_{31}\cos\theta\left[\varepsilon'_2 + \varepsilon'_1\cos^2\theta + \varepsilon'_5\sin 2\theta + \varepsilon'_3\sin^2\theta\right] \\ &+ e_{33}\cos\theta\left[\varepsilon'_3\cos^2\theta - \varepsilon'_3\sin 2\theta + \varepsilon'_1\sin^2\theta\right]. \end{aligned} \tag{8.78}$$

The strain components ε'_i are the ones after the θ rotation, with the *i* Cartesian coordinates defined in terms of the new system of reference. Equations 8.76 through 8.78 yield linear piezoelectric polarization only, but the extension to second order is straightforward.

8.8 Summary and Future Directions

Piezoelectricity in III-N semiconductors is an effect of very strong magnitude, with fields measured in the MV/cm a commonly reported feature. It affects device performance, and modeling must always take these fields into account. From the early linear models, nonlinear models have been developed, and even electromechanical coupling can now be included. DFPT and the Berry phase approach have been invaluable in obtaining reliable estimates of first- and second-order piezoelectric coefficients. At the time of this writing, a large variation in proposed values and discrepancies with experimental data still exists with regard to the values of e_{15} and spontaneous polarization, which must be addressed in the future.

The recent development in the experimental realization and characterization of polymorphic structures has already provided accurate and direct determination of the values of the spontaneous polarization of GaN (Lähnemann et al., 2012). In the near future, polymorphism might spear even more reliable experimental data to inform our theoretical understanding of piezoelectric polarization in semiconductor materials.

Though there is no reason to doubt the current models, it is still very difficult to reliably simulate full device behavior. Too often one still must rely on some level of fitting parameters to match experimental data. The reason for this is not clear. We mentioned earlier that an often empirical solution to the fitting of experimental optical data is to scale piezoelectric fields by multiplying times 0.7. Since there is hardly any scope in the models or calculated parameters for such a large correction, it is unlikely that the coefficients or models alone, whether linear or nonlinear, are responsible for less-than-accurate predictions of the optical properties. It is much more likely that the scaling of the piezoelectric field simply corrects other factors in what are otherwise complex calculations involving nonhomogeneous ternary alloys and estimates of the electronic bandstructure, quantum confined wavefunctions, and Coulomb repulsion. There might well be several factors that are incorrectly or not explicitly taken into account in the models. A nonexhaustive list includes:

1. Real quantum wells don’t have sharp interfaces as typically assumed in the models.
2. Alloy disorder can have a more complex influence on electro-optical properties than the models can at present incorporate.
3. LEDs suffer from efficiency droop, which is most often simulated based on the assumption of electron leakage, which may not be correct.
4. The relationship between piezoelectric fields and Auger recombination rates is unknown.

In relation to alloy disorder, the second item on the list, some interesting discussion can be found in Yang et al. (2014). An idea that could help understand the underlying mechanisms of real device behavior is that any conduction of electrical current in alloys takes place through percolation, i.e., carriers always moving from one end to the other of the material following the path of least resistance. This means that electrons do not see the alloy as a homogeneous medium but feel predominantly the presence of the clusters of atoms that most help them continue on their path. The pockets or clusters of atoms with the largest piezoelectric fields are the most helpful in this process. However, percolation also means that the effective path from one electrode to another is no longer the distance between them. The path becomes much longer! Device behavior depends on the balance between these effects.

Though this concept of percolation-driven carrier diffusion would be a major step forward in treating alloys, one of the aspects that is still missing in evaluating polarization properties is a full atomistic treatment over realistic size models of at least tens of nanometers. Since, almost invariably, the experimental data available involves alloys and disorders which can play a significant and underestimated role in affecting real-life measurements, future work in this field must take alloy disorder fully into account.

In relation to electron leakage models and Auger recombination, the fourth item on the list, this is a very important and current topic, but outside the scope of this chapter. Interested readers should consult, e.g., Piprek (2015) and Piprek et al. (2015).

References

Al-Zahrani HYS, Pal J, Migliorato MA (2013) Non-linear piezoelectricity in wurtzite ZnO semiconductors. *Nano Energy* 2:1214–1217.

Al-Zahrani HYS, Pal J, Migliorato MA, Tse G, Yu D (2015) Piezoelectric field enhancement in III–V core–shell nanowires. *Nano Energy* 14:382–391.

Ambacher O, Majewski J, Miskys C, Link A, Hermann M, Eickhoff M, Stutzmann M, Bernardini F, Fiorentini V, Tilak V, Schaff B, Eastman LF (2002) Pyroelectric properties of Al(In)GaN/GaN hetero—And quantum well structures. *Journal of Physics: Condensed Matter* 14:3399.

Bernardini F, Fiorentini V (2002) Nonlinear behavior of spontaneous and piezoelectric polarization in III–V nitride alloys. *Physica Status Solidi (A)* 190:65–73.

Bernardini F, Fiorentini V, Vanderbilt D (1997) Spontaneous polarization and piezoelectric constants of III-V nitrides. *Physical Review B* 56:R10024–R10027.

Bernardini F, Fiorentini V, Vanderbilt D (2001) Accurate calculation of polarization-related quantities in semiconductors. *Physical Review B* 63:193201.

Bester G, Wu X, Vanderbilt D, Zunger A (2006a) Importance of second-order piezoelectric effects in zinc-blende semiconductors. *Physical Review Letters* 96:187602.

Bester G, Zunger A, Wu X, Vanderbilt D (2006b) Effects of linear and nonlinear piezoelectricity on the electronic properties of InAs/GaAs quantum dots. *Physical Review B* 74:081305.

Beya-Wakata A, Prodhomme P-Y, Bester G (2011) First- and second-order piezoelectricity in III–V semiconductors. *Physical Review B* 84:195207.

Bimberg D, Grundmann M, Ledentsov NN (1998) *Quantum Dot Heterostructures*. Chichester: Wiley.

Caro MA, Schulz S, O'Reilly EP (2015a) Origin of nonlinear piezoelectricity in III–V semiconductors: Internal strain and bond ionicity from hybrid-functional density functional theory. *Physical Review B* 91:075203.

Caro MA, Zhang S, Riekkinen T, Ylilammi M, Moram MA, Lopez-Acevedo O, Molarius J, Laurila T (2015b) Piezoelectric coefficients and spontaneous polarization of ScAlN. *Journal of Physics: Condensed Matter* 27:245901.

Crutchley BG, Marko IP, Pal J, Migliorato MA, Sweeney SJ (2013) Optical properties of InGaN-based LEDs investigated using high hydrostatic pressure dependent techniques. *Physica Status Solidi (B)* 250:698–702.

Cullis AG, Norris DJ, Migliorato MA, Hopkinson M (2005) Surface elemental segregation and the Stranski–Krastanow epitaxial islanding transition. *Applied Surface Science* 244:65–70.

Cullis AG, Norris DJ, Walther T, Migliorato MA, Hopkinson M (2002) Stranski–Krastanow transition and epitaxial island growth. *Physical Review B* 66:081305.

Curie J, Curie P (1880) Contractions et dilatations produites par des tensions dans les cristaux hémièdres à faces inclinées *Comptes Rendus de l'Académie du Sciences* 93:40.

Ding F, Singh R, Plumhof JD, Zander T, Křápek V, Chen YH, Benyoucef M, Zwiller V, Dörr K, Bester G, Rastelli A, Schmidt OG (2010) Tuning the exciton binding energies in single self-assembled InGaAs/GaAs quantum dots by piezoelectric-induced biaxial stress. *Physical Review Letters* 104:067405.

Ediger M, Bester G, Gerardot BD, Badolato A, Petroff PM, Karrai K, Zunger A, Warburton RJ (2007) Fine structure of negatively and positively charged excitons in semiconductor quantum dots: Electron-hole asymmetry. *Physical Review Letters* 98:036808.

Garg R, Hüe A, Haxha V, Migliorato MA, Hammerschmidt T, Srivastava GP (2009) Tunability of the piezoelectric fields in strained III–V semiconductors. *Applied Physics Letters* 95:041912.

Grimmer H (2007) The piezoelectric effect of second order in stress or strain: Its form for crystals and quasicrystals of any symmetry. *Acta Crystallographica Section A* 63:441–446.

Harrison W (1989) *Electronic Structure and the Properties of Solids: The Physics of the Chemical Bond*. New York, NY: Dover Publications.

Jogai B, Albrecht JD, Pan E (2003) Effect of electromechanical coupling on the strain in AlGaN/GaN heterojunction field effect transistors. *Journal of Applied Physics* 94:3984.

Kopf RF (2003) State-of-the-Art Program on Compound Semiconductors XXXIX and Nitride and Wide Bandgap Semiconductors for Sensors, Photonics and Electronics IV. *Proceedings of the International Symposia*. Pennington, NJ: Electrochemical Society.

Lähnemann J, Brandt O, Jahn U, Pfüller C, Roder C, Dogan P, Grosse F, Belabbes A, Bechstedt F, Trampert A, Geelhaar L (2012) Direct experimental determination of the spontaneous polarization of GaN. *Physical Review B* 86:081302.

Łepkowski SP, Majewski JA (2006) Effect of electromechanical coupling on the pressure coefficient of light emission in group-III nitride quantum wells and superlattices. *Physical Review B* 74:035336.

Lippmann G (1881) Principe de la conservation de l'électricité, ou second principe de la théorie des phénomènes électriques. *Journal de Physique Théorique et Appliquée* 10:14.

Migliorato MA, Cullis AG, Fearn M, Jefferson JH (2002) Atomistic simulation of $In_xGa_{1-x}As/GaAs$ quantum dots with nonuniform composition. *Physica E: Low-dimensional Systems and Nanostructures* 13:1147–1150.

Migliorato MA, Pal J, Garg R, Tse G, Al-Zahrani HYS, Monteverde U, Tomić S, Li CK, Wu YR, Crutchley BG, Marko IP, Sweeney SJ (2014) A review of non linear piezoelectricity in semiconductors. *AIP Conference Proceedings* 1590:32–41.

Migliorato MA, Powell D, Cullis AG, Hammerschmidt T, Srivastava GP (2006) Composition and strain dependence of the piezoelectric coefficients in $In_xGa_{1-x}As$ alloys. *Physical Review B* 74 Article ID: 245332.

Migliorato MA, Powell D, Liew SL, Cullis AG, Navaretti P, Steer MJ, Hopkinson M, Fearn M, Jefferson JH (2004) Influence of composition on the piezoelectric effect and on the conduction band energy levels of $In_xGa_{1-x}As/GaAs$ quantum dots. *Journal of Applied Physics* 96:5169.

Migliorato MA, Powell D, Zibik EA, Wilson LR, Fearn M, Jefferson JH, Steer MJ, Hopkinson M, Cullis AG (2005) Anisotropy of the electron energy levels in $In_xGa_{1-x}As/GaAs$ quantum dots with non uniform composition. *Physica E: Low-Dimensional Systems and Nanostructures* 26:436–440.

Nobel (2014) *Laureates in Physics*. http://www.nobelprize.org/nobel_prizes/physics/laureates/2014/

Nye JF (1957) *Physical Properties of Crystals*. Oxford: Clarendon.

Pal J, Migliorato MA, Li CK, Wu YR, Crutchley BG, Marko IP, Sweeney SJ (2013) Enhancement of efficiency of InGaN-based light emitting diodes through strain and piezoelectric field management. *Journal of Applied Physics* 114:073104.

Pal J, Tse G, Haxha V, Migliorato MA, Tomić S (2011a) Second-order piezoelectricity in wurtzite III-N semiconductors. *Physical Review B* 84 Article ID: 085211.

Pal J, Tse G, Haxha V, Migliorato MA, Tomić S (2011b) Importance of non linear piezoelectric effect in Wurtzite III-N semiconductors. *Optical and Quantum Electronics* 44:9.

Pal J, Tse G, Haxha V, Migliorato MA, Tomić S (2011c) Erratum: Second-order piezoelectricity in wurtzite III-N semiconductors. *Physical Review B* 84:159902, 085211.

Park S-H, Chuang S-L (2000) Spontaneous polarization effects in wurtzite GaN/AlGaN quantum wells and comparison with experiment. *Applied Physics Letters* 76:1981.

Pedesseau L, Katan C, Even J (2012) On the entanglement of electrostriction and non-linear piezoelectricity in non-centrosymmetric materials. *Applied Physics Letters* 100:031903.

Piprek J (2015) How to decide between competing efficiency droop models for GaN-based light-emitting diodes. *Applied Physics Letters* 107:031101.

Piprek J, Römer F, Witzigmann B (2015) On the uncertainty of the Auger recombination coefficient extracted from InGaN/GaN light-emitting diode efficiency droop measurements. *Applied Physics Letters* 106:101101.

Prodhomme P-Y, Beya-Wakata A, Bester G (2013) Nonlinear piezoelectricity in wurtzite semiconductors. *Physical Review B* 88:121304.

Romanov AE, Baker TJ, Nakamura S, Speck JS, Group EJU (2006) Strain-induced polarization in wurtzite III-nitride semipolar layers. *Journal of Applied Physics* 100:023522.

Schulz S, Marquardt O (2015) Electronic structure of polar and semipolar (1122)-oriented nitride dot-in-a-well systems. *Physical Review Applied* 3:064020.

Shimada K, Sota T, Suzuki K, Okumura H (1998) First-principles study on piezoelectric constants in strained BN, AlN, and GaN. *Japanese Journal of Applied Physics* 37:L1421.

Singh J (1992) *Physics of Semiconductors and Their Heterostructures* (McGraw Hill Series in Electrical and Computer Engineering). Blacklick, OH: McGraw-Hill College.

Singh J (2003) *Electronic and Optoelectronic Properties of Semiconductor Structures*. Cambridge: Cambridge University Press.

Suski T, Łepkowski SP, Staszczak G, Czernecki R, Perlin P, Bardyszewski W (2012) Universal behavior of photoluminescence in GaN-based quantum wells under hydrostatic pressure governed by built-in electric field. *Journal of Applied Physics* 112:053509.

Tartakovskii AI, Kolodka RS, Liu HY, Migliorato MA, Hopkinson M, Makhonin MN, Mowbray DJ, Skolnick MS (2006) Exciton fine structure splitting in dot-in-a-well structures. *Applied Physics Letters* 88:131115.

Tomić S, Pal J, Migliorato MA, Young RJ, Vukmirović N (2015) Visible spectrum quantum light sources based on $In_xGa_{1-x}N/GaN$ quantum dots. *ACS Photonics* 2:958–963.

Tse G, Pal J, Monteverde U, Garg R, Haxha V, Migliorato MA, Tomicì S (2013) Non-linear piezoelectricity in zinc blende GaAs and InAs semiconductors. *Journal of Applied Physics* 114:073515.

Tsubouchi K, Mikoshiba N (1985) Zero-temperature-coefficient SAW devices on AlN epitaxial films. *IEEE Transactions on Sonics and Ultrasonics* 32:634–644.

Vanderbilt D (2000) Berry-phase theory of proper piezoelectric response. *Journal of Physics and Chemistry of Solids* 61:147–151.

Vegard L (1921) Die Konstitution der Mischkristalle und die Raumfüllung der Atome. *Zeitschrift für Physik* 5:10.

Voigt W (1910) *Lehrbuch der Kristallphysik*. Leipzig: B. G. Teubner.

Vurgaftman I, Meyer JR (2003) Band parameters for nitrogen-containing semiconductors. *Journal of Applied Physics* 94:3675–3696.

Wang ZL (2007) Nanopiezotronics. *Advanced Materials* 19:4.

Yan WS, Zhang R, Xie ZL, Xiu XQ, Han P, Lu H, Chen P, Gu SL, Shi Y, Zheng YD, Liu ZG (2009) A thermodynamic model and estimation of the experimental value of spontaneous polarization in a wurtzite GaN. *Applied Physics Letters* 94:042106.

Yang T-J, Shivaraman R, Speck JS, Wu Y-R (2014) The influence of random indium alloy fluctuations in indium gallium nitride quantum wells on the device behavior. *Journal of Applied Physics* 116:113104.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

9

Dilute Nitride Alloys

9.1 Introduction.. 251
9.2 Theory of the Electronic Structure of Dilute Nitride Alloys 254
First Principles and Empirical Pseudopotential Calculations • Band-Anticrossing: Empirical, Tight-Binding, and **k·p** Models • Tight-Binding and Alloy Disorder: Linear Combination of Isolated Nitrogen States
9.3 Theory of GaAs-Based 1.3-μm InGaNAs Quantum Well Lasers.. 269
Theoretical Model for Dilute Nitride Quantum Well Lasers • Impact of Nitrogen Incorporation on the Band Structure and Gain of Quantum Wells • Impact of Nitrogen Incorporation on Carrier Recombination: Theory versus Experiment
9.4 Alternative Applications of Dilute Nitride Alloys.................. 281
GaAs-Based 1.55-μm Quantum Well Lasers • Semiconductor Optical Amplifiers, Electro-Absorption Modulators, and Avalanche Photodiodes • Monolithic Integration of III–V Lasers on Silicon
9.5 Theory of the Density of States and Carrier Transport in Dilute Nitride Alloys 287
The Anderson Impurity Model: Self-Consistent Green's Function Method • Electron Transport in GaN$_x$As$_{1-x}$ Alloys
9.6 Summary and Conclusions.. 303

Christopher A. Broderick
Masoud Seifikar
Eoin P. O'Reilly
and
Judy M. Rorison

9.1 Introduction

The exponential growth of optical telecommunications and the internet has been underpinned by the development of semiconductor lasers emitting at 1.3 and 1.55 μm, the respective wavelengths at which dispersion is zero and losses minimized, in standard silica optical fibers [1,2]. Lasers designed to emit at these wavelengths are based primarily on the growth of quaternary InGaAsP or AlInGaAs quantum well (QW) structures on InP substrates. However, despite their widespread applications there are a number of significant limitations associated with incumbent 1.3- and 1.55-μm technologies, several of which are associated with the constraints of growing on InP substrates [3]. This has stimulated significant interest in the development of telecom-wavelength lasers grown on GaAs substrates, which has the potential to deliver a series of key advantages. First, because GaAs is a more robust material, growth can be carried out on larger substrates. Second, growth of reliable semiconductor lasers and related devices on GaAs opens up the possibility of monolithic integration of high-performance photonic components with GaAs-based high-speed electronics. Third, from a fundamental perspective, better optical confinement can be achieved in GaAs-based structures compared to those based on InP, because of the larger refractive index contrast between GaAs and AlGaAs compared to that between InP and the InGaAsP or AlInGaAs quaternary alloys. This means that, for example, vertical-cavity surface-emitting lasers (VCSELs) can be grown monolithically on GaAs, but only with great difficulty on InP. Additionally, because AlGaAs has a considerably larger band gap than InP and can be grown lattice-matched on GaAs, it is possible to achieve significantly better carrier confinement in a GaAs-based laser structure, thereby making it possible to overcome the losses associated with carrier leakage from which InP-based devices are known to suffer under high injection currents and at high temperatures.

Considerable advantages could therefore be gained if high-quality and reliable 1.3- and 1.55-μm lasers can be developed on GaAs substrates. However, this is difficult to achieve using conventional QW structures. Highly efficient lasers based on compressively strained InGaAs QWs are well established for emission at wavelengths close to 1 μm [4], but to extend the emission to longer wavelengths requires further incorporation of indium (In) leading to a large lattice mismatch relative to GaAs. As such, critical thickness issues become important in determining the material quality, leading to a degradation in performance and a general inability to achieve reliable InGaAs/GaAs QW lasers emitting at 1.3 μm and beyond [1]. A number of approaches have been proposed and investigated to overcome these issues, in order to extend the wavelength range accessible using the GaAs platform. These include the development of InGaAs quantum dot (QD) lasers [5,6], and metamorphic QW structures incorporating relaxed InGaAs buffer layers to facilitate the growth of heterostructures having lattice constants intermediate between those of GaAs and InP [7–9].

In this chapter, we focus on an alternative approach: the use of dilute nitride alloys—III–V semiconductor alloys containing small fractions of substitutional nitrogen (N) atoms—to achieve long-wavelength emission in materials having lattice constants compatible with growth on GaAs. Interest in dilute nitride alloys for applications in photonic devices originally grew from the observation that replacing a small fraction of the arsenic (As) atoms with N in (In)GaAs leads to an extremely rapid reduction of band gap, by up to 150 meV per % N in GaN_xAs_{1-x} [10–12]. Strong interest in the fundamental aspects of the physics of dilute nitride alloys, combined with their potential for practical applications, has stimulated a significant body of research on the theory and simulation of N-containing semiconductors over the past 15 or so years. In this chapter—the first of two on highly mismatched III–V semiconductor alloys—we trace the development of the theory of dilute nitride alloys, ranging from fundamental investigations of the electronic structure of ordered and disordered alloys at an atomistic level to continuum models suited to modeling the properties of dilute nitride–based photonic devices.

Since N is much smaller and more electronegative than arsenic (As), it acts as an isovalent impurity when substituted in dilute concentrations in (In)GaAs to form the $(In)GaN_xAs_{1-x}$ alloy. Furthermore, due to its smaller size relative to As, incorporation of N results in a simultaneous reduction of the band gap and lattice constant with increasing x in $(In)GaN_xAs_{1-x}$. This trend is uncommon in III–V semiconductor alloys, in which a reduction in band gap is typically associated with an increase in lattice constant. Dilute nitride alloys therefore offer the possibility not only to extend the emission wavelength beyond that accessible using InGaAs alloys grown on GaAs substrates, but also to allow the compressive strain due to In incorporation to be compensated. This opens the possibility to achieve long-wavelength emission on GaAs at acceptable levels of epitaxial strain, and has led to the demonstration of 1.3-μm edge-emitting lasers and VCSELs [13–15], as well as 1.55-μm edge-emitting lasers [16,17].

From a fundamental perspective, the impurity-like nature of substitutional N atoms leads to a strong perturbation of the material band structure, primarily in the conduction band (CB). As a result, commonly employed approaches to modeling alloy electronic properties, such as the virtual crystal approximation, break down and a more detailed theoretical treatment is required to understand the strong impact of N incorporation on the electronic structure of GaN_xAs_{1-x} and related alloys. As such, there has been considerable theoretical interest in the causes and consequences of the unusual electronic properties of dilute nitride alloys, detailed understanding of which has largely been developed through atomistic calculations. We therefore begin this chapter by presenting, in Section 9.2, a review of atomistic theoretical calculations of the electronic structure, focusing primarily on the GaN_xAs_{1-x} and GaN_xP_{1-x} alloys. These calculations quantify the impact of N incorporation on the electronic structure, highlight the strong role played by

N-related alloy disorder, and describe the consequences of these effects for technologically relevant material properties in real N-containing alloys.

While atomistic approaches have provided significant insight into the electronic structure of N-containing alloys, they are too computationally expensive to apply directly on the length scales required to model the properties of optoelectronic devices. The significant interest in developing dilute nitride alloys for applications in semiconductor lasers and related devices has motivated the development of continuum models which are suited to calculating the properties of realistically sized N-containing device heterostructures. It has been demonstrated that the main features of the band structure of GaN_xAs_{1-x} and related alloys can be understood phenomenologically in terms of a band-anticrossing (BAC) interaction, between the unperturbed extended (Bloch) states of the (In)GaAs host matrix CB edge and localized states associated with substitutional N impurities. In Section 9.2 we review the quantitative understanding of the BAC model of the dilute nitride CB structure developed through detailed atomistic supercell calculations and outline the derivation of a 10-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian for $(In)GaN_xAs_{1-x}$, which is suited to calculate the electronic properties of device heterostructures.

Next, in Section 9.3, we review the theory of dilute nitride semiconductor lasers. The majority of the research effort dedicated to developing GaAs-based dilute nitride semiconductor lasers has focused on using quaternary InGaNAs QW structures, to target 1.3-μm emission. As such, it is to these structures that we primarily devote our attention. We describe how the 10-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian presented in Section 9.2 can be used to quantify the impact of N incorporation on the electronic and optical properties, and review important results outlining key trends in the predicted properties and performance of InGaNAs QW lasers. This discussion is supplemented by an outline of further investigations that have quantified the importance of various recombination pathways—including defect-related, radiative, and nonradiative (Auger) recombination—in prototypical InGaNAs devices, with comparisons between theory and experiment highlighting the progress that has been made in developing the understanding of this novel class of semiconductor lasers.

Having outlined in detail the theory and modeling of 1.3-μm InGaNAs QW lasers grown on GaAs substrates, in Section 9.4 we provide an overview of a range of alternative applications of dilute nitride alloys. These further attempts to exploit the impact of N incorporation on the band structure and facilitate the development of devices with enhanced capabilities and characteristics include the development and use of (1) InGaNAsSb structures on GaAs for 1.55-μm lasers and semiconductor optical amplifiers (SOAs), (2) InGaNAs structures grown on GaAs to develop SOAs, electro-absorption modulators (EAMs), and avalanche photodiodes (APDs) operating at 1.3 μm, and (3) GaNAsP alloys and QW structures for monolithic integration of III–V lasers on silicon (Si).

Finally, in Section 9.5 we turn our attention to the topic of carrier transport in dilute nitride alloys. Incorporation of N, which strongly perturbs the CB structure via the formation of highly localized impurity states, can be expected to have important implications for electron transport. This has been reflected in experimental measurements, which have revealed significant reductions in electron mobility even at ultra-dilute N compositions. By considering the BAC model of Section 9.2 in detail, we highlight the shortcomings of this simple picture of the CB structure from the perspective of calculating the transport properties of real materials. We provide an overview of descriptions of the band structure, based on the Anderson impurity model (AIM), which have been developed to overcome the shortcomings of the BAC model and facilitate practical calculations of density of states (DOS) and carrier transport. We then review theoretical models and calculations of electron transport in GaN_xAs_{1-x} and highlight the importance of the fine details of the N-perturbed CB structure in determining the properties of real materials. We focus in this review on calculations of the electron mobility and demonstrate the significant level of detail that must be incorporated in transport calculations in order to reliably reproduce the results of experimental measurements, and in doing so demonstrate rigorous validation of theoretical models of the CB structure in GaN_xAs_{1-x} alloys. We further highlight that N incorporation places intrinsic limits on the electron mobility in dilute nitride alloys, emphasize the dominant role in determining the transport properties played by strong resonant scattering of electrons by N-localized impurity states (associated not only with isolated

N impurities but also with N–N pairs and larger clusters of N atoms), and discuss the implications of N incorporation for the existence of a negative differential velocity regime in GaN_xAs_{1-x}.

The overall purpose of this chapter is to present a comprehensive overview of the theory and modeling of dilute nitride alloys, with a focus on applications in photonic devices. Our aim is that this chapter should provide the reader with a comprehensive introduction to the theory and simulation of dilute nitride alloys, in a manner that should enable them to engage meaningfully with the existing and emerging literature on this interesting and unusual class of semiconductor alloys.

9.2 Theory of the Electronic Structure of Dilute Nitride Alloys

In this section, we present a review of the theory and modeling of the electronic structure of dilute nitride alloys. Due to the fact that N acts as an isovalent impurity when incorporated into (In)GaAs(P) in dilute concentrations, a significant body of analysis has been undertaken to quantify the resulting strong perturbations to the band structure. A detailed exposition of the results of all of these studies is beyond the scope of this chapter. Here, we instead highlight key results from the literature, following a path that enables us to build toward a quantitative description of the properties of real, disordered dilute nitride alloys. The results and models reviewed in this section will underpin our discussions of the theory and simulation of dilute nitride materials and devices throughout the remainder of this chapter.

We begin in Section 9.2.1 with an overview of initial *ab initio* analysis of the generic properties of GaN_xAs_{1-x} alloys based on density functional theory (DFT), which served as motivation for the development of large supercell analyses of the alloy electronic structure undertaken using the empirical pseudopotential (EP) method. As described above, while atomistic analyses have provided detailed insight into the causes and consequences of the unusual properties of dilute nitride alloys, they are too computationally expensive to be applied on the length scales required to model realistic semiconductor devices. This has prompted the development of phenomenological descriptions of the dilute nitride CB structure based on the BAC model. In Section 9.2.2 we review the BAC model, beginning with the original empirical motivation and parametrization in response to experimental data before reviewing the results of atomistic tight-binding (TB) calculations, which allow the BAC and related models to be placed on a quantitative footing. We then describe the extension of the BAC approach, which has been accomplished by using TB supercell calculations to derive and parametrize an extended basis set 10-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian for dilute nitride alloys. This 10-band model will form the basis of our discussion of the theory of $(In)GaN_xAs_{1-x}$ semiconductor lasers in Section 9.3 and has been widely employed to calculate and analyze the properties of dilute nitride heterostructures and devices.

In Section 9.2.3 we outline the linear combination of the isolated nitrogen states (LCINS) approach to describing the band structure of N-containing alloys, and review that explicit consideration and construction of the N-related localized states allows for rigorous description of the electronic structure of real, disordered dilute nitride alloys. We demonstrate that the calculations undertaken using the LCINS method are in excellent quantitative agreement with the results of experimental measurements of a range of key material properties, in particular capturing the strong impact of N-related alloy disorder and non-monotonic variation with composition of important band properties.

9.2.1 First Principles and Empirical Pseudopotential Calculations

Before discussing the electronic structure of dilute nitride alloys in detail, it is useful first to state explicitly what is meant by describing a semiconductor alloy as “highly mismatched.” In conventional semiconductor alloys, such as $Al_xGa_{1-x}As$ or $In_xGa_{1-x}As_yP_{1-y}$, the material properties vary smoothly with composition. For example, the lattice and elastic constants, and deformation potentials, typically follow Végard’s law. That is, they can be determined straightforwardly by interpolating linearly between those of the binary compounds constituting the alloy. For example, the lattice constant a in $In_xGa_{1-x}As$ can be determined to a high degree of accuracy as $a(x) = (1 - x)\,a(\text{GaAs}) + x\,a(\text{InAs})$, where $a(\text{GaAs})$ and $a(\text{InAs})$ are the GaAs and InAs lattice constants, respectively. For some material parameters, typically the band gap and valence-band (VB) spin-orbit-splitting energy, experimental measurements reveal slight deviation from the linear variation described by Végard's law, which can be described in terms of a bowing parameter b—for example, the band gap E_g in $\text{In}_x\text{Ga}_{1-x}\text{As}$ can be determined as $E_g(x) = (1 - x)\,E_g(\text{GaAs}) + x\,E_g(\text{InAs}) - b\,x\,(1 - x)$, where b is independent of the alloy composition x and typically takes a small positive value <1 eV. The variation of the properties with alloy composition in conventional semiconductor alloy can then be described theoretically in a straightforward manner using, for example, the virtual crystal approximation in which the group-III atoms in $\text{In}_x\text{Ga}_{1-x}$ As can be treated as though they have the properties of an average "$\text{In}_x\text{Ga}_{1-x}$" atom.

In highly mismatched semiconductor alloys—such as the dilute nitride and bismide alloys we discuss here and in Chapter 10—the situation is markedly different. Highly mismatched semiconductor alloys are typically characterized by the fact that the constituent elements forming the alloy have significantly different sizes (covalent radii) and chemical properties (electronegativities). For example, in the dilute nitride (bismide) alloy $\text{GaN}_x\text{As}_{1-x}$ ($\text{GaBi}_x\text{As}_{1-x}$), the N (Bi) atoms are significantly smaller (larger) and more electronegative (electropositive) than the As atoms they replace. In these cases, the N and Bi atoms act as isovalent impurities when incorporated substitutionally in dilute concentrations, strongly perturbing the electronic structure. As a result, the material properties are found to vary rapidly, and in some cases non-monotonically, with alloy composition, leading to a breakdown of simple descriptions such as the virtual crystal approximation. This is typically characterized by the band gap E_g, which has two general properties not found in conventional semiconductor alloys: (1) a rapid variation of the magnitude of E_g with alloy composition, and (2) a large bowing parameter b which itself depends strongly upon the alloy composition. In a simple sense, the observation in experiment of one or more large, composition-dependent bowing parameters serves as a useful indication that a given semiconductor alloy is highly mismatched and is hence likely to possess a range of unusual material properties. The presence of such behavior in the dilute nitride alloys $\text{GaN}_x\text{As}_{1-x}$ and $\text{GaN}_x\text{P}_{1-x}$ has been known for some time: experimental measurements have revealed the presence due to N incorporation of (1) large, composition-dependent band gap bowing, (2) large and strongly composition-dependent CB effective masses, (3) significantly reduced pressure deformation potentials, and (4) a progression of sharp impurity lines in spectroscopic measurements, lying energtically within the host matrix (GaAs or GaP) band gap.

The presence of strongly composition-dependent band gap bowing in the dilute nitride alloy $\text{GaN}_x\text{As}_{1-x}$ was first theoretically analyzed in detail by Wei and Zunger [18]. Using a series of DFT supercell calculations, they identified two distinct regimes in the variation of the material properties with alloy composition, which serve as an indication of the general nature of a given semiconductor alloy: (1) an impurity-like composition regime, within which the material properties vary abruptly and strongly in response to small changes in composition, and (2) a band-like regime, within which the material properties vary smoothly and gradually in response to changes in composition. A comparative analysis of the band gap bowing in $\text{GaN}_x\text{As}_{1-x}$ and $\text{GaAs}_{1-x}\text{P}_x$ first revealed that the strong, composition-dependent bowing in $\text{GaN}_x\text{As}_{1-x}$ persists to large N compositions >25%, suggesting strong impurity-like behavior over a wide range of alloy compositions. The band gap bowing was calculated to be as large as $b = 15.69$ eV in a small supercell containing 12.5% N ($x = 0.125$), which was found to reduce to 7.6 eV by 25% N ($x = 0.25$). In constrast, such behavior was not calculated for $\text{GaAs}_{1-x}\text{P}_x$, with b calculated to vary from $b = 0.23$ eV at 12.5% P to 0.18 eV by 25% P. This allowed Wei and Zunger to rigorously distinguish between conventional and highly mismatched semiconductor alloys on the basis of the impurity- and band-like regimes described above. In a conventional alloy, substitutional incorporation of a dilute composition of atoms of a different species—for example, P atoms in GaAs to form $\text{GaAs}_{1-x}\text{P}_x$ at low x—leads to impurity-like behavior only over an extremely limited range of alloy compositions, with conventional band-like behavior emerging for compositions as low as $x \approx 1\%$. Distinct to this calculated behavior for $\text{GaAs}_{1-x}\text{P}_x$, Wei and Zunger demonstrated that highly mismatched alloys such as $\text{GaN}_x\text{As}_{1-x}$ can be characterized by the fact that the impurity-like regime extends to much higher alloy compositions.

Further analysis suggests that this persistence of the impurity-like regime is, as outlined above, associated with the large differences in orbital energies and covalent radius between an alloy's constituent elements. That this is the case has been generally known since the seminal work of, among others, Slater and Koster [19] and Hjalmarson et al. [20] on the electronic structure of deep-level impurities in semiconductors—subsequent interest in dilute nitrides and related materials has helped bring these concepts to the forefront of contemporary semiconductor research. As a general point, it is interesting to note that this unusual material behavior—i.e. strong qualitative changes in the material properties—are typically associated with one of several classes of phase transitions (for example, structural transitions between crystal phases or topological transitions in the electronic band structure), while in highly mismatched alloys such changes can be brought about by incorporating a dilute concentration of an impurity atom, suggesting significant potential to control, design, and optimize desirable material properties.

Following this general classification of conventional versus highly mismatched alloys, Wei and Zunger performed a systematic analysis of the large, composition-dependent band gap bowing in $\text{GaN}_x\text{As}_{1-x}$. Specifically, they analyzed the contributions to b associated with N incorporation arising from (1) volume deformation, (2) structural relaxation of the crystal, as well as (3) charge exchange between atoms. Each of these factors contributes to the observed strong reduction of the $\text{GaN}_x\text{As}_{1-x}$ band gap with increasing x, associated, respectively, with (1) compression (extension) of the Ga–As (Ga–N) bonds compared to their equilibrium lengths an a binary compound, (2) strain due to local relaxation of the crystal lattice about substitutional N impurities, and (3) a change in the spatial localization of the ground state charge density of VB (bonding) electrons in $\text{GaN}_x\text{As}_{1-x}$ as compared to that calculated for unrelaxed GaAs and GaN having the $\text{GaN}_x\text{As}_{1-x}$ lattice constant [18]. In this interpretation of the $\text{GaN}_x\text{As}_{1-x}$ alloy electronic structure, the local structural relaxation and charge exchange depend directly on the differences in atomic size and electronegativity between. As and N. Using a series of supercell calculations, and again comparing the calculated results for both N and P impurities in $\text{GaN}_x\text{As}_{1-x}$ and $\text{GaAs}_{1-x}\text{P}_x$, they demonstrated that while the volume deformation was comparable between the two alloys, the local structural relaxation and charge exchange were significantly more pronounced in $\text{GaN}_x\text{As}_{1-x}$. The prominence of these effects is then associated with the large ~20% lattice mismatch between GaAs and GaN, as well as the large differences ~4 and 2 eV between the s- and p-like valence orbitals in As and N. The calculated strong contribution of the relaxation of the local crystal structure about isolated N impurities, as well as the observed transfer of valence (conduction) charge from N to As (As to N), confirmed that the large, composition-dependent band gap bowing $\text{GaN}_x\text{As}_{1-x}$ is a direct result of the large chemical and size differences between As and N. This fundamental insight into the nature of the $\text{GaN}_x\text{As}_{1-x}$ electronic structure has, as we will see throughout the remainder of this section, motivated and informed a rich body of subsequent work on the theory of $\text{GaN}_x\text{As}_{1-x}$ and related alloys.

Following this initial study of the electronic structure of ordered $\text{GaN}_x\text{As}_{1-x}$ alloys, Bellaiche et al. [21] undertook a more detailed investigation of the electronic structure of $\text{GaN}_x\text{As}_{1-x}$ and the related dilute nitride alloy $\text{GaN}_x\text{P}_{1-x}$. Through a combination of DFT and EP calculations on supercells containing up to 512 atoms, they calculated, compared, and contrasted the electronic structure of $\text{GaN}_x\text{As}_{1-x}$ and $\text{GaN}_x\text{P}_{1-x}$ in terms of the impact of N incorporation on the CB structure. First, their calculations indicated the presence of impurity states associated with a substitutional N atom, which lie energetically within (and are hence resonant with) the CB in GaAs, but form a bound state in GaP (lying energetically within the GaP band gap). In both GaAs and GaP, the N-related impurity states were found to be highly localized about the N atomic site(s). Later calculations by Kent et al. [22], which we discuss further below, indicate that isolated N impurities are responsible for introducing strong N-mediated coupling of the extended host matrix CB states which creates a set of A_1-symmetric impurity states from a basis formed of a combination of GaAs CB states from the Γ-, L- and X-valleys, in addition to A_1-symmetric states associated directly with N. This leads to the N-derived localized states being formed of an admixture of a significant number of large wave vector host matrix states, with the A_1-like N-localized state associated with an isolated N impurity in $\text{GaN}_x\text{As}_{1-x}$ ($\text{GaN}_x\text{P}_{1-x}$) calculated to lie approximately 150 meV (30 meV) above (below) the GaAs (GaP) Γ-point CB states. As we describe in further detail throughout the remainder of this section, it is this N-mediated coupling which (1) is responsible for the strong perturbation to the GaAs(P) CB structure caused by N incorporation and (2) forms the basis of simplified theoretical models which have been used to understand and model the properties of dilute nitride alloys and heterostructures. A key success of this theoretical approach has been the explanation of the unusual pressure dependence of the GaN_xAs_{1-x} electronic structure: in the ultra-dilute limit, Kent et al. used a series of EP calculations to confirm that the localized states associated with isolated N impurities—which have intrinsically low-pressure dependence compared to the CB states of the host matrix semiconductor—can be pushed into the band gap under the application of hydrostatic pressures ∼2 GPa. This is consistent with a range of experimental measurements and confirms the role of the N-mediated coupling described above in determining the observed low-pressure deformation potentials in GaN_xAs_{1-x} alloys.

Given the presence of N-related localized states in GaN_xP_{1-x} lying energetically within the GaP band gap, incorporation of dilute concentrations of N is found to bring about an abrupt transition in the nature of the band structure: the unperturbed GaP band structure is an indirect gap, with the lowest energy CB lying at the X-point in the Brillouin zone. Incorporation of N gives rise to N-related localized bound states which lie below the GaP X-point CB minimum at Γ, therefore suggesting an abrupt change from an indirect to a direct band gap in GaN_xP_{1-x} even at ultra-dilute N compositions [23–25]. This unusual material behavior is particularly appealing from a technological perspective, as it brings about the opportunity to develop direct-gap III–V alloys which have lattice constants compatible with epitaxial growth on GaP or Si substrates (cf. Section 9.4.3).

Analysis of the bond lengths in $GaAs_{1-x}P_x$ and GaN_xAs_{1-x} alloys further revealed strong local relaxation of the crystal structure about the N atomic site(s), and comparison to calculations performed for isolated P impurities in GaAs again emphasized that these calculated material properties are distinct from those associated with conventional alloys. Following this, Bellaiche et al. investigated the impact of increasing the N composition x on the structural properties of GaN_xAs_{1-x} and GaN_xP_{1-x}, demonstrating the persistence and increase of strong structural relaxation with increasing x. Specifically, these calculations demonstrated the emergence of a binomial distribution of nearest-neighbor bond lengths corresponding, for example, to longer (shorter) Ga—As (Ga—N) bonds in disordered GaN_xAs_{1-x}. These calculations clearly demonstrated the emergence of percolation in GaN_xAs_{1-x} alloys for $x > 0.19 \equiv x_p$, associated with the energetically preferential formation of Ga—N—Ga—N—... nearest-neighbor "chains" throughout the crystal. EP calculations on disordered 512-atom supercells further confirmed the strong deviation of the GaN_xAs_{1-x} band gap from the variation predicted on the basis of Végard's law, throughout the entire composition range $0 \leq x \leq 1$—the electronic structure of GaN_xAs_{1-x} was found to divide roughly into three distinct regions: (1) $x < x_p$ in which N acts as an impurity primarily perturbing the CB structure, (2) $x_p \leq x \leq 0.4$ in which the material properties varying relatively smoothly with N composition, and (3) $x > 0.4$ in which As acts as an impurity primarily perturbing the VB structure. Overall, these calculations indicate highly unusual material behavior consisting of a range of complicated and correlated factors, including the presence of localized impurity states and their strong perturbation of the band structure—leading to strong band mixing and localization effects—as well as structural relaxation, charge transfer, and, ultimately, anion percolation.

Further theoretical analysis of the electronic structure of GaN_xAs_{1-x} and GaN_xP_{1-x} alloys was presented by Kent et al. in Refs. [26] and [22]. In these studies, Kent et al. applied the EP method to investigate the electronic structure of large, disordered GaN_xAs_{1-x} and GaN_xP_{1-x} supercells containing up to 14,000 atoms. Key results of these theoretical calculations include (1) the identification and quantitative analysis of a range of N-related localized states—associated not only with isolated N impurities, but also with N–N pairs (in which two N atoms share a common Ga nearest neighbor) and larger N clusters (in which multiple N atoms are second-nearest neighbors)—features associated with which had previously been observed in spectroscopic measurements, and (2) the identification of a localized-to-delocalized transition both in GaN_xAs_{1-x} and GaN_xP_{1-x}, related to strong, composition-dependent hybridization between the extended states of the host matrix (GaAs or GaP) CB states and highly localized N-related impurity states, and associated with a breakdown of the Bloch character of the alloy CB edge states.

Beginning with (1)—the existence and impact of N-related localized states—the detailed analysis undertaken by Kent et al. distinguished the impact of clustering of substitutional N impurities on the electronic structure of GaAs and GaP as again being distinct from that observed in conventional alloys such as $\mathrm{GaAs}_{1-x}\mathrm{P}_x$. First, while the formation of P–P pairs and larger clusters of P atoms tends not to form impurity states in $\mathrm{GaAs}_{1-x}\mathrm{P}_x$—thus leaving the electronic structure relatively unperturbed—as noted above, the formation of N–N pairs and larger clusters of N atoms in $\mathrm{GaN}_x\mathrm{As}_{1-x}$ and $\mathrm{GaN}_x\mathrm{P}_{1-x}$ produces a series of strongly localized impurity states. EP calculations indicate that these localized states typically become more strongly bound as the number of N atoms associated with a given localized perturbation increases, thereby giving rise to the progression of sharp impurity lines observed in spectroscopic measurements. Recalling that the localized states associated with isolated N impurities in $\mathrm{GaN}_x\mathrm{As}_{1-x}$ are resonant with the GaAs CB, the presence of N–N pairs and larger N clusters gives rise to more strongly bound localized states lying lower in energy, accounting for the observed progression of impurity lines observed within the $\mathrm{GaN}_x\mathrm{As}_{1-x}$ band gap at dilute N compositions.

The calculations of Kent et al. further suggested that these individual N cluster states are sufficiently localized that they tend not to significantly overlap or interact with one. Further analysis highlighted that the N-induced downward shift in energy of the alloy CB edge with increasing x is sufficiently large that it results in strong hybridization between CB edge states and N cluster states—as the former pass through the latter in energy—which, in conjunction with experimental measurements, suggests that N cluster states play a key role in bringing about the observed large CB edge effective masses and anomalously low pressure deformation potentials in $\mathrm{GaN}_x\mathrm{As}_{1-x}$ [22,40,44]. These results suggest that some of the more detailed features of the $\mathrm{GaN}_x\mathrm{As}_{1-x}$ (and $\mathrm{GaN}_x\mathrm{P}_{1-x}$) CB structure then depend on the precise N-related disorder present in a given material sample. It turns out, as we discuss in Sections 9.2.3 and 9.5.1, that this is indeed the case: a quantitative understanding of the dilute nitride CB structure, and in particular the CB edge effective mass and g factor, is possible only when the localized states associated with a full distribution of N clusters are treated explicitly. This again distinguishes dilute nitride alloys from conventional semiconductors, indicating that sophisticated, highly-specialized and computationally demanding theoretical approaches are required in order to understand observed trends in many important material parameters. A detailed review of the quantitative analysis of various N cluster states undertaken by Kent et al. is beyond the scope of this chapter. We refer the reader to Ref. [22] for the relevant details.

Turning our attention to (2)—the existence of a localized-to-delocalized transition—the EP calculations of Kent et al. further quantified the evolution of the alloy CB edge states in $\mathrm{GaN}_x\mathrm{As}_{1-x}$ and $\mathrm{GaN}_x\mathrm{P}_{1-x}$ as the N composition x is increased. At low N composition, the aforementioned strong N-induced hybridization between the extended states of the host matrix CB and N-related localized impurity states leads to alloy CB edge states having significant localized character (corresponding to a loss of the Bloch character that is largely preserved in conventional alloys). This behavior is characteristic of the extended impurity-like composition regime identified above, within which the material properties are strongly influenced by the nature of the localized states present in a given alloy. As the N composition is increased sufficiently to emerge from the impurity-like regime to the band-like regime, the presence and influence of these localized states is diminished, leading to a transition from alloy CB states which have strong localized (N-derived) character to extended alloy states which have significant Bloch character and more closely resemble those present in conventional alloys. We again refer the reader to Ref. [22] for further details and a quantitative discussion of this aspect of the dilute nitride electronic structure.

Having reviewed at a high level the impact of N incorporation on the structural and electronic properties of the prototypical dilute nitride alloys $\mathrm{GaN}_x\mathrm{As}_{1-x}$ and $\mathrm{GaN}_x\mathrm{P}_{1-x}$, we defer for now further discussion of the impact of N incorporation on the detailed features of the electronic structure. Instead, we turn our attention to simplified models of the dilute nitride CB structure, which have proved highly useful for (1) the interpretation of the general features of the band structure as revealed through experimental measurements, as well as for (2) the theory and modeling of dilute nitride semiconductor devices. The material we discuss in Section 9.2.2 will then informs our further discussion of the dilute nitride electronic structure in Sections 9.2.3 and 9.5, in which we respectively address the impact of N incorporation and N-related alloy disorder on the CB structure and transport properties of $\mathrm{GaN}_x\mathrm{As}_{1-x}$ and related alloys.

9.2.2 Band-Anticrossing: Empirical, Tight-Binding, and k·p Models

On the basis of the DFT and EP electronic structure calculations described above, it was established that the localized, impurity-like nature of the states associated with substitutional N atoms in (In)GaAsP alloys was responsible for the strong perturbation of the CB structure brought about by N incorporation. While the interpretation of these atomistic calculations was in general complex, in Ref. [27] Shan et al. demonstrated that the impact of N incorporation on the CB structure in (In)Ga$N_x$$As_{1-x}$ could in general be understood in a straightforward, phenomenological manner. Specifically, Shan et al. performed pressure-dependent photo-modulated reflectance (PR) measurements on a series of (In)Ga$N_x$$As_{1-x}$ alloys which revealed a number of important trends. First, experiments undertaken at ambient temperature and pressure suggested that N incorporation splits the Γ-point CB minimum into two subbands, clearly identifiable as distinct optical transitions in measured PR spectra. Second, performing PR measurements under applied hydrostatic pressure, Shan et al. observed a strongly non-linear variation of the alloy CB edge energy with applied hydrostatic pressure, a clear deviation from the linear variation which is characteristic of conventional semiconductor alloys. Furthermore, under high applied pressure saturation of the alloy CB edge energy was observed, clearly indicating anticrossing behavior.

On the basis of their experimental measurements, Shan et al. demonstrated that these unusual trends could be explained quantitatively by a simple two-band model of the (In)Ga$N_x$$As_{1-x}$ CB structure in which the extended states of the (In)GaAs host matrix undergo a strong anticrossing interaction with an energetically narrow band of localized N-related states that are resonant with the (In)GaAs CB. This is the well-known two-band BAC model, in that the CB structure is obtained by diagonalizing the Hamiltonian [27]

$$H(k) = \begin{pmatrix} E_N & V_{Nc} \\ V_{Nc}^* & E_{CB}^{(0)} + \frac{\hbar^2 k^2}{2m_e^*(0)} \end{pmatrix}, \tag{9.1}$$

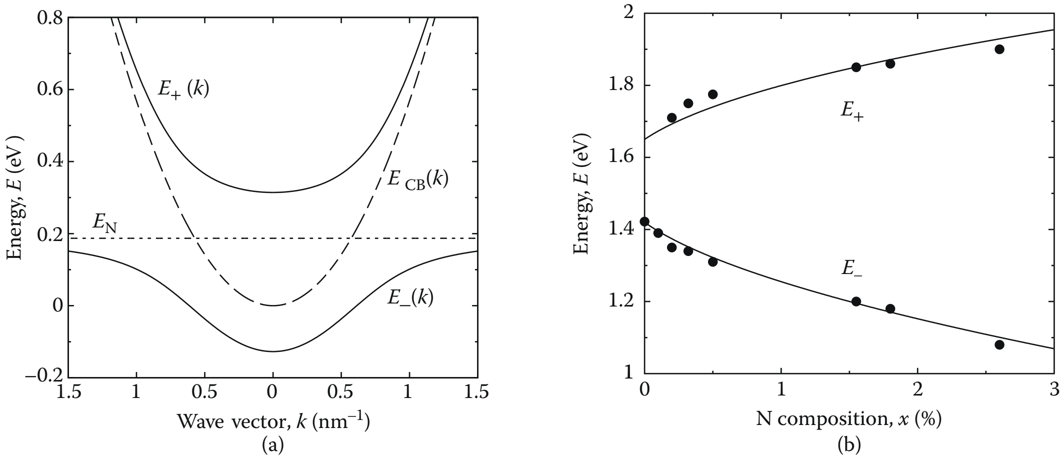
where $E_{CB}^{(0)}$ and $m_e^*(0)$ are the CB edge energy and electron effective mass of the (unperturbed) (In)GaAs host matrix, and E_N and V_{Nc} are the energy of the N-related localized states, and a matrix element describing a repulsive interaction between the (In)GaAs CB and N-related resonant states. Diagonalizing Equation 9.1 to obtain the (In)Ga$N_x$$As_{1-x}$ CB structure demonstrates the splitting of the (In)GaAs CB into two subbands $E_\pm(k)$ given by [27]

$$E_\pm(k) = \frac{E_{CB}(k) + E_N}{2} \pm \sqrt{\left(\frac{E_N - E_{CB}(k)}{2}\right)^2 + |V_{Nc}|^2}, \tag{9.2}$$

where $E_{CB}(k) = E_{CB}^{(0)} + \frac{\hbar^2 k^2}{2m_e^*(0)}$ describes the dispersion of the unperturbed (In)GaAs CB.

Comparing the CB structure calculated using Equation 9.2 with the results of experimental measurements, Shan et al. demonstrated that this simple model is capable of describing the observed variation of (1) the energy of the optical transitions associated with $E_-(0)$ and $E_+(0)$ with N composition x, and (2) the alloy CB edge energy $E_-(0)$ with applied hydrostatic pressure. In order to obtain a quantitative description, Shan et al. determined on the basis of experimental data the matrix element V_{Nc} describing the N-induced coupling/hybridization of the (In)GaAs CB and N-related localized states depends strongly on the N composition, varying from 0.12 eV at $x = 0.9\%$ to 0.40 eV at $x = 2.3\%$. This composition dependence then describes that the observed strong, N-composition-dependent bowing of (In)Ga$N_x$$As_{1-x}$ band gap is associated with an N composition-dependent BAC interaction between the (In)GaAs CB and N-localized resonant states—that is, the strongly N composition-dependent reduction in the alloy CB edge energy $E_-(0)$ is the origin of the observed rapid reduction of the (In)Ga$N_x$$As_{1-x}$ band gap with increasing x (cf. Equation (9.2)).

The (In)Ga$N_x$$As_{1-x}$ band structure calculated using the BAC model is depicted schematically in Figure 9.1a [28], which demonstrates that the interaction of the (In)GaAs CB (dashed line) with the N-related resonant states (dotted line) splits the (In)Ga$N_x$$As_{1-x}$ CB into the two subbands $E_\pm(k)$ described



(a)

(b)

FIGURE 9.1 (a) Band structure of $\mathrm{GaN_{0.01}As_{0.99}}$ ($x = 1\%$) calculated using the two-band BAC model of Equation 9.1. The BAC model describes that the GaAs host matrix CB (dashed line) undergoes an N composition-dependent anticrossing interaction with a set of localized N-related resonant states at energy E_N (dotted line), splitting the CB into two bands $E_\pm(k)$ (solid lines). The zero of energy is taken at the GaAs CB edge. (b) Calculated (solid lines) and measured (closed circles) room temperature energies of the optical transitions from the VB edge to the $E_\pm$ band edge states (at $k = 0$), as a function of N composition x for a series of $\mathrm{GaN_xAs_{1-x}}$ samples. (Adapted from P.J. Klar et al., *Appl. Phys. Lett.*, 76, 3439, 2000.) The excellent agreement between theory and experiment confirms that the BAC model describes the observed strong reduction and composition-dependent bowing of the band gap in terms of the BAC-induced reduction of the alloy CB edge energy $E_-(0)$.

by Equation 9.2 (solid lines). We note that the BAC interaction (1) pushes the alloy CB edge downward in energy, accounting for the strong N-induced reduction of the band gap, and (2) pushes down the dispersion of the lower band at fixed k, leading to a strong N-induced increase in the effective mass (cf. Section 9.2.3). An important point to note regarding the BAC band structure is that the energy $E_-(k)$ of the alloy CB edge approaches E_N asymptotically from below as $k \to \infty$ (cf. Equation 9.2), indicating the presence of an unphysical singularity in the DOS at $E = E_\mathrm{N}$. Despite its success in describing the evolution of the main features of the (In)$\mathrm{GaN_xAs_{1-x}}$ band structure, this inability to describe the DOS is an intrinsic limitation of the BAC model, which impairs its ability to provide a comprehensive description of the impact of N incorporation on the CB structure. While this issue is of particular importance for analysis of the transport properties of dilute nitride alloys, it is not of significant consequence for analysis of the optoelectronic properties (where we are generally concerned with carriers lying only in the immediate vicinity of the CB and VB edges in energy). As such, we defer discussion of the DOS until our review of the theory of the impact of N incorporation on carrier transport in Section 9.5.

Turning our attention to the parameterization of the BAC model, we note that the predictive capability of the model is limited due to the approach followed by Shan et al. in Ref. [27]. That this is the case can be understood by noting that the calculated transition energies $E_-(0)$ and $E_+(0)$ associated with the optical features observed in the experiment depend on two independent parameters: (1) the assumed energy E_N of the N-related resonant states, and (2) the magnitude V_Nc of the BAC interaction matrix element. In their initial study, Shan et al. determined E_N in (In)GaAs via an extrapolation of the available experimental data for the energy of the N-related states in GaAsP alloys, which for sufficiently large P compositions are bound states lying energetically within the host matrix band gap. For each composition considered, V_Nc was then determined by fitting to the measured transition energy associated with the $E_-(0)$ optical transition (which, by Equation 9.1, simultaneously determines $E_+(0)$). From a theoretical perspective, this approach has two significant shortcomings. First, the fitted value of V_Nc at a given N composition x depends on the energy E_N assumed for the N-related resonant states, leading to parametric ambiguity in the absence of detailed information regarding E_N and its dependence on composition. Second, the strong composition dependence of V_{Nc} means that—in the absence of experimental data against which to fit V_{Nc}—it is not possible to use Equation 9.1 to predict N composition-dependent trends in the (In)GaN$_x$As$_{1-x}$ CB structure.

In order to overcome these issues, O'Reilly et al. developed an atomistic TB model to study the electronic structure of dilute nitride alloys [29]. This approach has proved useful in elucidating the impact of N incorporation on the electronic structure of GaN$_x$As$_{1-x}$ and related alloys, including quantitative predictions of several key material properties and their dependence on N composition [30], as well as providing detailed information regarding the nature of the localized states associated with substitutional N impurities [31]. A full description of the TB model is beyond the scope of this chapter—we refer the reader to Ref. [29] for full details. In this section, we focus on one aspect of the TB analysis of the GaN$_x$As$_{1-x}$ electronic structure which is relevant to the BAC model. Using the TB method, O'Reilly et al. demonstrated that it is possible to explicitly construct the N-related localized resonant states associated with substitutional N impurities in supercell calculations, thereby allowing the N-related parameters of the BAC model to be computed directly and unambiguously for a given host matrix, and hence delivering predictive power to the BAC model. In the BAC model, the (In)GaN$_x$As$_{1-x}$ alloy CB edge state $|\psi_{CB}^{(1)}\rangle$ is hybridized—i.e. it is formed of a linear combination of the unperturbed (In)GaAs CB edge state $|\psi_{CB}^{(0)}\rangle$ and the localized resonant state $|\psi_N\rangle$ associated with an isolated, substitutional N impurity. As a result, $|\psi_N\rangle$ can be constructed directly via TB calculations carried out on N-free and N-containing supercells as [29,31]

$$|\psi_N\rangle = \frac{|\psi_{CB}^{(1)}\rangle - \langle\psi_{CB}^{(0)}|\psi_{CB}^{(1)}\rangle\,|\psi_{CB}^{(0)}\rangle}{\sqrt{1 - |\langle\psi_{CB}^{(0)}|\psi_{CB}^{(1)}\rangle|^2}}, \tag{9.3}$$

where $|\psi_{CB}^{(0)}\rangle$ and $|\psi_{CB}^{(1)}\rangle$ are, respectively, the CB edge eigenstates of the N-free and N-containing supercells, and $f_{\Gamma,CB} = |\langle\psi_{CB}^{(0)}|\psi_{CB}^{(1)}\rangle|^2$ is the fractional Γ character—that is, the squared modulus of the overlap between the CB edge state $|\psi_{CB}^{(1)}\rangle$ in an N-containing alloy supercell and the unperturbed host matrix CB edge state $|\psi_{CB}^{(0)}\rangle$ [29,31].

Using Equation 9.3 to construct $|\psi_N\rangle$ for a large supercell (containing ~2000 atoms in the presence of a single, isolated substitutional N impurity) then allows E_N and V_{Nc} to be computed directly as $E_N = \langle\psi_N|\widehat{H}|\psi_N\rangle$ and $V_{Nc} = \langle\psi_{CB}^{(0)}|\widehat{H}|\psi_N\rangle$, where $\widehat{H}$ is the full TB Hamiltonian for the N-containing supercell. By applying this approach to a series of ordered GaN$_x$As$_{1-x}$ supercells, O'Reilly et al. directly determined the N composition dependence of the BAC interaction matrix element as $V_{Nc} = \beta\sqrt{x}$, where the coupling parameter β depends on (1) the type of N-related localized state—i.e. whether it is associated with an isolated N impurity, an N–N pair, or larger cluster of N atoms—and (2) the host matrix semiconductor into which N is incorporated.

The calculated values of E_N and β for isolated, substitutional N impurities are summarized in Table 9.1 for a range of III–V host matrix semiconductors [12]. The energies of these N–related localized states are given relative to the host matrix VB edge (as E_N) and relative to the host matrix CB edge (as ΔE_N). Positive (negative) values of E_N indicate that a substitutional N impurity forms resonant (bound) localized states, which lie energetically within the host matrix CB (band gap). Also given in Table 9.1 are the calculated energies E_{NN} of the localized states associated with N–N pairs—in which two N atoms share a group-III (cation) nearest neighbor—and are again given relative to the host matrix VB edge (as E_{NN}) and relative to the host matrix CB edge (as ΔE_{NN}). We note that the values of E_N calculated using the TB method are in quantitative agreement with experiment. In particular, the TB calculations correctly describe that an isolated N impurity forms a resonant (bound) state in GaAs (GaP), lying energetically within the CB (band gap) [23]. Furthermore, we note that $\Delta E_{NN} < \Delta E_N$ for all host matrices considered, indicating that the localized states associated with N–N pairs are more strongly bound than those associated with isolated N impurities. We note also the impact of temperature on the nature of the N-related localized states: given that the N-related (impurity-like) are relatively temperature insensitive, the reduction in the host matrix CB edge energy brought about by an increase in temperature means that, for certain host matrices, an

TABLE 9.1 Energies of the Localized States Associated with Isolated Substitutional N Impurities (E_N) and N–N Pairs(E_{nn}), as well as the Coupling Strength β of the BAC Interaction between $|\psi_N\rangle$ and the Unperturbed Host Matrix CB Edge Calculated at Low Temperature using the TB method for a Range of III–V Semiconductor Host Matrices

Host Matrix	E_g (eV)	E_N (eV)	ΔE_N (meV)	β (eV)	E_{NN} (eV)	ΔE_{NN} (meV)
GaP	2.886 (2.777)	2.306	−580 (−471)	1.737	2.180	−706 (−597)
GaAs	1.519 (1.422)	1.706	187 (284)	2.000	1.486	−33 (64)
GaSb	0.812 (0.727)	0.770	−42 (43)	2.400	0.540	−272 (−187)
InP	1.424 (1.353)	1.903	479 (550)	0.901	1.770	346 (417)
InAs	0.417 (0.354)	1.386	969 (1032)	1.212	1.201	784 (847)
InSb	0.235 (0.174)	0.720	485 (546)	1.970	0.520	285 (346)

Values in parentheses are those computed at room temperature. E_N and E_{NN} are given relative to the unperturbed host matrix VB edge. The band gap E_g of the unperturbed host matrix is provided for reference. ΔE_N (ΔE_{NN}) denotes the difference in energy between E_N (E_{NN}) and the unperturbed host matrix CB edge. As such, positive (negative) values of ΔE_N or ΔE_{NN} indicate that an isolated N impurity or N–N pair forms a resonant (bound) localized state lying energetically within the host matrix CB (band gap).

isolated N impurity forms a bound state at low temperature but a resonant state at room temperature. For example, the calculated negative (positive) value of ΔE_N indicates that an isolated N impurity forms a bound (resonant) state in GaSb at low (room) temperature. A similar trend is observed in GaAs, in which the N–N pair states lie close in energy to the unperturbed host matrix CB edge.

A final important point to note is that the N-related localized states in In-containing compounds lie significantly higher in energy than the host matrix CB edge. As we see when we discuss the impact of alloy disorder on the electronic structure in Section 9.2.3, the presence of N-related states lying close in energy to the CB edge leads to extremely strong perturbation of the electronic structure. As such, it is generally found that the BAC model—which treats only the impact of isolated N impurities on the CB structure—is generally more accurate when applied to In-containing dilute nitride alloys due to the large energy separation between the host matrix CB edge and the localized states associated with N pairs and clusters [12]. It is for this reason that the BAC-based 10-band **k·p** Hamiltonian we describe below is found to be sufficiently detailed to accurately describe the electronic and optical properties of InGaNAs alloys and heterostructures (as we see when we discuss the theory of 1.3-μm dilute nitride semiconductor lasers in Section 9.3).

The success of the BAC model—parameterized explicitly via TB supercell calculations—in describing the main features of the (In)GaN$_x$As$_{1-x}$ CB structure has been explicitly confirmed by comparison of the calculated alloy band edge energies $E_\pm(0)$ to PR measurements of the associated optical transition energies performed on a range of samples having different N compositions [28]. Using the values of E_N and β given for GaN$_x$As$_{1-x}$ in Table 9.1, it is possible to calculate the variation with x of $E_\pm(0)$, which can then be compared to experiment. This theory–experiment comparison is shown in Figure 9.1b, which compares the measured transition energies $E_\pm(0)$ (closed circles) to those calculated using Equation 9.2 with $k = 0$ (solid lines). We note the excellent agreement between theory and experiment, confirming that the BAC model is capable of quantitatively describing the evolution of the main features of the GaN$_x$As$_{1-x}$ band structure with N composition x.

For the purposes of modeling the electronic and optical properties of dilute nitride semiconductor lasers and related photonic devices, it is desirable to develop appropriate **k·p** models which treat the states lying close in energy to the CB and VB edge on an equal footing. For conventional III–V semiconductor alloys an eight-band model—taking as a basis the spin degenerate zone-center Bloch states associated with the lowest energy CB, as well as those of the heavy-hole (HH), light-hole (LH), and spin-split-off-hole (SO) VBs—is typically found to be sufficiently detailed to accurately describe the electronic and optical properties of quantum-confined heterostructures. However, as we have seen, experimental measurements and atomistic calculations suggest that N incorporation introduces localized impurity states lying close in energy to the host matrix CB edge which, due to their strong composition-dependent interactions with the host matrix CB edge, must be treated explicitly in order to understand the impact of N on the band structure. This suggests that the conventional eight-band **k·p** model is insufficient to describe the dilute nitride band structure, and hence that an extended **k·p** model must be developed in which the N-related localized resonant states are included explicitly in the basis set. Given the success of the BAC model, the natural choice of basis set is then to augment the zone-center Bloch states of the eight-band model for the host matrix with a pair of spin degenerate N-related localized states (having energy E_{N}). This leads us to a 10-band **k·p** Hamiltonian for (In)GaN$_x$As$_{1-x}$ and related alloys.

The TB model introduced by O'Reilly et al. provides a natural framework within which to derive and parameterize the 10-band model. In the interest of brevity, the manner by which this parameterization proceeds—full details of which can be found in Refs. [29] and [32]—is described in more detail in Chapter 10, when we consider a similar extended basis set **k·p** Hamiltonian for the band structure of the dilute bismide alloy (In)GaBi$_x$As$_{1-x}$. Essentially, by using the zone-center eigenstates calculated for the host matrix in conjunction with the N-related localized states constructed following Equation 9.3 to construct the 10-band basis set, one can directly evaluate the matrix elements of the full TB Hamiltonian for the N-containing supercell and hence fully parametrize the 10-band model without recourse to post hoc fitting via experimental data. The 10-band **k·p** Hamiltonian obtained in this manner is similar to the eight-band model for the host matrix semiconductor, but is distinguished by the addition of (1) virtual crystal (conventional alloy) contributions describing the linear variation of the band edge energies with N composition x, and (2) N composition-dependent BAC interactions $V_{\text{Nc}} = \beta\sqrt{x}$ between the host matrix CB edge and the N-related localized states. TB calculations reveal that the N-related states $|\psi_{\text{N}}\rangle$ do not couple to either of the HH, LH, or SO VB edge states [29,32]. The resulting 10-band **k·p** Hamiltonian for dilute nitride alloys is then of the form

$$H = \begin{pmatrix} E_{\text{N}} & V_{\text{Nc}} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ & E_{\text{CB}} & -\sqrt{3}T & \sqrt{2}U & -U & 0 & 0 & 0 & -T^{*} & -\sqrt{2}T^{*} \\ & & E_{\text{HH}} & \sqrt{2}S & -S & 0 & 0 & 0 & -R & -\sqrt{2}R \\ & & & E_{\text{LH}} & Q & 0 & T^{*} & R & 0 & \sqrt{3}S \\ & & & & E_{\text{SO}} & 0 & \sqrt{2}T^{*} & \sqrt{2}R & -\sqrt{3}S & 0 \\ & & & & & E_{\text{N}} & V_{\text{Nc}} & 0 & 0 & 0 \\ & & & & & & E_{\text{CB}} & -\sqrt{3}T^{*} & \sqrt{2}U & -U \\ & & & & & & & E_{\text{HH}} & \sqrt{2}S^{*} & -S^{*} \\ & & & & & & & & E_{\text{LH}} & Q \\ & & & & & & & & & E_{\text{SO}} \end{pmatrix}, \tag{9.4}$$

where we have shown only the upper triangle since H is a Hermitian matrix—the diagonals below can be obtained by Hermitian conjugation. The definition of the basis $|u_1\rangle, \ldots, |u_{10}\rangle$ of zone-center crystal eigenstates in which Equation 9.4 is written can be found in Ref. [33]. Taking the zero of energy at the unperturbed host matrix VB edge, the matrix elements of Equation 9.4 are defined at N composition x , and in the presence of pseudomorphic strain corresponding to growth along the (001) direction, as [33]

$$E_{\text{CB}} = E_g + (\kappa - \alpha)\,x + \frac{\hbar^2}{2m_0} s_c \left(k_{\parallel}^2 + k_z^2\right) + \delta E_{\text{CB}}^{\text{hy}}\,, \tag{9.5}$$

$$E_{\text{HH}} = \kappa\,x - \frac{\hbar^2}{2m_0}\left(\left(\gamma_1 + \gamma_2\right)k_{\parallel}^2 + \left(\gamma_1 - 2\gamma_2\right)k_z^2\right) + \delta E_{\text{VB}}^{\text{hy}} - \delta E_{\text{VB}}^{\text{ax}}\,, \tag{9.6}$$

$$E_{\text{LH}} = \kappa\,x - \frac{\hbar^2}{2m_0}\left(\left(\gamma_1 - \gamma_2\right)k_{\parallel}^2 + \left(\gamma_1 + 2\gamma_2\right)k_z^2\right) + \delta E_{\text{VB}}^{\text{hy}} + \delta E_{\text{VB}}^{\text{ax}}\,, \tag{9.7}$$

$$E_{\text{SO}} = \kappa\,x - \Delta_{\text{SO}} - \frac{\hbar^2}{2m_0}\gamma_1 \left(k_{\parallel}^2 + k_z^2\right) + \delta E_{\text{VB}}^{\text{hy}}\,, \tag{9.8}$$

$$Q = -\frac{1}{\sqrt{2}}\frac{\hbar^2}{m_0}\gamma_2 k_\parallel^2 + \sqrt{2}\frac{\hbar^2}{m_0}\gamma_2 k_z^2 - \sqrt{2}\,\delta E_{\mathrm{VB}}^{\mathrm{ax}}\,, \tag{9.9}$$

$$R = \frac{\sqrt{3}}{2}\frac{\hbar^2}{m_0}\left(\gamma_{\mathrm{av}} k_-^2 - \mu k_+^2\right), \tag{9.10}$$

$$S = \sqrt{\frac{3}{2}}\frac{\hbar^2}{m_0}\gamma_3 k_z k_-\,, \tag{9.11}$$

$$T_\pm = \frac{P}{\sqrt{6}}k_\pm\,, \tag{9.12}$$

$$U = \frac{P}{\sqrt{3}}k_z\,, \tag{9.13}$$

where $k_\parallel = \sqrt{k_x^2 + k_y^2}$, $k_\pm = k_x \pm ik_y$, m_0 is the free electron mass, E_g and Δ_{SO} are the host matrix band gap and VB spin-orbit-splitting energy, P is the Kane interband momentum matrix element, and $\gamma_{1,2,3}$ are the modified Luttinger parameters. γ_{av} and μ are defined in terms of the modified Luttinger parameters as $\gamma_{\mathrm{av}} = \frac{1}{2}(\gamma_2 + \gamma_3)$ and $\mu = \frac{1}{2}(\gamma_3 - \gamma_2)$, and the definition of the parameter s_c can be found in Ref. [33]. For brevity, the definitions of the energy shifts $\delta E_{\mathrm{CB}}^{\mathrm{hy}}$, $\delta E_{\mathrm{VB}}^{\mathrm{hy}}$, and $\delta E_{\mathrm{VB}}^{\mathrm{ax}}$ to the CB and VB edges, resulting from hydrostatic and axial strain, are provided in Chapter 10. The composition-dependent N-localized state energy E_{N} of Equation 9.4 is given relative to the host matrix VB edge as $E_{\mathrm{N}} = E_g + \Delta E_{\mathrm{N}} + (\kappa - \gamma)\,x$ and, as above, the BAC interaction matrix element is $V_{\mathrm{Nc}} = \beta\sqrt{x}$.

A detailed discussion of the N-related parameters α, γ, and κ, which describe the aforementioned virtual crystal contributions to the N-induced shifts of the band edge energies, can be found in Ref. [34]. For the calculations based on the 10-band **k·p** model to be presented in Section 9.3, we have used the parameters provided by Tomić et al. in Ref. [33]. The band dispersion calculated in the vicinity of the CB and VB edges in energy using this 10-band model is in excellent agreement with full TB supercell calculations [32,35]. The 10-band model then describes the strong perturbation of the CB structure brought about by N incorporation (due to strong, composition-dependent BAC), while simultaneously describing that N has relatively little impact on the VB structure (described purely in terms of minor shifts in the band edge energies associated with strain and conventional alloy effects).

Having described that the main features of the CB structure of (In)GaNAs(P) alloys can be straightforwardly described by a strongly composition-dependent BAC interaction between the extended states of the host matrix CB edge and highly localized states associated with substitutional N impurities, we now turn our attention to more detailed aspects of the electronic structure of GaN$_x$As$_{1-x}$ alloys, focusing in particular on the impact of N-related alloy disorder. This highlights conditions under which the BAC model can be expected to break down and demonstrates that a quantitative description of the dilute nitride CB structure depends upon a direct treatment of the full range of N-related localized states appearing in realistic, disordered alloys.

9.2.3 Tight-Binding and Alloy Disorder: Linear Combination of Isolated Nitrogen States

As we have described so far in this section, initial investigations of the electronic structure of dilute nitride alloys were primarily based on two complementary approaches: (1) detailed electronic structure calculations using DFT, EP and TB methods, and (2) the empirically motivated BAC model. Detailed comparisons between theory and experiment have demonstrated that the BAC model is generally capable of describing the dependence of the band edge energies in bulk and QW (In)GaN$_x$As$_{1-x}$ samples on N composition and applied hydrostatic pressure. However, despite the successes of these approaches, there are some important experimental data which were not satisfactorily described by either the atomistic supercell or the BAC models described above. First, the BAC model fails to describe the anomalously large values m_e^* and g_e^* of the CB edge electron effective mass and g factor measured in some $\text{GaN}_x\text{As}_{1-x}$ samples even for ultra-dilute N compositions $x \approx 0.2\%$ [36–38]. Second, a feature associated with the higher-energy N-derived E_+ state appearing in the BAC model is only observable in spectroscopic measurements on $\text{GaN}_x\text{As}_{1-x}$ alloys over a limited range of compositions, appearing for $x > 0.2\%$ and growing in sharpness and prominence with increasing x, before undergoing rapid energy broadening and vanishing for $x > 3\%$ [28,39]. Description of these detailed features of the $\text{GaN}_x\text{As}_{1-x}$ electronic structure and, in particular, description of the observed strong, nonmonotonic variation of m_e^* and g_e^* with N composition then serves as a stringent test of theoretical models of the $\text{GaN}_x\text{As}_{1-x}$ CB structure.

These previously unexplained aspects of the electronic properties of $\text{GaN}_x\text{As}_{1-x}$ alloys motivated further theoretical investigations of the CB structure beyond the BAC model. Of critical importance is to note the fact that the BAC model assumes that all N-related states have the same energy E_N, the energy of the localized state associated with an isolated N impurity (giving rise to the aforementioned singularity in the BAC DOS at the energy of the N-localized states). This is not the case in a real, disordered $\text{GaN}_x\text{As}_{1-x}$ alloy: as described for isolated N versus N–N pairs in Section 9.2.2, the localized states associated with different types of N impurities have different energies. This is confirmed by experimental measurements that indicate the presence of a series of N-related levels in the $\text{GaN}_x\text{As}_{1-x}$ CB, lying close in energy to the unperturbed GaAs CB edge, suggesting the presence of a broad collection of N-related localized states in a disordered alloy. Such states have also been observed in theoretical calculations based on the EP and TB methods. Two important points then become clear. First, given that the BAC model is capable of describing the dependence of the band edge energies on N composition and applied hydrostatic pressure to a reasonable degree of accuracy, it can be concluded that N-related localized states associated with N–N pairs and larger clusters of N atoms play only a secondary role in determining the variation of the band edge energies (and hence band gap) with x, but play an important role in determining the electron effective mass (and hence the transport properties). Second, any quantitative description of the details of $\text{GaN}_x\text{As}_{1-x}$ electronic structure must explicitly take account of the full distribution of N-related localized states, including those associated not only with isolated N impurities, but also with N–N pairs and larger clusters of N atoms.

In Ref. [40], O'Reilly et al. developed an approach that explicitly treats the N-related localized states in realistic, disordered $\text{GaN}_x\text{As}_{1-x}$ alloys. This approach proceeds on the basis that the N-related states associated with different N complexes are highly localized about the N atomic sites, suggesting that the N-related states in a disordered alloy can be represented as a LCINS—that is, as a linear combination of the localized states associated with each individual N complex in a given alloy (supercell). Each of these localized states can be constructed explicitly using the TB method (cf. Sec. 9.2.2) allowing their energies and interactions with the unperturbed GaAs CB edge states to be quantified. Taking as a basis these N-related localized states in conjunction with the unperturbed GaAs CB edge state $|\psi_\text{CB}^{(0)}\rangle$, the LCINS approach then allows a suitable Hamiltonian to be constructed which explicitly treats the interaction between the unperturbed GaAs CB edge and the full distribution of N-related localized states present in a realistic, disordered alloy. More specifically, given the fact that the localized state associated with each N complex is highly localized about the N site(s), one can associate a separate localized state with each isolated N impurity, N–N pair, or larger N cluster. To verify that this is the case, O'Reilly et al. examined the N-related states in a series of ordered, $2M$-atom $\text{Ga}_M\text{N}_n\text{As}_{M-n}$ supercells ($x = \frac{n}{M}$), in which a total of n substitutional N impurities were placed on a regular grid of sites on the anion sublattice. The N-related states in such a structure were constructed as an LCINS via [40]

$$|\psi_\text{N}\rangle = \frac{1}{\sqrt{n}} \sum_{m=1}^{n} |\psi_\text{N}^{(m)}\rangle\,, \tag{9.14}$$

where $|\psi_\text{N}^{(m)}\rangle$ is the N-related localized state associated with an *isolated* substitutional N impurity located at the mth N site. For all bar one of the supercells considered, the calculated overlap of the LCINS state described by Equation 9.14 with the N-related state computed directly for the ordered supercell via

Equation 9.3 was $\geq$ 94%. A smaller overlap ($\approx$ 70%) was obtained only for an 8-atom $Ga_4N_1As_3$ supercell having a large N composition $x = 25\%$, and in which each N atom had an N second-nearest neighbor (manifested as a chain of Ga–N nearest-neighbor pairs due to the Born-von Karman supercell boundary conditions). Furthermore, the energy $E_N = \langle \psi_N | \widehat{H} | \psi_N \rangle$ computed for each supercell using the N-related states described by Equations 9.3 and 9.14—in conjunction with the full TB supercell Hamiltonian $\widehat{H}$—was found to be in excellent agreement for all supercells considered. These comparisons confirmed that the N-related states in a GaN_xAs_{1-x} alloy supercell can in general be accurately described via an LCINS of the form of Equation 9.14.

In order to investigate the accuracy of the LCINS approach for describing the electronic structure of disordered alloy supercells, O'Reilly et al. next considered a series of disordered, 1000-atom $Ga_{500}N_{n+2p}As_{500-n-2p}$ supercells ($x = \frac{n+2p}{M}$) containing a total of $n + 2p$ substitutional N impurities in the form of n isolated N atoms and p N–N pairs. In general, the GaN_xAs_{1-x} alloy CB edge state is expected to be formed of an admixture of the unperturbed GaAs CB edge state $|\psi_{CB}^{(0)}\rangle$ and the LCINS with which it interacts. On this basis they defined a $(1 + n + 2p) \times (1 + n + 2p)$ Hamiltonian matrix describing the interaction between $|\psi_{CB}^{(0)}\rangle$ and (1) the n localized states associated with isolated N impurities, and (2) the $2p$ localized states associated with N–N pairs. This Hamiltonian matrix, $H_{ij} = \langle \psi_i | \widehat{H} | \psi_j \rangle$ where $i, j = 1, \ldots, n + 2p$, is constructed explicitly using the full TB supercell Hamiltonian. The energies of the LCINS eigenstates are then calculated as the eigenvalues of the generalized eigenvalue equation $H\, u_i = E_i\, S\, u_i$, where $S_{ij} = \langle \psi_i | \psi_j \rangle$ is the overlap matrix between the $1 + n + 2p$ basis states (which, in general, has nonzero off-diagonal elements due to the partial overlap between N-related localized states centered on different lattice sites). The LCINS eigenstates are then computed as [40]

$$|\phi_i\rangle = u_{CB,i}\, |\psi_{CB}^{(0)}\rangle + \sum_{m=1}^{n} u_{N,i}^{(m)} |\psi_N^{(m)}\rangle + \sum_{m=1}^{n} \left(u_{NN+,ia}^{(m)}\, |\psi_{NN+}^{(m)}\rangle + u_{NN-,i}^{(m)}\, |\psi_{NN-}^{(m)}\rangle \right), \tag{9.15}$$

where $|\psi_{NN+}^{(m)}\rangle$ and $|\psi_{NN-}^{(m)}\rangle$ are the localized states associated with isolated N–N pairs which are, respectively, symmetric and antisymmetric about the Ga (cation) nearest neighbor shared by the N atoms forming the pair.

By applying this model to the aforementioned series of disordered $Ga_{500}N_{n+2p}As_{500-n-2p}$ supercells and computing (1) the eigenstate energies E_i, and (2) the fractional Γ character associated with each of the LCINS eigenstates $|\phi_i\rangle$, it was confirmed that the LCINS approach is in excellent agreement with the results of full TB supercell calculations. These detailed comparisons for both ordered and disordered alloys emphatically demonstrate that the GaN_xAs_{1-x} electronic structure can be described straightforwardly and accurately in terms of an extended BAC-like approach in which the interactions between the unperturbed GaAs CB edge state $|\psi_{CB}^{(0)}\rangle$ and the full distribution of localized states associated with different N complexes are treated explicitly via an LCINS.

As described above, the LCINS approach appears to be supercell-dependent—that is, it appears to depend strongly on the assumed distribution of isolated N impurities, N–N pairs, and largers clusters of N atoms in a given supercell—which draws into question its general predictive capability. However, O'Reilly et al. demonstrated that this apparent limitation can be overcome by treating the N-related localized states in disordered alloy supercells containing large numbers ($\sim 10^6 - 10^7$) of atoms and hence large numbers ($\sim 10^4$) of substitutional N impurities, even at ultra-dilute N compositions, with the occurrence of N–N pairs and larger clusters of N atoms determined statistically. This then minimizes the impact of statistical fluctuations on the calculated electronic structure and, in this sense, the LCINS approach is capable of providing an unambiguous description of the electronic structure at any given N composition x. The following approach to analyzing the LCINS electronic structure of a large supercell for a given N composition x provides significant insight into the nature of the CB structure. First, the full distribution of N-related states is considered independently of $|\psi_{CB}^{(0)}\rangle$ in order to determine the energies ϵ_i of the (interacting) N-related localized states centered on different lattice sites. For a supercell containing a total of M substitutional N impurities this consists of diagonalizing an $M \times M$ Hamiltonian matrix $\langle \psi_{\text{N},i} | \hat{H} | \psi_{\text{N},j} \rangle$ $(i, j = 1, ..., M)$ constructed in the basis $\{ |\psi_{\text{N},i}\rangle \}$ of N-related localized states associated with all N complexes present in the supercell. Second, the eigenvectors of this $M \times M$ Hamiltonian are used to construct the N-related LCINS basis states $|\phi_{\text{N},i}\rangle$, as in Equation 9.15 but omitting $|\psi_{\text{CB}}^{(0)}\rangle$. Third, the interaction $V_i = \langle \phi_{\text{N},i} | \hat{H} | \psi_{\text{CB}}^{(0)} \rangle$ between $|\psi_{\text{CB}}^{(0)}\rangle$ and each of the N-related LCINS basis states $|\phi_{\text{N},i}\rangle$ is computed explicitly for each $|\phi_{\text{N},i}\rangle$. Finally, the energies ϵ_i of the N-related LCINS basis states, and their interactions V_i with $|\psi_{\text{CB}}^{(0)}\rangle$, are used to generate an $(M+1) \times (M+1)$ Hamiltonian matrix in the basis $\{ |\phi_{\text{N},i}\rangle, |\psi_{\text{CB}}^{(0)}\rangle \}$, which can then be diagonalized to obtain the energies E_i and eigenstates $|\phi_i\rangle$ of the large, disordered N-containing supercell. This procedure enables the interactions (1) between the N-related localized states $|\psi_{\text{N},i}\rangle$, and (2) between each N-related localized state $|\psi_{\text{N},i}\rangle$ and $|\psi_{\text{CB}}^{(0)}\rangle$, to be quantified explicitly, thereby identifying the contributions of distinct classes of N isolated impurity, N–N pair, and N cluster states to determining the electronic structure of the disordered alloy.

As discussed above, a stringent test of any theoretical model of the GaN$_x$As$_{1-x}$ electronic structure is the accuracy with which it is capable of reproducing the measured (1) anomalously large values of m_e^* and g_e^* and their nonmonotonic variation with x, and (2) emergence of an "E_+" feature associated with N-derived states which are resonant with the GaAs CB for $x > 0.2\%$, and its energy broadening and disappearance for $x > 3\%$. We briefly summarize the analysis of these detailed aspects of the electronic structure here, both of which are described quantitatively by the LCINS approach.

The GaN$_x$As$_{1-x}$ electron effective mass at N composition x can be computed as [40]

$$m_e^*(x) = \frac{m_e^*(0)\, E_g(x)}{E_g(0) f_{\Gamma,\text{CB}}(x) f_{\Gamma,\text{VB}}(x)}\,, \tag{9.16}$$

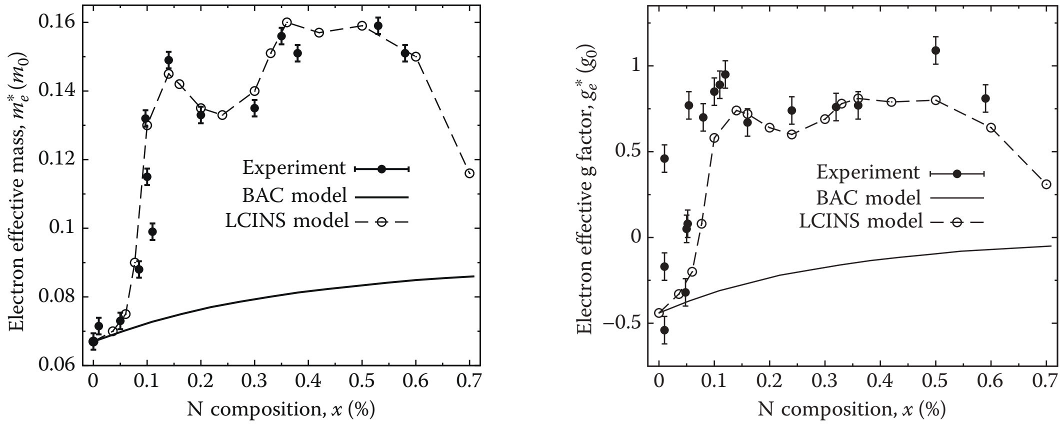
where $m_e^*(0)$ and $E_g(0)$ are the GaAs ($x = 0$) electron effective mass and band gap, $E_g(x)$ is the GaN$_x$As$_{1-x}$ band gap at N composition x, and $f_{\Gamma,\text{CB(VB)}}(x)$ is the Γ character of the GaN$_x$As$_{1-x}$ alloy CB (VB) edge state at N composition x. $f_{\Gamma,\text{CB}}(x)$ can be calculated directly by projecting the alloy CB edge eigenstate computed via the LCINS approach onto $|\psi_{\text{CB}}^{(0)}\rangle$ (cf. Section 9.2.2), while it can be assumed—in line with detailed supercell calculations—that $f_{\Gamma,\text{VB}}(x) = 1 - x$ since the GaN$_x$As$_{1-x}$ VB edge eigenstates evolve continuously from those of the unperturbed GaAs VB edge as in a conventional semiconductor alloy [12,29].

Closely related to the effective mass is the Landé effective g factor g_e^* of CB electrons. g_e^* is a key parameter in semiconductor materials, describing the response of the electron spins to externally applied magnetic fields as well as providing useful information regarding the symmetry of the crystal eigenstates at the Γ point. In addition to providing information pertinent to evaluating the potential of specific materials for spintronic applications, analysis of the electron spin properties and g factor therefore provides valuable insight into the material band structure. In particular, g_e^* is extremely sensitive to the separation in energy between zone-center eigenstates, as well as hybridzation between eigenstates caused by, e.g. a reduction of symmetry due to strain [41] or quantum confinement [42] or, in this case, incorporation of N. As a result, calculations of g_e^* also provide a stringent test of theoretical models of the electronic structure [43]. The GaN$_x$As$_{1-x}$ electron effective g factor at N composition x can be computed as [12]

$$g_e^*(x) = g_e^*(0)\left(1 - \frac{P^2 f_{\Gamma,\text{CB}}(x) f_{\Gamma,\text{VB}}(x)}{3}\left(\frac{1}{E_g(x)} - \frac{1}{E_g(x) - \Delta_{\text{SO}}(x)}\right)\right), \tag{9.17}$$

where P is the Kane interband momentum matrix element in GaAs, and where $E_g(x)$ and $\Delta_{\text{SO}}(x)$ are, respectively, the GaN$_x$As$_{1-x}$ band gap and spin-orbit-splitting energy at composition x.

Figure 9.2a and 9.2b [12,44] show, respectively, the variation of m_e^* and g_e^* with x in GaN$_x$As$_{1-x}$ calculated using the two-band BAC model of Section 9.2.2 (solid line) and the LCINS method (open circles), compared to experimental measurements (closed circles) [12]. First, we note the large deviation of the measured values of m_e^* and g_e^* from those calculated using the two-band BAC model, even at ultra-dilute



(a) (b)

FIGURE 9.2 Variation of (a) the electron effective mass m_e^*, and (b) the electron effective g factor g_e^*, as a function of N composition x in $\mathrm{GaN}_x\mathrm{As}_{1-x}$, calculated via Equation 9.16 using the two-band BAC (solid line) and LCINS (open circles, dashed lines) models, compared to experimental measurements (closed circles). The strong influence of N pair and cluster states on the alloy CB edge leads to high values of m_e^*, a change in the sign of g_e^*, and strong non-monotonic variation of both m_e^* and g_e^* with x, none of which is captured by the two-band BAC model. The values of m_e^* and g_e^* calculated using the LCINS approach are in good, quantitative agreement with experiment, verifying the strong impact of N-related alloy disorder on the $\mathrm{GaN}_x\mathrm{As}_{1-x}$ CB structure and emphasizing that a quantitative theoretical description requires direct treatment of N pair and cluster states. (Adapted from E.P. O'Reilly et al., *Semicond. Sci. Technol.*, 24, 033001, 2009; A. Lindsay and E.P. O'Reilly, *Phys. Rev. Lett.*, 93, 196402, 2004.)

N compositions $x < 0.1\%$. Second, we note the excellent quantitative agreement between the measured values of m_e^* and g_e^* and those calculated using the LCINS method. The LCINS approach accurately captures the strong increase in m_e^* between $x = 0.1\%$ and 0.2%, as well as the associated change in the sign of g_e^*, neither of which is captured by the BAC model. This quantitative agreement with experiment highlights that the LCINS method describes the nature and evolution of the CB structure of disordered $\mathrm{GaN}_x\mathrm{As}_{1-x}$ alloys with a sufficient level of accuracy to account for the observed nonmonotonic variation of m_e^* and g_e^* with x. Furthermore, the fractional Γ character $f_{\Gamma,\mathrm{CB}}(x)$ of the $\mathrm{GaN}_x\mathrm{As}_{1-x}$ CB eigenstates calculated using the LCINS approach provides insight into the origin of these unusual material trends.

We recall that the LCINS model directly describes the hybridzation of GaAs CB edge eigenstates with a full distribution of N-related localized states, including states which lie below the GaAs CB edge in energy (e.g., the N–N pair states described in Table 9.1). Detailed analysis of LCINS calculations on ultralarge supercells reveals that it is the hybridization between localized states associated with N–N pairs and larger clusters of N atoms that largely brings about the observed large values of m_e^* and g_e^*. As the alloy CB edge moves downward in energy with increasing x it passes through these N pair and cluster states, undergoing strong hybridization when its energy is close to the energy ϵ_i of N-related LCINS states having appreciable (but not necessarily large) V_i. This results in low calculated values $f_{\Gamma,\mathrm{CB}} < 50\%$ for the $\mathrm{GaN}_x\mathrm{As}_{1-x}$ CB edge eigenstates across certain ranges of composition, leading to the presence of several distinct maxima in the variation of m_e^* and g_e^* at select N compositions, and thereby accounting for both the large measured values of m_e^* and g_e^* as well as their nonmonotonic variation with x. It should be noted that this is in contrast to the two-band BAC model of $\mathrm{GaN}_x\mathrm{As}_{1-x}$, in which all N-related states are assumed to lie above the unperturbed GaAs CB edge in energy (cf. Table 9.1) and, hence, in which $f_{\Gamma,\mathrm{CB}}$ must exceed 50% for all x. These results emphasize that while the evolution of the E_g with x is primarily governed by the interaction of the GaAs CB with localized states associated with isolated N impurities (with N pair and cluster states playing a secondary role), and can hence be described by a simple BAC approach, the evolution of m_e^* and g_e^* is strongly impacted by alloy disorder and can hence only be described accurately via theoretical approaches which explicitly account for the impact of N pair and cluster states on the alloy CB edge eigenstates [12, 40,44]. Additional analysis using this approach has elucidated the variation of m_e^* as a function of applied hydrostatic pressure: the upward movement in energy of the host matrix CB edge in response to applied pressure causes strong hybridization as described above, resulting in strong, nonmonotonic variation of m_e^*, with increasing pressure [40].

Further calculations using the LCINS method have also quantitatively described the observed trends associated with the higher-energy, N-derived "E_+" feature in GaN_xAs_{1-x}, in terms of the evolution with increasing x of the GaAs CB edge Γ character $f_{\Gamma,CB}$ acquired by N-related localized resonant states [40]. At ultra-dilute N compositions $x < 0.2\%$, the composition-dependent hybridization of the N-related resonant states is sufficiently weak to ensure that the E_+ feature acquires little appreciable GaAs CB edge Γ character, indicating that it is not likely to be resolvable as a distinct, optically active feature in spectroscopic measurements. As the N composition increases, the N-related resonant states hybridize more strongly with $|\psi_{CB}^{(0)}\rangle$, acquiring larger $f_{\Gamma,CB}$ and leading to the emergence of the distinct, energetically narrow E_+ feature observed in experiment. As the N composition is further increased above $x \approx 1.4\%$, the repulsive BAC interaction between $|\psi_{CB}^{(0)}\rangle$ and the N-related resonant states pushes the E_+ feature sufficiently high in energy that it becomes resonant with the large density of CB states associated with the L-minima. The resonance of these localized N-related states with the large density of L-valley states then leads to a strong energetic broadening of the E_+ feature (cf. Section 9.5.1), manifest in the LCINS calculations by a distribution of $f_{\Gamma,CB}$ over a large number of higher-energy alloy CB states. This redistribution of $f_{\Gamma,CB}$ brought about by resonance with L-valley states continues with increasing x, leading to the E_+ feature weakening and diminishing in prominence. For N compositions $x > 2.5\%$ this broadening of the E_+ feature becomes sufficiently large as to become indistinguishable as a clear feature in the CB structure. The LCINS approach, therefore, not only accounts for the inability to observe a distinct E_+ feature in spectroscopic measurements at N compositions $x > 3\%$, but also elucidates the mechanism by which this feature vanishes.

We note that it is also possible to calculate the band dispersion of N-containing alloys using the LCINS method. We reserve discussion of the LCINS band dispersion until Section 9.5.1, in which we review the use of Green's function methods in conjunction with the AIM to facilitate quantitative analysis of the CB structure and DOS in disordered GaN_xAs_{1-x} alloys, accurate calculations of which are essential to predicting the transport properties of N-containing alloys. Overall, the LCINS approach highlights (1) the breakdown of the simple two-band BAC model of the dilute nitride CB structure in the presence of N-related alloy disorder as well as (2) the ability to quantitatively describe the electronic structure of a disordered N-containing alloy by explicitly treating the interaction between the unperturbed GaAs CB edge and the full distribution of N-related localized states. In particular, these results highlight the important role played by short-range alloy disorder in determining the details of the electronic structure. This is a general characteristic of highly-mismatched alloys, clearly distinguishing their properties from those of conventional semiconductors. While we have outlined the main principles and some key results of the LCINS model here, full details of this approach are beyond the scope of this chapter. We refer the reader to Refs. [40] and [44] for further information on the development of the LCINS approach and its application to GaN_xAs_{1-x} alloys, and to Refs. [24]and [25] for details of its application to GaN_xP_{1-x} alloys.

9.3 Theory of GaAs-Based 1.3-μm InGaNAs Quantum Well Lasers

Having focused so far on developing a consistent approach to model the $(In)GaN_xAs_{1-x}$ band structure, in this section we turn our attention to applying this theory to analyze photonic devices. Specifically, we consider the theory and modeling of 1.3-μm InGaNAs QW lasers grown on GaAs. We begin Section 9.3.1 by outlining a theoretical model for the electronic and optical properties of InGaNAs QW lasers based on the 10-band $\mathbf{k}\cdot\mathbf{p}$ model of Section 9.2.2. Next, in Section 9.3.2, we review a series of theoretical calculations of the band structure and optical gain in an $In_{0.36}Ga_{0.64}N_{0.018}As_{0.982}$ single QW laser structure, containing 1.8% N and chosen on the basis of the laser device investigated in Refs. [45] and [33]. Through this analysis we elucidate the impact of N incorporation on the electronic and optical properties of ideal N-containing QWs, allowing us to identify key trends in the material and device parameters for this exemplar 1.3-μm dilute nitride device. Finally, in Section 9.3.3, we discuss the performance of real InGaNAs QW lasers. We focus on the impact of N incorporation on the loss mechanisms present in devices of this type, by reviewing calculations of the temperature dependence of defect-related, radiative, and nonradiative (Auger) recombination in 1.3-μm devices. We first outline how experimental measurements can be undertaken to identify and quantify the recombination pathways present in a QW laser, before comparing the results of theoretical calculations of the recombination rates directly to experiment.

The comparisons to experimental data we review display the predictive capability of the theory, highlighting in particular the ability of the 10-band **k·p** model to describe the impact of N incorporation on the band structure and optical properties. This confirms the validity of the theoretical model described in Section 9.3.1 for use in the design and optimization of future dilute nitride semiconductor lasers and related devices.

9.3.1 Theoretical Model for Dilute Nitride Quantum Well Lasers

Here, we provide a brief outline of a theoretical model that has been developed to study the electronic and optical properties of InGaNAs QW lasers. Full details of this model can be found in Refs. [33,34]. The theoretical model is based on the 10-band **k·p** Hamiltonian presented in Section 9.2.2. The QW band offsets are calculated following the procedure outlined in Ref. [33], while the QW eigenstates are calculated in the envelope function approximation using a numerically efficient plane wave expansion method (similar to that commonly employed in *ab initio* electronic structure calculations). An overview of the plane wave expansion method within the context of **k·p** theory can be found in Ref. [46].

In the plane wave approach, the QW eigenstate $|\psi_n(\mathbf{k}_\parallel, z)\rangle$ at position z and in-plane wave vector $\mathbf{k}_\parallel$, having energy $E_n(\mathbf{k}_\parallel)$, is written as

$$|\psi_n(\mathbf{k}_\parallel, z)\rangle = \frac{1}{\sqrt{L}} \sum_{b=1}^{10} \sum_{m=-M}^{M} a_{nbm}(\mathbf{k}_\parallel)\, e^{iG_m z} |u_b\rangle \,, \tag{9.18}$$

where L is the periodicity (length) of the calculational supercell ($-\frac{L}{2} \leq z \leq \frac{L}{2}$), b indexes the bands of the bulk **k·p** Hamiltonian, and m indexes the set of plane wave basis states with which the components of the envelope function associated with $|\psi_n(\mathbf{k}_\parallel, z)\rangle$ are represented. The discrete wave vectors G_m are defined as $G_m = \frac{2m\pi}{L}$, and $|u_b\rangle$ is the zone center Bloch states of the bulk 10-band **k·p** Hamiltonian. We note that this description of the QW eigenstates, which consists of $2M + 1$ independent Fourier components, is formally exact as $M \to \infty$. However, for most practical calculations a basis set containing ~ 70 plane waves ($M \sim 35$) is sufficient to ensure that the calculated eigenstates have converged with respect to further increases in the basis set size [34].

By substituting Equation 9.18 into the multiband Schrödinger equation it is straightforward to derive analytical expressions for the matrix elements of the QW Hamiltonian in the plane wave basis set defined by the set of wave vectors $\{G_m\}$ [34,46], leading to a reciprocal space Hamiltonian matrix of size $10(2M+1) \times 10(2M + 1)$. In this manner, an independent Hamiltonian matrix is obtained at each in-plane wave vector $\mathbf{k}_\parallel$, and diagonalizing these matrices yields the QW band structure $E_n(\mathbf{k}_\parallel)$ and eigenstates $|\psi_n(\mathbf{k}_\parallel, z)\rangle$. We note that, by definition, the eigenstates described by Equation 9.18 satisfy periodic boundary conditions over the calculational supercell, meaning that the plane wave expansion method can readily be used to investigate the electronic and optical properties of superlattices.

In order to facilitate accurate calculation of the optical properties, it is important to include the strong impact of N on the QW eigenstates. For this reason, the QW band structure and envelope functions (calculated in the axial approximation—that is, by setting $\mu = 0$ in Equation 9.10 [47]) are used *directly* to compute the optical properties of a given laser structure. Key to this analysis is the calculation of the optical (momentum) matrix elements. Following Ref. [48], the optical matrix element of an electron–hole pair at in-plane wave vector $\mathbf{k}_{\parallel}$ is given in terms of the QW eigenstates described by Equation 9.18 as

$$P_{n_c,n_v}\left(\mathbf{k}_{\parallel}\right) = \frac{m_0}{\hbar}\left\langle \psi_{n_v}\left(\mathbf{k}_{\parallel},z\right)\left|\frac{\partial\widehat{H}}{\partial\mathbf{k}_{\parallel}}\cdot\widehat{e}\right|\psi_{n_c}\left(\mathbf{k}_{\parallel},z\right)\right\rangle, \tag{9.19}$$

where $\widehat{e}$ is a unit vector denoting the polarization of the emitted/absorbed photon and n_c (n_v) indexes the conduction (valence) subbands. This approach to calculating the optical matrix elements then explicitly incorporates key band structure effects in the computation of the optical properties, including N-induced BAC and carrier localization in addition to temperature- and injection-dependent carrier spillout from the QW(s), band mixing, and pseudomorphic strain. When discussing optical transition "strengths" below, we refer for simplicity to the quantity $\frac{1}{2m_0}|P_{n_c,n_v}\left(\mathbf{k}_{\parallel}\right)|^2$, which has units of energy.

Using the optical matrix elements computed in this manner, the calculation of the optical (material) gain spectrum at a given carrier density follows the approach outlined in Ref. [49], which is based upon a transformation of the spontaneous emission (SE) spectra. We note that we are concerned in this section solely with the transverse electric (TE)-polarized optical gain, which is of significantly larger magnitude than the transverse magnetic (TM)-polarized gain in the compressively strained structures under investigation. Comparison of the SE spectra calculated using this approach with temperature-dependent experimental measurements suggests that the spectral broadening for InGaNAs QWs is best described using a hyperbolic secant lineshape of the form $S(\hbar\omega) = \frac{1}{\pi\delta}\operatorname{sech}\left(\frac{E_0-\hbar\omega}{\delta}\right)$, where δ is the spectral linewidth [33]. Analysis of the SE spectra at threshold for the devices investigated in Ref. [33] demonstrated that incorporation of N leads to significant spectral broadening. This strong broadening of the optical spectra is primarily attributable to the impact of N-related alloy disorder on the CB structure, highlighting the important role played by the full distribution of N-related localized states in determining the details of the material properties (cf. Section 9.2). The determined value $\delta = 18$ meV for the spectral linewidth at room temperature is approximately three times larger than the typical value $\delta = 6.6$ meV (corresponding to an interband relaxation time of approximately 0.1 ps) observed in N-free devices at the same wavelength [33].

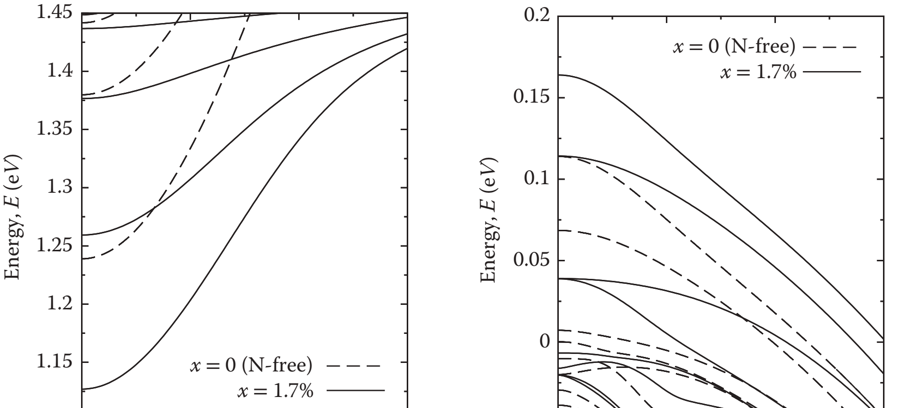
As we demonstrate in Sections 9.3.2 and 9.3.3, the results of calculations undertaken using this theoretical model are in good agreement with a range of experimental measurements, confirming the strong impact of N incorporation on the device performance and quantifying key trends in the expected device properties. We note that alternative theoretical approaches have been developed by several other groups. These alternative models are generally based on the 10-band **k·p** model of Section 9.2.2, which is used to calculate the QW band structure and single-particle eigenstates that serve as the starting point for the computation of the optical properties. For example, Koch et al. have performed many-body calculations of the optical gain [50] and carrier recombination [51] in 1.3-μm InGaNAs lasers using the semiconductor Bloch equations [52]. The device properties and performance computed using this detailed approach are largely in accordance with those calculated using that outlined in this chapter, and have been shown to be in good quantitative agreement with experimental measurements [53].

9.3.2 Impact of Nitrogen Incorporation on the Band Structure and Gain of Quantum Wells

In order to quantify the effects of N incorporation on the electronic and optical properties of QW laser structures, we compare the calculated gain characteristics of an $\mathrm{In_{0.36}Ga_{0.64}N_{0.017}As_{0.983}}$/GaAs ($x = 1.7\%$) single QW laser structure to those of an N-free $\mathrm{In_{0.36}Ga_{0.64}As}$/GaAs QW laser structure [33]. We review here the influence of N incorporation on the calculated QW band structure, optical transition strengths, radiative current density, material gain, and differential gain, and identify general trends in the electronic and optical properties of N-containing laser structures. Following this, we review the impact of electrostatic confinement on the optical gain by summarizing the results of calculations in which the QW band structure and eigenstates are computed self-consistently by solving the multiband Schrödinger equation for the laser structure coupled to Poisson's equation for the (assumed) thermalized carrier density associated with injected electrons and holes [34].

The calculated CB, HH, and LH band offsets are $\Delta E_{\mathrm{CB}} = 245$ meV, $\Delta E_{\mathrm{HH}} = 130$ meV, and $\Delta E_{\mathrm{LH}} = 5$ meV in the compressively strained N-free QW. The incorporation of 1.7% N reduces the compressive in-plane strain from 2.51% in the N-free QW to 2.19% in the $\mathrm{In_{0.36}Ga_{0.64}N_{0.017}As_{0.983}}$ QW and, when BAC effects are taken into account, this leads to calculated CB, HH, and LH band offsets $\Delta E_{\mathrm{CB}} = 354$ meV, $\Delta E_{\mathrm{HH}} = 181$, meV, and $\Delta E_{\mathrm{LH}} = 7$ meV in the N-containing QW.

The solid lines in Figures 9.3a and 9.3b show, respectively, the CB and VB structures calculated using the 10-band **k·p** Hamiltonian in conjunction with the plane wave expansion method for ideal, 7-nm-thick $\mathrm{In_{0.36}Ga_{0.64}As/GaAs}$ (dashed lines; N-free) and $\mathrm{In_{0.36}Ga_{0.64}N_{0.017}As_{0.983}/GaAs}$ (solid lines; $x = 1.7\%$) QWs. The calculated transition energy between the lowest energy bound electron state ($e1$) and the highest energy hole state ($hh1$) is calculated to be 1.124 eV in the N-free QW and 0.962 eV in the N-containing QW. Of this 161 meV reduction in the QW band gap, only 50 meV is due to the N-induced changes in the VB edge energy, while the remaining 111 meV results from the BAC-induced downward shift of the CB edge energy. The impact of N incorporation on the CB structure, described in terms of the BAC interaction (cf. Section 9.2.2), is clearly visible: (1) incorporation of N leads to a strong decrease in the energy of the electron states at the QW zone center, $k_{\parallel} = 0$, and (2) at nonzero in-plane wave vector the BAC interaction pushes the CB dispersion downward in energy leading to strong CB nonparabolicity, increased band edge electron effective masses, and a strong increase in the CB DOS. These calculations indicate that the DOS



(a)

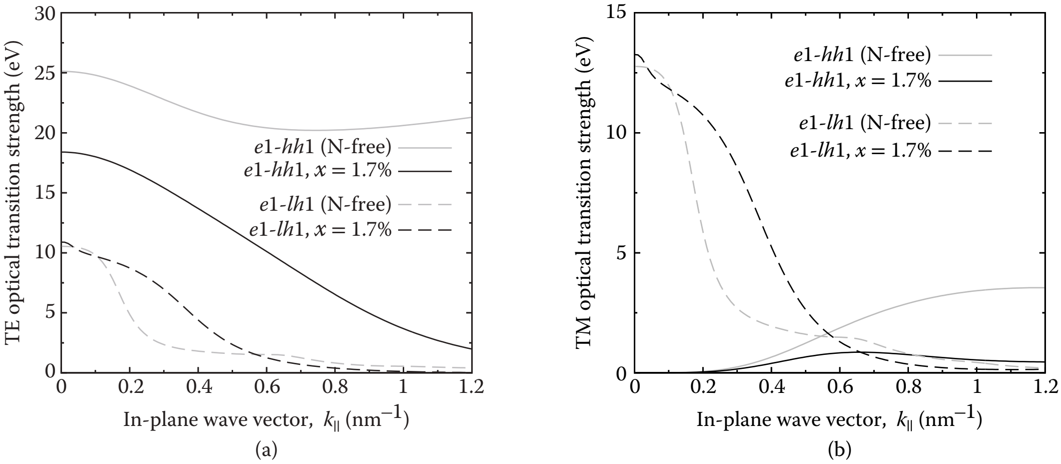
(b)

FIGURE 9.3 Calculated room temperature (a) CB and (b) VB structure for 7-nm-thick $\mathrm{In_{0.36}Ga_{0.64}As}$ (dashed lines; N-free) and $\mathrm{In_{0.36}Ga_{0.64}N_{0.017}As_{0.983}}$ (solid lines; $x = 1.7\%$) QWs having GaAs barriers. The zero of energy in each case is taken at the VB edge of the unstrained GaAs barrier. Incorporation of N leads to a strong anticrossing effect and reduction (increase) in the energy (effective mass and DOS) of the CB states, while leaving the VB states relatively unaffected, aside from an upward shift in energy brought about by a reduction of the in-plane compressive strain in the N-containing QW. The primary result of the impact of N on the band structure is a considerable redshift of the QW emission wavelength. The consequences of the N-induced modifications to the band structure and DOS for laser operation are described in detail in the text.

effective mass at $k_{\parallel} = 0$ for the lowest energy CB increases from $0.046m_0$ in the N-free QW, to $0.060m_0$ in the $In_{0.36}Ga_{0.64}N_{0.017}As_{0.983}$ QW (where m_0 is the free electron mass).

By comparison, the impact of N incorporation on the VB structure is relatively minor. The primary effect on the VB structure is associated with the N-induced reduction in compressive strain in the N-containing QW, which acts to slightly increase the energies of the bound hole states at fixed in-plane wave vector. The ordering of the hole states is calculated to be the same in both the N-free and N-containing QWs: the three highest-energy hole states have predominately HH character at $k_{\parallel} = 0$, with the ordering of the hole states being $hh1$, $hh2$, $hh3$, and $lh1$. Given that the VB edge DOS is largely unaltered by incorporation (aside from a relatively rigid strain-related energy shift), the impact of N incorporation on the gain characteristics is almost entirely attributable to the N-induced modifications to (1) the CB structure and DOS, (2) the interband optical transition strengths.

The N-induced hybridization between the host matrix CB edge states and N-related localized states leads to a portion of the CB eigenstates consisting of an admixture of N-related localized states which do not couple optically to the p-like states at the VB edge [29]. Despite the tendency of N incorporation to enhance the electron confinement—due to (1) the increase in CB offset ΔE_{CB}, and (2) enhanced spatial localization of the electron charge density, in the InGaNAs QW layer—the net effect of the N-induced hybridization is to decrease the interband optical transition strengths compared to those in a conventional N-free structure. Figure 9.4a and 9.4b shows, respectively, the calculated variation of the TE- and TM-polarized optical transition strengths as a function of in-plane wave vector $k_{\parallel}$, for transitions from the lowest-energy electron states $e1$ to the highest-energy HH- and LH-like states $hh1$ (solid) and $lh1$ (dashed), in the $In_{0.36}Ga_{0.64}N_{0.017}As_{0.983}$/GaAs (N-containing; black lines) and $In_{0.36}Ga_{0.64}As$/GaAs (N-free; gray lines) QW structures. As described above, N-induced hybridization leads to an overall decrease in the optical transition strengths. For example, the optical transition strength $|M_{e1-hh1}|^2$ calculated at $k_{\parallel} = 0$ for the TE-polarized $e1$–$hh1$ transition in the $In_{0.36}Ga_{0.64}N_{0.017}As_{0.983}$ QW is approximately 70% of that calculated for the N-free structure. It is also interesting to note that the calculated reduction in $|M_{e1-hh1}|^2$ and increase in the CB edge effective mass m_e^* causes the product $|M_{e1-hh1}|^2 \times m_e^*$ to be reduced when N



(a) (b)

FIGURE 9.4 Calculated variation of the (a) TE-polarized and (b) TM-polarized optical transition strengths as a function of the in-plane wave vector $k_{\parallel}$ for the $e1$-$hh1$ (solid lines) and $e1$-$lh1$ (dashed lines) transitions in $In_{0.36}Ga_{0.64}As$/GaAs (N-free; gray lines) and $In_{0.36}Ga_{0.64}N_{0.017}As_{0.983}$/GaAs ($x = 1.7\%$; black lines) QWs. Note the difference in scale between (a) and (b). It can be clearly seen that N-induced hybridization of CB states leads to a significant reduction of the strength of the TE-polarized $e1$-$hh1$ transitions, which account for the majority of the material gain g in these compressively strained QWs. This suggests that N incorporation can be expected to reduce the magnitude of g compared to that which can be achieved an N-free QW.

is incorporated in the QW. This is expected on the basis of **k·p** theory, in which the dominant contribution to the inverse bulk CB effective mass is directly proportional to $|M_{e1-hh1}|^2$, which describes the coupling of the *s*-like CB edge states to the *p*-like VB edge states via the momentum operator, and hence the large increase in the CB DOS associated with N incorporation.

Having investigated the N-induced modifications to the band structure and optical transition strengths we now turn our attention to the calculated gain characteristics, which we again compare for the N-containing and N-free laser structures. The band structure and optical transition strengths presented in Figures 9.3a and 9.3b, and , and in Figures 9.4a and 9.4b were used to calculate the variation of the peak TE-polarized material gain g as a function of carrier density n, and of radiative current density J_{rad}, for temperatures $T = 200$, 300 and 400 K. The solid (dashed) lines in Figures 9.5a and 9.5b denote the calculated variation of g with n and J_{rad} for the laser structure having an N-containing (N-free) $\text{In}_{0.36}\text{Ga}_{0.64}\text{N}_{0.017}\text{As}_{0.983}$ ($\text{In}_{0.36}\text{Ga}_{0.64}\text{As}$) QW [54].

Examining first the calculated variation of g with n in Figure 9.5a, we note that incorporation of N leads to an increase in the carrier density n_{tr} at transparency. Further analysis reveals that this behavior arises largely due to two factors: (1) the general reduction of the optical transition strengths due to N-induced hybridization with localized impurity states, and (2) the N-induced increase in the CB edge effective mass m_e^*. The increase in m_e^* is associated with an increase in the DOS at the CB edge, which leads to a reduction in the energy separation ΔF between the electron and hole quasi-Fermi levels at fixed carrier density. As a result of these N-induced changes to the band structure and optical transition strengths, the peak gain in the N-containing structure at fixed n is expected to be reduced compared to that in the N-free structure. This is confirmed by the numerical calculations, which compared to the N-free case show (1) an overall reduction of g, and (2) a saturation of the material gain at a lower level.

Turning our attention to Figure 9.5b, we note that the impact of N incorporation on the calculated variation of the peak material gain with J_{rad} is broadly similar to that observed for the variation of g with n. The calculated reduction of g at fixed J_{rad} in the N-containing structures reflects the fact that, for fixed quasi-Fermi level separation ΔF, the radiative current density J_{rad} in a QW laser is approximately proportional to $m_r^* \times |M_{e1-hh1}|^2$, where m_r^* is the reduced effective mass of a band edge electron–hole pair. Because the effective mass of a hole at the VB edge is larger than that of a CB edge electron, the reduced effective mass m_r^* is determined primarily by m_e^*. Based on our analysis of the band structure and optical transition

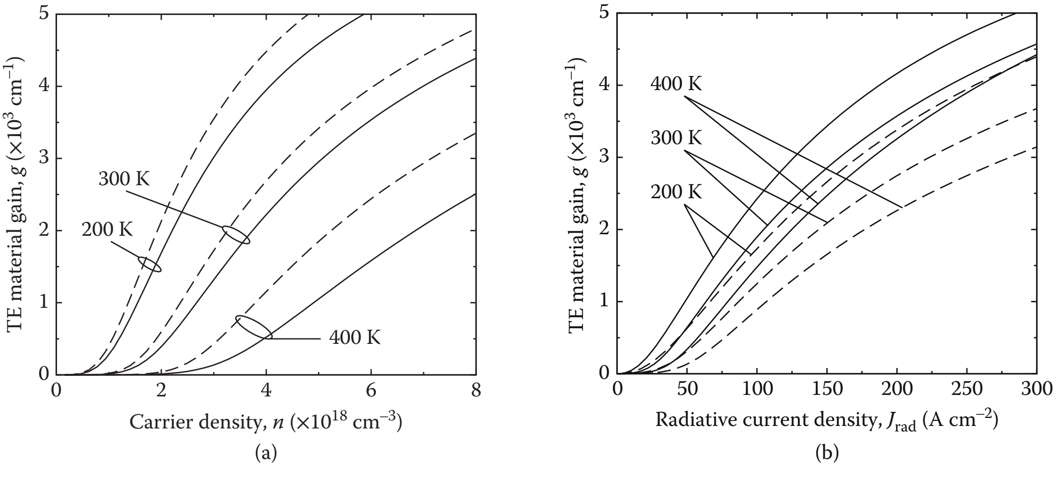


FIGURE 9.5 Calculated variation of the TE-polarized material gain g as a function of (a) carrier density n and (b) radiative current density J_{rad}, for $\text{In}_{0.36}\text{Ga}_{0.64}\text{As}/\text{GaAs}$ (N-free; dashed lines) and $\text{In}_{0.36}\text{Ga}_{0.64}\text{N}_{0.017}\text{As}_{0.983}/\text{GaAs}$ ($x =$ 1.7%; solid lines) QWs at temperatures $T = 200$, 300, and 400 K. Incorporation of N brings about a reduction of g at fixed n and J_{rad} and leads to gain saturation at lower g compared to the N-free QW.

strengths above, we conclude that the N-induced increase in m_r^* is insufficient to overcome the corresponding reduction in $|M_{e1-hh1}|^2$, meaning that the product $m_r^* \times |M_{e1-hh1}|^2$ undergoes a moderate reduction when N is incorporated into the QW. The result of this is that $J_{\mathrm{rad}} \propto m_r^* \times |M_{e1-hh1}|^2$ is reduced at fixed n compared to that in an N-free QW, thereby accounting for the calculated reduction of the material gain at fixed J_{rad}, primarily in terms of the impact of N on the TE-polarized interband optical transition strengths.

The differential gain $\frac{dg}{dn}$ plays a key role in determining both the threshold carrier density and bandwidth of a directly modulated semiconductor laser. For example, the modulation response frequency ω_{r} is proportional to the square root of the differential gain with respect to the carrier density, $\omega_{\mathrm{r}} \propto \sqrt{\frac{dg}{dn}}$ [55,56]. Figure 9.6 shows the calculated variation of $\frac{dg}{dn}$ at transparency (squares) and at threshold (circles) in a series of ideal 7-nm-thick $In_yGa_{1-y}N_xAs_{1-x}$ QW structures, calculated using linewidths $\delta = 6.6$ meV (closed symbols, solid lines) and 18 meV (open symbols, dashed lines) [57]. In order to maintain a constant QW ground-state $e1$–$hh1$ transition energy (emission wavelength) of 0.954 eV (1.3 μm) in the range of N compositions from $x = 0$ to 3%, the In composition must be reduced in order to compensate for the reduction in the QW band gap brought about by N incorporation. The calculated In compositions y required to maintain 1.3-μm emission ranges from 55% in the N-free case to 30% at $x = 3\%$ [57]. For fixed spectral linewidth δ, the calculated differential gain decreases with increasing N composition, with the most rapid decrease observed at low x. As expected, $\frac{dg}{dn}$ also decreases with increasing spectral linewidth. Due to the fact that the measured spectral broadening in InGaNAs QWs increases with increasing x [58], we conclude that the calculated value $\frac{dg}{dn} \sim 0.86 \times 10^{-15}$ cm^{-1} at threshold for $x = 2\%$ is approximately about 40% of that which could be achieved in an ideal InGaAs 1.3-μm QW laser [9,57].

As can be seen from the calculated variation of the peak material gain with carrier density in Figure 9.5a, the initial rapid decrease in $\frac{dg}{dn}$ is due to the strong hybridization between the N-resonant states and the CB edge states, which occurs even for small N compositions x. Figure 9.6 then suggests that an optimized InGaNAs/GaAs QW laser device should contain minimal N, ideally being N-free. However, this is not possible due to the excessively large strain required to achieve 1.3-μm emission from an InGaAs/GaAs

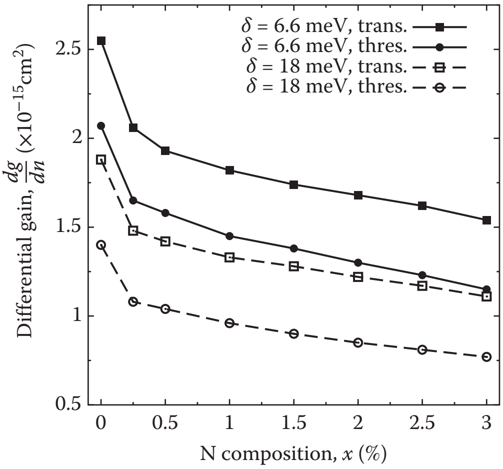


FIGURE 9.6 Calculated variation of the room temperature differential gain $\frac{dg}{dn}$ at transparency (closed squares) and at threshold (closed circles) as a function of N composition x in a series of (In)GaN_xAs_{1-x}/GaAs QW structures, using homogenous linewidths $\delta = 6.6$ meV (solid lines, closed symbols) and $\delta = 18$ meV (dashed lines, open symbols) for the computation of the material gain g. At each N composition x the In composition in the QW has been adjusted to maintain the emission wavelength at 1.3 μm. It can be seen that both N incorporation and N-induced broadening of the gain spectrum lead to a reduction of $\frac{dg}{dn}$ for devices having a fixed 1.3-μm emission wavelength (Adapted from S. Tomić and E. P. O'Reilly, *IEEE Photon. Tech. Lett.*, 15:6, 2003.).

QW [9]. Using the strain-thickness criteria presented in Ref. [59], it is estimated on the basis of the calculated trends for this series of InGaNAs QWs that one needs N and In composition $x > 1.5\%$ and $y < 39\%$ to achieve 1.3-μm emission in a pseudomorphic $In_yGa_{1-y}N_xAs_{1-x}$/GaAs QW, compared to approximately 55% In in an N-free InGaAs structure [9]. The N composition required to achieve 1.3-μm emission can be further reduced by adding more In to narrower QWs and/or by growing tensile strained barrier layers above and below the QW, as demonstrated by Tansu et al., who achieved 1.3-μm emission with $x = 0.5\%$ in a 6-nm-thick QW by utilizing strain compensating tensile strained GaAsP [60,61] and GaAsN [62] barrier layers.

Experimental analysis has led to some controversy as to the origin of the magnitude and temperature sensitivity of the threshold current density J_{th} in 1.3-μm InGaNAs/GaAs QW lasers. Tansu et al. have suggested that the observed large values of J_{th} are associated with hole leakage, which impacts upon the device performance at high temperature [63]. That this might be the case appears consistent with the analysis reviewed above, which demonstrates that the BAC interaction leads to strong electron confinement, but low VB offsets and comparatively weak confinement of holes. Given the large mismatch in the CB and VB offsets in InGaNAs/GaAs QWs, electrostatic confinement of holes by strongly bound and localized electrons can be expected to play an important role in determining the electronic and optical properties of lasers under injection. In order to investigate the potential for hole leakage in InGaNAs laser structures, it is then important to calculate the electron and hole charge densities self-consistently, by solving Poisson's equation to obtain the net electrostatic potential at each injected carrier density n. This task can be accomplished efficiently using a plane wave basis set, since Poisson's equation can be solved analytically in reciprocal space [34].

This self-consistent approach has been applied to theoretically analyze the impact of hole leakage on the gain characteristics of InGaNAs/GaAs(P) QW laser structures. Full details of this self-consistent analysis of the gain characteristics, which are beyond the scope of this chapter, can be found in Ref. [34]. Here we summarize that these temperature- and injection-dependent calculations of the material gain suggest that electrostatic confinement of holes by electrons significantly reduces hole leakage at higher carrier densities, a conclusion which is confirmed by comparing the hole states calculated for structures having GaAs barriers to those having GaAsP barriers (the latter having larger VB offsets and enhanced hole confinement). Taking electrostatic confinement into account leads to a slight increase in the hole quasi-Fermi level at fixed injection, decreasing the quasi-Fermi level separation ΔF and hence increasing the material gain at fixed carrier density. Aside from these minor qualitative changes, the gain characteristics calculated self-consistently for InGaNAs/GaAs(P) QWs are largely the same as those described above. This analysis then suggests that hole leakage does not play a major role in determining the threshold characteristics and performance of 1.3-μm InGaNAs QW lasers, and that other mechanisms must therefore be considered in order to describe the measured magnitude and temperature sensitivity of J_{th}. It is to this question that we turn our attention in Section 9.3.3.

Overall, we conclude that despite the tendency of N to slightly degrade the gain characteristics, 1.3-μm InGaNAs QW lasers offer good (theoretical) performance compared to idealized highly strained InGaAs/-GaAs QWs at the same wavelength. While incorporating N entails trading off some performance compared to equivalent N-free laser structures, utilizing InGaNAs QWs allows to overcome the strain limitations associated with InGaAs QWs, and extends the range of emission wavelengths accessible using a GaAs substrate. Comparing these results for the exemplar GaAs-based InGaNAs laser structure considered above with theoretical analysis of conventional 1.3-μm InP-based InGaAsP and AlInGaAs QW laser structures [33,64–66], we find that InGaNAs QWs offer (1) higher optical gain, and (2) larger differential gain, with the value at threshold $\frac{dg}{dn} \sim 8 \times 10^{-14}$ cm^2 (calculated at $T = 300$ K) being approximately 33% larger than the value $\sim 6 \times 10^{-14}$ cm^2 calculated for an InP-based device having the same compressive strain, optical gain, and QW thickness. We note that this is the case despite the significantly larger spectral linewidth $\delta = 18$ meV in the InGaNAs structure, compared to the value $\delta = 6.6$ meV assumed for the InP-based structure, highlighting the intrinsically superior characteristics of GaAs-based heterostructures. Furthermore, improved optical confinement can be achieved in GaAs-based 1.3-μm InGaNAs laser structures compared to InP-based devices, because of the larger refractive index contrast achievable through use of AlGaAs cladding layers. This reduces the number of QWs required in an optimized edge-emitting device, while also opening the possibility of using InGaNAs alloys to develop GaAs-based 1.3- and 1.55-μm VCSELs. The theoretical analysis reviewed here therefore confirms the potential of dilute nitride alloys for applications in long-wavelength GaAs-based photonics, and in particular for edge- and surface-emitting laser applications in the telecommunication wavelength ranges.

9.3.3 Impact of Nitrogen Incorporation on Carrier Recombination: Theory versus Experiment

While there has been considerable interest in the influence of N incorporation on the electronic structure and optical transitions in InGaNAs/GaAs QWs, there have been comparatively few quantitative investigations of the primary loss mechanisms in lasers based on these material systems. Such studies are of interest since, for example, they allow a direct comparison between the relative importance of radiative and non-radiative recombination pathways in dilute nitride lasers versus their conventional InP-based (N-free) counterparts. Non-radiative recombination processes make a significant contribution to the threshold current in conventional InP-based lasers, leading to strong temperature dependence of the threshold current [3,67]. A number of additional factors can also contribute to this temperature dependence, such as carrier leakage due to poor confinement and thermionic emission [68,69], or intervalence band absorption (IVBA) [64,70]. However, the dominant loss mechanism in 1.3- and 1.55-μm InP-based lasers is generally acknowledged to be non-radiative Auger recombination [67,71]. We therefore begin our discussion of the performance of InGaNAs QW lasers with a description of the quantitative analysis of recombination pathways in 1.3-μm InGaNAs/GaAs lasers, highlighting the importance of defect-related and Auger recombination [45]. We provide an interpretation of the observed device characteristics in terms of the detailed theoretical analysis of the impact of N incorporation on Auger recombination rates in InGaNAs QWs [72]. This is then used as a basis to understand the threshold characteristics measured for a range of InGaNAs QW lasers across the 1.2- to 1.6-μm wavelength range.

Under the assumptions of (1) charge neutrality in the active region, that is, the electron density n being equal to the hole density p, and (2) negligible carrier leakage, the current density J in a semiconductor laser can be written as

$$J = J_{\mathrm{mono}} + J_{\mathrm{rad}} + J_{\mathrm{Auger}} + J_{\mathrm{leak}} = e\left(An + Bn^2 + Cn^3\right) + J_{\mathrm{leak}}\,, \tag{9.20}$$

where J_{mono}, J_{rad}, and J_{Auger}, respectively refer to the current densities associated with defect-related (monomolecular), radiative, and Auger recombination, e is the magnitude of the electrical charge of an individual carrier (electron or hole), and J_{leak} is the current density associated with leakage of carriers from the active region. In the Boltzmann approximation, the current density associated with each of these recombination mechanisms is proportional to the products of the densities of each type of carrier involved in the process. For example, the radiative component of J—describing SE due to recombining electron–hole pairs—is proportional to np which, under the aforementioned assumption of overall charge neutrality, is equal to n^2. Similarly, the component of J associated with Auger recombination processes—three carrier processes involving two electrons and one hole, or one electron and two holes—is proportional to n^3. The coefficients A, B, and C on the right-hand side of Equation 9.20, which describe the monomolecular, radiative, and Auger recombination rates, respectively, depend strongly on the details of the active region band structure and, in a real device, on the material quality. We note that the assumption of negligible carrier leakage, $J_{\mathrm{leak}} \approx 0$, is found to be generally valid for 1.3-μm InGaNAs laser structures, due to good carrier confinement brought about by the large CB and VB offsets.

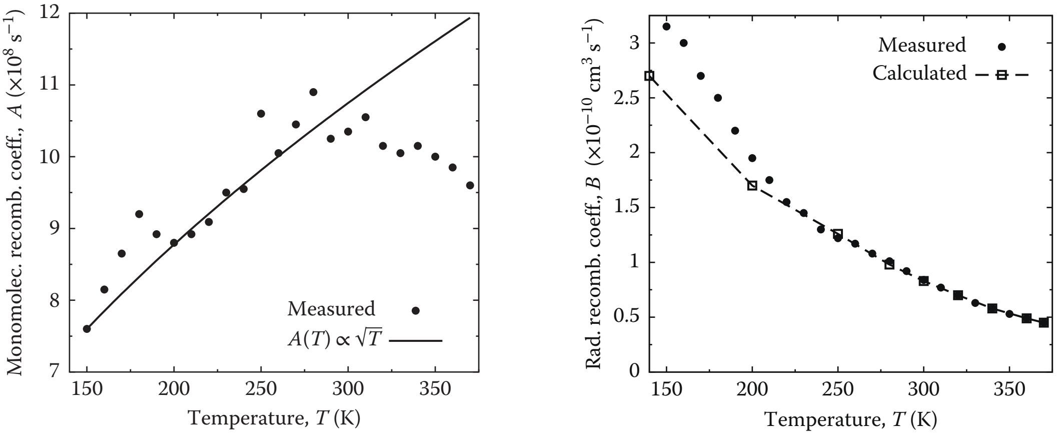
Experimental determination of the individual components J_{mono}, J_{rad}, and J_{Auger} of Equation 9.19 relies on the fact that, over a limited range of current densities, Equation 9.20 can be written as $J \propto n^z$, where $1 \leq z \leq 3$, and with the value of z varying depending on whether the dominant contribution to the current density is due to defect-related ($z = 1$), radiative ($z = 2$), or Auger ($z = 3$) recombination [67,73].

Measuring the integrated SE rate $L \propto Bn^2$ then allows the total current density J to be related to L as $J \propto L^{\frac{z}{2}}$, which can be rewritten as $\log(J) \propto z \log(\sqrt{L})$. This enables z to be quantified experimentally for a given laser structure by plotting $\log(J)$ against $\log(\sqrt{L})$, with the power factor z then given as the slope of the log–log plot [45,74]. Based on this approach, it is possible to directly determine the individual components of J described by Equation 9.20. We note that this analysis is typically undertaken at threshold, which is straightforward to identify due to the pinning of the carrier density n, and hence the integrated SE L, in QW lasers. This procedure provides a useful method by which to compare the performance of different laser structures at a given wavelength, since it enables the relative importance of different loss mechanisms to be compared and contrasted for different material systems and laser structures [45]. Full details of this approach to quantifying and analyzing loss mechanisms in semiconductor lasers can be found in Refs. [45] and [74].

Using the temperature-dependent components of the threshold current density J_{th} extracted in this manner, it is possible to determine the recombination coefficients A, B, and C at fixed temperature for a given laser structure using the calculated threshold carrier density n_{th} in Equation 9.20. Using $J_{\mathrm{mono}} = eAn_{\mathrm{th}}$ allows A to be determined straightforwardly at fixed temperature, with B and C determined similarly using $J_{\mathrm{rad}} = eBn_{\mathrm{th}}^2$ and $J_{\mathrm{Auger}} = eCn_{\mathrm{th}}^3$. Repeating this procedure across a range of temperatures enables direct determination of the temperature-dependent recombination coefficients, which can be compared directly to theoretical calculations. This provides quantitative insight into the loss mechanisms present in the laser, as well as their relative importance and dependence on temperature. Comparison of the measured recombination coefficients to theoretical calculations then enables further analysis and interpretation on the basis of the calculated material band structure providing, in the case of dilute nitride alloys, valuable insight into the impact of N incorporation on carrier recombination. An analysis of this type has been undertaken by Fehse et al. in Ref. [45], for a GaAs-based laser structure containing a single $\mathrm{In_{0.36}Ga_{0.64}N_{0.017}As_{0.983}/GaAs}$ QW (cf. Section 9.3.2). We summarize the results of this analysis here and highlight the ability of the theoretical model outlined in Section 9.3.1 to quantitatively describe trends in real dilute nitride laser devices.

The closed circles in Figure 9.7a show the values of the monomolecular recombination coefficient A extracted from temperature-dependent measurements of current density J_{mono} [45]. However, A cannot readily be computed within the framework of the theoretical models outlined in this chapter—since it is in practice primarily associated with the material quality—making a direct comparison between theory and experiment difficult. Nonetheless, a comparison can be made based on the following simple model: we assume that the defect-related recombination occurs when free carriers having (thermal) velocity v are captured in the InGaNAs material by defects of uniform cross-section σ at a rate $A = \tau_{\mathrm{mono}}^{-1}$. Assuming there are n_{D} such defects per unit volume, the monomolecular recombination rate is given as $\tau_{\mathrm{mono}}^{-1} = \sigma v n_{\mathrm{D}}$. The dependence of the free carrier velocity v on temperature is given by the relationship between the kinetic and thermal energy as $\frac{1}{2}mv^2 \propto k_{\mathrm{B}}T$, so that $v \propto \sqrt{T}$. Using this relationship and assuming that σ is independent of T, we conclude that $A = \sigma v n_{\mathrm{D}} \propto \sqrt{T}$. The solid line in Figure 9.7a shows the variation of A with temperature predicted by this simple model, normalized to the measured value of A at $T = 150$ K—that is, $A(T) = A(150\ \mathrm{K})\sqrt{\frac{T}{150\ \mathrm{K}}}$ for $T \geq 150$ K. This simple description of the temperature dependence of A is in good agreement with experiment up to room temperature.

On the other hand, the radiative and Auger recombination coefficients B and C can be computed explicitly from the QW band structure and eigenstates, and therefore compared directly to experiment, allowing the recombination pathways in a given device to be analyzed quantitatively. This provides a stringent test of theoretical models of the electronic and optical properties. The calculation of B is straightforward and consists at each temperature of integrating over the calculated SE spectrum at threshold with respect to the photon energy in order to determine the radiative current density at threshold $J_{\mathrm{rad}}^{\mathrm{th}} = eBn_{\mathrm{th}}^2$, and hence B. Figure 9.7b compares the measured (closed circles) and calculated (open squares) values of B, for temperatures between 150 and 300 K. The calculated values of B are in good overall agreement with experiment: the theoretical model describes well the magnitude of B at low temperature, and the calculated variation of



(a)

(b)

FIGURE 9.7 (a) Measured (at threshold; closed circles) variation of the monomolecular (defect-related) recombination coefficient A at threshold as a function of temperature T for an $In_{0.36}Ga_{0.64}N_{0.017}As_{0.983}/GaAs$ ($x = 1.7\%$) 1.3-μm single QW laser structure. The solid black line shows the calculated $A(T)$ based on the simple model described in the text, with the value of A in the theoretical calculation normalized to the measured value at $T = 150$ K. (b) Measured (at threshold; closed circles) and calculated (open squares) variation of the radiative recombination coefficient B as a function of temperature T for the same device as in (a). The calculated values of B—obtained in the Boltzmann approximation, that is, assuming $J_{\text{rad}}^{\text{th}} = eBn_{\text{th}}^2$, where $J_{\text{rad}}^{\text{th}}$ is calculated directly by integrating over the calculated temperature-dependent SE spectrum at threshold for the device—display the $B(T) \propto T^{-1}$ temperature dependence characteristic of an ideal QW (Adapted from R. Fehse et al., *IEEE J. Sel. Topics Quantum Electron.*, 8:801, 2002).

B with temperature agrees well with experiment. We further note that the calculated values of B between 150 and 350 K follow the $B(T) \propto T^{-1}$ temperature dependence expected for an ideal QW [75].

The calculation of the Auger recombination coefficient C is significantly more complicated due to the large number of matrix elements describing three-body (Coulombic) scattering processes that must be evaluated and integrated over in order to determine the Auger current density $J_{\text{Auger}} = eCn^3$. Full details of the calculation of C are beyond the scope of this chapter. We refer the reader to Ref. [72] and references therein for more details on various approaches to the calculation of Auger recombination rates in quantum-confined heterostructures. Figure 9.8a compares the measured (closed circles) and calculated (open squares) values of C, for temperatures between 200 and 300 K. Given the well-known challenges associated with quantitative calculations of Auger recombination rates—the values of which are known to vary by at least an order of magnitude based on the approximations employed in the calculation—we conclude that the calculated values of C can be considered to be in good agreement with experiment. Further analysis of C indicates that the dominant Auger recombination mechanism, accounting for the majority of the calculated value of C, is the so-called CHSH pathway. The CHSH Auger recombination mechanism occurs when an electron in the CB recombines with a hole in the VB, leading to the excitation of a hole from the VB edge to the SO band (cf. Chapter 10). The "hot" SO hole generated by this process dissipates its excess energy and crystal momentum primarily via scattering within the lattice, which increases the lattice temperature and generates significant waste heat. The other notable Auger recombination mechanism is the so-called CHCC pathway, in which an electron in the CB recombining with a hole in the VB leads to the excitation of an electron to a high-energy state in the CB. Given the strong impact of N on the CB structure, it could be expected that N incorporation has significant impact on the Auger recombination rates. However, calculations of C for the InGaNAs/GaAs QW considered in Section 9.3.2 demonstrate that incorporation of 1.7% N leads to a decrease of only ~ 5% in the total Auger recombination rate compared to that in an equivalent N-free QW [72]. This indicates that the nature of Auger recombination in

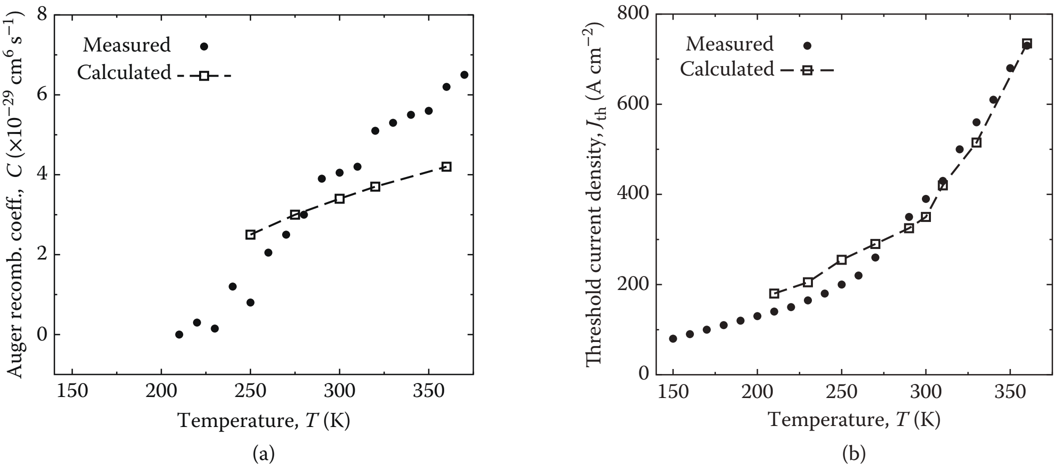


FIGURE 9.8 (a) Measured (at threshold; closed circles) and calculated (open squares) variation of the Auger recombination coefficient C as a function of temperature T for the same device as in Figures 9.7a and 9.7b. Despite the difficulties associated with direct calculations of the Auger current density, we note that the measured and calculated values of C are in reasonably good quantitative agreement. (b) Measured (closed circles) and calculated (open squares) threshold current density J_{th} as a function of temperature for the same device as in (a) and Figures 9.7a and 9.7b. J_{th} was computed at each temperature using the calculated values of A, B, and C from Figure 9.7a, 9.7b and 9.8a. We note good, quantitative agreement between theory and experiment across a broad range of temperatures from 200 to 360 K (Adapted from R. Fehse et al., *IEEE J. Sel. Topics Quantum Electron.*, 8, 801, 2002).

GaAs-based InGaNAs laser structures is largely the same as that in conventional InP-based InGaAsP and AlInGaAs structures. We note that typical values of C in InGaAsP/InP 1.3-μm QWs are in the range of 1–8 $\times10^{29}$ cm^6 s^{-1} [76,77], which are broadly similar to those calculated by Andreev and O'Reilly for a 1.3-μm InGaNAs/GaAs laser structure [72].

Figure 9.8b compares the measured (closed circles) and calculated (open squares) variation of the threshold current density J_{th} with temperature for this $In_{0.36}Ga_{0.64}N_{0.017}As_{0.983}$/GaAs single QW laser structure [45] where, overall, we note good agreement between theory and experiment. Based on the measured and calculated current densities and recombination coefficients, we conclude that defect-related, radiative, and Auger recombination account, respectively, for approximately 55%, 20%, and 25% of J_{th} at room temperature. This suggests that defect-related recombination, as opposed to carrier leakage or Auger recombination, is primarily responsible for the observed large magnitude of J_{th}, as well as the measured variation of J_{th} with temperature.

The strong role played by defect-related recombination in 1.3-μm InGaNAs QW lasers has been confirmed by a systematic study of a range of devices, where experimental measurements indicate that devices having higher J_{th} tend to have higher values of the characteristic temperature† T_0 [79]. As described above, the contribution of Auger recombination to J_{th} in InGaNAs devices is broadly similar to that in conventional InGaAsP lasers at the same wavelength. Furthermore, the relatively weak temperature dependence of $J^{\mathrm{th}}_{\mathrm{mono}}$ suggests that the large values of J_{th} measured in devices having high measured T_0 arise due primarily due to a dominant contribution to J_{th} from defect-related recombination. This strong defect-related recombination is primarily associated with the impact of N incorporation, as well as a degradation in material quality brought about by the nonequilibrium growth conditions required to facilitate N incorporation in the laser active region. We note that similar trends have been observed in 1.3-μm metamorphic (Al)InGaAs QW lasers grown on relaxed InGaAs buffer layers, in which case the strong defect-related contribution to J_{th} is associated with the formation of threading dislocations [9].

---

† The threshold current density J_{th} of a semiconductor laser at two temperatures T_1 and T_2 can typically be related in the simple form $J_{th}(T_2) = J_{th}(T_1)\exp(\frac{T_2-T_1}{T_0})$. Assuming that T_2 is greater (less) than T_1, it follows from this relation that changing the temperature by an amount T_0 leads to an increase (decrease) of J_{th} by a factor of e. In this manner, the "characteristic" temperature T_0 describes the temperature dependence of J_{th}, with a large (small) value of T_0 indicating that a large (small) change in temperature is required to bring about a given change in J_{th}. T_0 therefore describes the temperature stability of a given laser device, with a large value of T_0 generally indicating temperature stable operation. For example, InP-based 1.55-μm QW lasers—in which J_{th} is dominated by strongly temperature-dependent Auger recombination—have low characteristic temperatures $T_0 \sim \frac{T}{4}$ [78].

Overall, while N incorporation can be used to successfully realize 1.3-μm emission on a GaAs substrate, we conclude that (1) the performance of these devices suffers due to the prevalence of defect-related recombination, and (2) Auger recombination plays an important role in determining the high-temperature performance, as in conventional InP-based devices. The theoretical approach outlined in this and the previous section has provided detailed insight into the impact of N incorporation on the properties and performance of InGaNAs QW lasers and has enabled quantitative analysis of the recombination pathways in dilute nitride alloys and laser structures. This confirms the potential of theory and simulation for use in the design and optimization of dilute nitride semiconductor lasers, both at 1.3-μm and at longer wavelengths. The above analysis also highlights that the main issues to be overcome for the realization of high-performance dilute nitride semiconductor lasers are related to material growth, which must be refined in order to minimize the impact of defect-related losses. Despite these issues, there has been significant success achieved in the development of 1.3-μm InGaNAs/GaAs QW lasers, including realization of low-threshold [80,81] and high-speed [82,83] devices, whose properties and performance are competitive with existing InP-based technologies.

9.4 Alternative Applications of Dilute Nitride Alloys

So far in this chapter, we have focused primarily on the development of InGaNAs alloys to realize GaAs-based semiconductor lasers emitting at the 1.3-μm optical communications wavelength. While the development of 1.3-μm QW lasers on GaAs is the application which attracted a significant share of the initial research effort on dilute nitride alloys, the possibilities brought about by the impact of N incorporation on the band structure of a range of III–V compounds and alloys have stimulated a broader program of research to address related and alternative applications. In this section, we provide an overview of applications of dilute nitride alloys beyond 1.3-μm semiconductor lasers. The applications for which dilute nitride alloys are under investigation are diverse, spanning from the use of InGaNAs(Sb) alloys to develop 1.55-μm semiconductor lasers (Section 9.4.1), through to applications including 1.3- and 1.55-μm SOAs, EAMs, and APDs (Section 9.4.2), and monolithic integration of III–V lasers on Si (Section 9.4.3). An important focus of research interest has been the use of dilute nitride alloys to facilitate the development of 1.3- and 1.55-μm vertical-cavity devices on GaAs, including VCSELs, vertical-external-cavity surface-emitting lasers (VECSELs), and vertical-cavity semiconductor optical amplifiers (VCSOAs) [84]. Such devices have a number of key advantages over conventional InP-based edge-emitting/absorbing devices, including lower fabrication costs, power consumption and noise, as well as enhanced temperature stability. Vertical-cavity devices also open up a range of potential applications in, e.g. on-chip and chip-to-chip optoelectronics, particularly in key components such as optical switches and interconnects. The range of applications described in this section is not exhaustive—for example, there have been significant activity and much success in developing dilute nitride multijunction solar cells and VECSELs, which we do not discuss in detail here—but should give the reader an appreciation of the significant potential to exploit the unique properties of dilute nitride alloys to deliver photonic and photovoltaic devices with enhanced performance and capabilities.

9.4.1 GaAs-Based 1.55-μm Quantum Well Lasers

As described in Section 9.1, a major goal for the III–V semiconductor community is to develop reliable 1.3- and 1.55-μm semiconductor lasers on GaAs substrates to take advantage of the enhanced material properties offered by the GaAs platform (compared to InP), and to exploit technological advantages (such as the ability to utilize vertical-cavity architectures and to monolithically integrate photonic components with high-speed microelectronics). While there has been much success in developing GaAs-based 1.3-μm edge-emitting lasers and VCSELs using InGaNAs alloys, in general it has not been possible to readily and reproducibly grow InGaNAs-based devices having emission wavelengths close to 1.55-μm. This limitation of the InGaNAs material system is primarily associated with the large In and N compositions (∼40% and 4%, respectively) required to extend the emission wavelength beyond 1.5 μm. The existence of a narrow window of suitable growth conditions for InGaNAs alloys with emission wavelengths ∼1.55 μm means that the growth of InGaNAs alloys with increased In and N compositions is generally associated with significant reductions in material quality and optical efficiency. While some success has been had in developing 1.5-μm InGaNAs broad area lasers [85], there has been less success in developing InGaNAs VCSELs emitting at > 1.5-μm. As such, InGaNAs alloys grown on GaAs are commonly considered as not well suited for the development of 1.55-μm lasers on GaAs.

These limitations of InGaNAs/GaAs alloys have motivated interest in related materials, in order to extend the wavelength range accessible on GaAs beyond 1.3 μm. The approach which has primarily been pursued is to extend the emission wavelength by co-alloying N and Sb in InGaAs to form the InGaNAsSb alloy. During epitaxial growth of this penternary alloy Sb acts as a reactive surfactant enabling N incorporation at higher growth temperatures, and ultimately leading to materials with improved morphology compared to equivalent Sb-free alloys [86]. InGaNAsSb QWs are generally found to have improved material quality and optical efficiency compared to InGaNAs, and have the additional benefits that (1) the window of optimal growth parameters is wider than in InGaNAs, and (2) there is generally a simplification of the alloy growth by which the N composition (and hence emission wavelength) is found to be controllable directly via the total growth rate [87]. As a result, the use of InGaNAsSb alloys to develop GaAs-based 1.55-μm edge-emitting lasers and VCSELs—a review of which can be found in Ref. [87]—has generally been quite successful. For example, Korpijärvi et al. [88] recently demonstrated the first 1.55-μm VECSEL, in which the active region consisted of InGaNAsSb QWs and the laser structure was grown on a GaAs substrate. Recent developments such as this highlight the continuing strong interest in and potential of dilute nitride alloys to facilitate novel long-wavelength devices on GaAs.

One significant benefit of co-alloying N and Sb to form InGaNAsSb is that, contrary to trends in InGaNAs alloys, there is little degradation of the optical efficiency with increasing N composition [87]. As a result, it has been possible to develop InGaNAsSb QW lasers operating across a broad range of wavelengths, from 1.45 to 1.64 μm, which display low J_{th}. For example, a value $J_{\text{th}} = 373$ A cm^{-2} was measured for a 1.55-μm edge-emitting device under continuous wave operation at room temperature, which is competitive with the lowest values observed for 1.3-μm InGaNAs devices [80,81,87]. The fact that such low threshold current densities have been achieved in InGaNAsSb devices—despite the impact of (1) significantly higher N compositions than in 1.3-μm InGaNAs devices, and (2) Auger recombination, which is known to increase strongly with increasing wavelength [89]—highlights the benefits of co-alloying N and Sb to extend the wavelength range accessible on GaAs to 1.55 μm and beyond.

However, despite the largely successful development of InGaNAsSb 1.55-μm edge-emitting lasers and VCSELs on GaAs, there are three intrinsic issues associated with this approach that limit the device performance. First, in order to compensate the relatively large compressive strain brought about by co-alloying In and Sb, InGaNAsSb QW laser structures are typically strain compensated by growing the QW(s) between tensile strained GaNAs barrier layers. While this approach has been successful in terms of facilitating the growth of QWs displaying high optical efficiencies, it comes at the cost of associated effects that impact negatively upon the device performance. For example, the tensile strain present in the GaNAs barrier layers pushes the barrier LH band edge upward in energy, leading to a reduction in the VB offset seen by holes in the QW. As such, thermal leakage of holes limits the high-temperature performance of InGaNAsSb QW lasers [90]. Second, while Sb incorporation tends to decrease the defect density in the QW, in a strain-compensated InGaNAsSb QW structure there remains a significant density of N-related defects in the tensile strained GaNAs barriers. Carriers which leak from the QWs then tend to undergo defect-related recombination in the GaNAs layers of the structure. It is possible that these first two issues could be mitigated by, for example, using GaAsP as the barrier material, in order to provide strain compensation without (1) degrading the carrier confinement, or (2) introducing a high density of monomolecular recombination centers [90]. Third, temperature-dependent analysis of the power factor z (cf. Section 9.3.3) suggests that Auger recombination is an important loss mechanism in 1.55-μm InGaNAsSb QW lasers and plays a dominant role in limiting the high-temperature performance, as in conventional InP-based devices [87].

Recalling the analysis of 1.3-μm InGaNAs lasers in Section 9.3.3, we note that the impact of Auger recombination and IVBA is determined largely by the nature of the active region band structure, with the CHSH mechanism being dominant in materials where the magnitude of VB spin-orbit-splitting energy is less than the band gap [3]. The nature of Auger and IVBA losses in semiconductor lasers operating in the 1–2 μm wavelength range is known to be largely independent of whether the device is grown on GaAs or InP, but instead depends upon the details of the VB structure, which are broadly similar both in conventional InGaAsP or AlInGaAs materials and in highly mismatched InGaNAs(Sb) alloys. On this basis, it is not surprising that the performance of InGaNAsSb QW lasers is degraded and limited by Auger recombination. Moreover, the presence of strong Auger recombination in these devices is reflective of the fact that while developing 1.3- and 1.55-μm semiconductor lasers on GaAs substrates is desirable from a technological point of view (due to the significant benefits offered by the GaAs material platform), many of the approaches which have been used to reach this remain encumbered by fundamental issues that need to be addressed in order to overcome the limitations of existing 1- to 2-μm semiconductor laser technologies. These issues, which are associated primarily with Auger recombination and IVBA, account for the low wall-plug efficiencies and poor high-temperature performance of existing GaAs- and InP-based devices. Ultimately, addressing these issues requires a detailed consideration of the band structure of the materials forming the active region. Such considerations have led to a recent surge of interest in dilute bismide alloys—containing bismuth (Bi)—and in particular the GaBiAs alloy, which offers the potential to engineer the band structure to directly mitigate issues associated with Auger recombination and IVBA in GaAs-based 1.55-μm laser structures. An introduction to the theory and simulation of dilute bismide alloys, a class of highly mismatched III–V semiconductor alloys which can, in several ways, be considered naturally complementary to dilute nitride alloys, is provided in Chapter 10.

9.4.2 Semiconductor Optical Amplifiers, Electro-Absorption Modulators, and Avalanche Photodiodes

Based on the strong interest in dilute nitride alloys for applications in semiconductor lasers, several of the measured properties of InGaNAs(Sb) devices have suggested that there is significant potential to use dilute nitrides to deliver new capabilities and improved performance in related devices such as 1.3- and 1.55-μm SOAs, EAMs, and APDs. In this section, we briefly review research on using dilute nitride alloys to develop each of these classes of devices, highlighting the principles motivating this research in terms of the potential benefits of dilute nitride alloys, and summarizing the progress that has been made toward developing devices having improved characteristics compared to conventional InP-based technologies.

In addition to the benefits of developing 1.3- and 1.55-μm optoelectronic devices on GaAs as opposed to InP, for applications in SOAs and EAMs one of the most attractive properties of InGaNAs(Sb) alloys and heterostructures is their reduced sensitivity to temperature. Incorporation of N, which acts as an isovalent impurity, is known to lead to a reduction in the temperature sensitivity of the band gap of bulk materials [91] and QWs [92] compared to that of N-free materials. This reduced temperature dependence of the band gap has its origins in the hybridization of the extended states of the host matrix CB edge with N-related localized states (cf. Section 9.2)—the properties of the latter are largely independent of temperature, so that N-induced hybridization leads to InGaNAs(Sb) alloy CB edge states containing an admixture of N-related localized character and hence having reduced sensitivity to temperature [93]. This property is particularly appealing for the development of SOAs and EAMs: the temperature dependence of the operation wavelength in conventional InP-based devices, as well as the degree of degradation in performance with increasing temperature, mandates the use of external cooling equipment to maintain operational stability which dramatically increases the power consumption of the overall system. The large CB offsets present in dilute nitride QWs compared to those in conventional InGaAsP/InP or AlInGaAs/InP structures lead to reduced thermal leakage of electrons, which can be expected to contribute further to improving high-temperature device performance. The ability to (1) prevent degradation of the optical absorption/amplification at high temperature, and (2) reduce temperature-dependent drift of the operational wavelength make InGaNAs(Sb) structures ideal candidates for the development of GaAs-based 1.3- and 1.55-μm SOAs and EAMs. Such devices additionally offer the potential to achieve uncooled operation, and hence, bring about significant energy savings.

From the perspective of SOA development, further strong motivations for the exploration of dilute nitride alloys include the presence of broad-band optical gain and a significant degree of polarization tunability in InGaNAs(Sb) QWs, both of which can be expected to lead to SOAs displaying enhanced performance [94,95]. These potential benefits of InGaNAs(Sb) alloys, as well as the possibility to grow long-wavelength SOAs on GaAs, have stimulated the development of 1.3-μm InGaNAs SOAs and VCSOAs, as well as 1.55-μm InGaNAsSb SOAs. Initial investigations by Hashimoto et al. [96] on a 1.3-μm traveling-wave InGaNAs/GaAs SOA confirmed the potential of dilute nitride SOAs, by demonstrating significantly enhanced temperature stability compared to InP-based devices. Further studies have developed this device concept further, leading to the demonstration of multiwavelength amplification at 1.3 μm [97], as well as high optical gain and favorable gain dynamics. Following the confirmation of the reduced sensitivity of dilute nitride SOAs to temperature, uncooled operation of a 1.3-μm device at 10 Gbit s^{-1} has been demonstrated [98]. Continued refinement of InGaNAs SOA design and fabrication has since led to further improvements in performance. Recently, Fitsios et al. [99,100] demonstrated error-free operation at 10 Gbit s^{-1} of a 1.3-μm InGaNAs/GaAs SOA displaying high gain (∼28 dB), fast gain recovery (∼100 ps), and temperature-stable gain (between 20°C and 50°C). The high performance of this device, combined with its reduced cost compared to incumbent 1.3-μm InP-based SOAs, suggests that existing dilute nitride SOAs are already suitable for deployment in, e.g. metro and optical access networks.

Substantial progress has also been made in the development of 1.3-μm InGaNAs/GaAs VCSOAs, from initial theoretical analysis [101] through to investigations of optically pumped devices [102–104] and electrically pumped “HELLISH” devices [105,106], to the recent realization of room temperature gain in an electrically pumped device [107]. These advances clearly illustrate the progress that has been made in developing the materials growth and device engineering of GaAs-based dilute nitride SOAs. The high performance of these devices and potential for on-chip integration further demonstrate that dilute nitride VCSOAs have significant potential for deployment in applications such as optical switches and interconnects. While the main focus of the research effort on GaAs-based dilute nitride SOAs has centered on 1.3-μm devices incorporating InGaNAs QWs, in line with research on dilute nitride QW lasers, there has also been some effort dedicated to using InGaNAsSb QWs to facilitate the development of 1.55-μm GaAs-based SOAs. Dilute nitride SOAs operating at 1.55-μm have also displayed fast gain recovery [108,109] and high temperature stability [109,110]. While these 1.55-μm devices are not as developed as their 1.3-μm counterparts, existing results at 1.55 μm highlight that there is scope for further development, and hence significant potential to develop high-performance GaAs-based (VC)SOAs in this technologically important wavelength range.

By comparison to the ongoing research effort on SOAs, there has been very little development of dilute nitride EAMs. Initial electric-field-dependent measurements of the optical absorption both in 1.3- and 1.55-μm InGaNAs(Sb) QW structures have demonstrated a strong quantum-confined Stark effect (QCSE) [111], suggesting that there is broad scope for the development of GaAs-based EAMs at optical communication wavelengths. Detailed theoretical analysis of a prototypical 1.3-μm InGaNAs structure suggests that dilute nitride EAMs operating at 1.3 μm should offer extinction ratios slightly improved over those expected for N-free devices [112]. These favorable characteristics result primarily from the CB offsets in InGaNAs(Sb) QWs, which are substantially larger than those available in the QWs employed in conventional InP-based EAMs. These large band offsets offer improved carrier confinement, and hence an enhancement of the QCSE. On this basis, reduced thermal leakage of electrons from dilute nitride QWs at higher temperatures is expected to lead to significant temperature stability of the QCSE in such structures. This has been confirmed by experimental measurements undertaken on a series of multi-QW 1.3-μm InGaNAs/GaAs EAMs, in which no appreciable degradation in the QCSE was measured between 25°C and 100°C [113]. These measurements additionally confirmed improved overall high-temperature performance of dilute nitride EAMs compared to InP-based devices, including the expected reduced temperature dependence of the absorption edge wavelength on temperature. Overall, initial studies suggest that dilute nitride alloys are a promising material system for the development of GaAs-based EAMs and, in addition to the benefits associated with growth on GaAs, dilute nitride EAMs should offer high-temperature performance which exceeds that of existing 1.3- and 1.55-μm InP-based devices.

Additional research interest in InGaNAs alloys has been motivated by their potential for applications in APDs. Of key importance in determining the performance of an APD are the electron and hole impact ionization rates[†] α and β. In Ref. [114], Adams identified the potential to use InGaNAs alloys to overcome the limitations associated with the nature of impact ionization in conventional III–V semiconductors by providing a low-noise medium for hole multiplication in APDs. As we discuss in detail in Section 9.5, the strong impact of N incorporation on the CB structure leads to a significant degradation of the electron mobility in InGaNAs alloys. Adams proposed that this N-induced limitation of high-field electron transport, combined with the fact that N incorporation leaves the VB relatively unchanged, should produce a bottleneck for electrons by limiting their ability to accelerate to sufficiently high energies to initiate impact ionization. Hence, incorporation of N can be expected to dramatically reduce α while leaving β relatively unchanged, and thereby leading to low-noise hole multiplication in a device having $k \gg 1$. Such devices are also expected to have reduced band-to-band tuneling due in part to the tendency of N incorporation to introduce intervalley mixing of CB states, which should lead to improved performance in long-wavelength devices [114].

While a favorable ratio $k \ll 1$ of the impact ionization rates can be obtained in Si-based APDs, the operation of these low-noise electron multiplying devices is limited to wavelengths <1.1 μm by the magnitude of the Si band gap. The reduction in band gap brought about by N incorporation in InGaNAs alloys then further adds the potential to realize low-noise APDs operating at 1.3 and 1.55 μm. This has the potential to overcome key limitations associated with existing InP-based technologies in this wavelength range (which must typically incorporate separate absorbing and carrier multiplying regions) by offering improved performance in devices having a single gain medium which offers efficient long-wavelength absorption and low-noise hole multiplication. When combined with the benefits of fabrication on GaAs substrates, including robust fabrication and reduced cost, the development of GaAs-based 1.3- and 1.55-μm APDs represents a promising opportunity to overcome fundamental issues associated with the InP-based devices deployed in existing optical communication networks. However, initial experimental analysis of the impact ionization rates in a series of InGaNAs diode structures has demonstrated that while N incorporation does cause a decrease in α leading to an increase of k by a factor of approximately four, this is accompanied by a reduction of both α and β by close to two orders of magnitude [115]. This strong suppression of the impact ionization rates resulting from N incorporation suggests that dilute nitride alloys may not be suitable for use as a multiplication medium in telecom-wavelength APDs, but may instead have potential for deployment in, for example, GaAs-based heterojunction bipolar transistors. Further investigations are required to ascertain whether or not this is the case, and to definitively demonstrate or rule out the promise of dilute nitrides for applications in long-wavelength APDs.

[†] The nature of avalanche multiplication in an APD depends on the relative values of the electron and hole impact ionization rates α and β, characterized by the ratio $k = \frac{\beta}{\alpha}$. For a low-noise APD it is important that α and β differ significantly in magnitude in order to allow for large, stable electrical gain. Otherwise, in the case where $k \approx 1$—that is, where the electron and hole multiplication rates are roughly equal—the impact ionization process leads to noisy and uncontrollable avalanche breakdown. The latter of these two situations is the case in most III–V semiconductors, in which α and β typically differ by no more than a factor of approximately two in magnitude.

9.4.3 Monolithic Integration of III–V Lasers on Silicon

The ongoing exponential growth of the internet, mobile/cloud computing, and "big data" has created strong demand for new hardware architectures to deliver both increased bandwidth and decreased energy consumption. Key to the continued development of data processing in computing and communications is the need to prevent the formation of bottlenecks by increasing the available bandwidth from chip to server level. Electrical data transmission reaches practical limits at a bandwidth-distance product of approximately 100 Gb m s^{-1} due to the intrinsic limitations of electrical interconnects. A promising approach to overcome these issues is direct integration of optoelectronic devices on silicon (Si) to enable the development of high bandwidth optical interconnects for on-chip and chip-to-chip data transfer. This photonics-based approach has significant advantages over traditional electronics: optical data transfer has been heavily developed for optical communications, leading to interconnects offering significantly improved bandwidth compared to their electronic counterparts, over distances ranging from as low as several meters to long-distance links of the order of thousands of kilometers in length. In addition to offering higher bandwidth, optoelectronic integration on Si has the additional benefit that optical interconnects do not suffer from the losses due to resistive heating that play a significant role in limiting the performance of electrical interconnects.

Compressively strained SiGe/Si alloys and heterostructures have been used to develop many of the components required to realize practical integrated optoelectronic architectures, including photodetectors, filters, waveguides, and switches. However, Si-compatible group-IV materials and heterostructures are generally limited by their inability to efficiently emit/absorb photons—due to their indirect band gaps—meaning that while some progress has been made there are, as yet, no efficient, electrically pumped Si-compatible light sources available to facilitate the widespread adoption of Si-based optoelectronic integrated circuits. Given these limitations, as well as the comparably high optical emission efficiencies offered by direct-gap III–V materials, there has been a growth of interest in integrating III–V optoelectronic devices on Si to deliver monolithically integrated emitters and absorbers. A review of the complete range of approaches being investigated to integrate III–V light sources on Si is beyond the scope of this chapter—we focus here on the potential of dilute nitride alloys to realize electrically pumped lasers sources on Si and refer the reader to Ref. [116] for a recent overview of alternative approaches.

The primary requirements for Si-compatible III–V materials are that the III–V material has (1) a lattice constant close to that of Si, to facilitate the growth of high-quality epitaxial layers and heterostructures, and (2) a direct band gap, to facilitate efficient optical emission/absorption. Of the III–V semiconductor compounds, GaP has the closest lattice constant to Si. This suggests that GaP offers a suitable starting platform for the development of Si-compatible optoelectronic devices. However, GaP also has an indirect band gap, placing severe limitations on its potential for the development of efficient light-emitting devices. Incorporation of a large fraction of As in GaAsP alloys is required to bring about a direct band gap, but this then leads to a large increase in lattice constant which limits the ability to grow direct-gap GaAsP alloys directly on Si. These issues can be overcome by the incorporation of N, which simultaneously lowers the lattice constant and the energy of the Γ-point CB edge. This therefore raises the possibility of using N incorporation to produce direct-gap GaNAsP alloys which can be grown directly on Si with acceptable levels of compressive strain. Analysis of the GaNAsP band structure suggests that As compositions >70% and N compositions ~4% are required to produce suitable direct-gap alloys which are compatible with growth on Si substrates [117].

Direct growth of GaNAsP alloys on Si is a challenging task for epitaxial growth. For example, in addition to the difficulties associated with the growth of high-quality dilute nitride alloys, the mismatch in thermal expansion coefficients between Si and GaNAsP can lead to the development of strain-induced dislocations, and even cracking, during epitaxial growth. Nonetheless, significant progress has been made in developing GaNAsP QW lasers on Si, including the development of novel boron (B) containing alloys to serve as suitable strain compensating barrier layers and mitigate the effects of the aforementioned thermal expansion issues during growth [118]. These developments led to the demonstration in 2011 of the first monolithically integrated III–V laser on Si [119]. The initial single QW device of Ref. [119] demonstrated electrically pumped lasing at a wavelength of 861 nm up to $T = 165$ K, with a threshold current density $J_{\text{th}} \sim 1.6$ kA cm^{-2}. Detailed experimental analysis, similar to that described in Section 9.3.3, highlights that GaNAsP/Si lasers offer broadly comparable performance to equivalent test devices grown on GaP [120,121]. These Si-based lasers were shown to suffer from increased defect-related recombination at low temperature compared to the GaP-based devices [116]. However, taking into account the fact that (1) electrically pumped lasing has already been demonstrated using this novel class of semiconductor lasers, and (2) that significant development that has occurred in the epitaxial growth of related dilute nitride alloys and heterostructures suggests that direct-gap GaNAsP laser structures grown directly on Si represent a promising approach to the development of monolithically integrated Si-based light sources suitable for applications in optoelectronic integration. Research on this novel device concept is ongoing, with the aim of realizing room temperature electrically pumped lasing from a monolithically integrated III–V device on Si.

9.5 Theory of the Density of States and Carrier Transport in Dilute Nitride Alloys

Having focused so far in this chapter on the electronic and optical properties of dilute nitride alloys, we turn our attention in this section to the impact of N incorporation on the carrier transport in GaN$_x$As$_{1-x}$. Since N incorporation primarily affects the CB structure while leaving the VB structure virtually unchanged, transport of VB holes in GaN$_x$As$_{1-x}$ is akin to that in GaAs while N impacts extremely strongly on the transport of CB electrons. Carrier transport plays a key role in the operation in several of the classes of dilute nitride–based devices described in Section 9.4, including in particular APDs and multijunction solar cells. As such, developing a detailed understanding of the impact of N incorporation on the carrier transport in GaN$_x$As$_{1-x}$ and related alloys is crucial to facilitate the design and optimization of a range of novel optoelectronic and photovoltaic devices. Investigation of electron transport in dilute nitride alloys has proven to be particularly enlightening from a theoretical perspective, as attempts to describe, e.g., the dramatic reductions in electron mobility observed in experimental measurements, have stimulated detailed considerations of the impact of N incorporation on the CB structure.

The simple BAC model of Section 9.2.2—which proved to be extremely useful and surprisingly accurate for the description of the electronic and optical properties of InGaNAs semiconductor lasers (cf. Section 9.3)—is insufficient to analyze the details of the CB structure and DOS at energies close to N-related localized states. In particular, the unphysical nature of the CB DOS predicted by the BAC model has significant implications for calculations of the transport properties in GaN$_x$As$_{1-x}$ and related alloys. In order to lay the foundations for a detailed understanding of the carrier transport in dilute nitride alloys, we therefore begin our discussion in Section 9.5.1 by reviewing approaches to calculating the CB structure and DOS beyond the BAC model. We discuss the development of Green's function techniques to obtain solutions of the AIM, and review that a self-consistent approach to calculating the AIM Green's function can be used to produce a description of the DOS which is in quantitative agreement with the results of large-scale supercell calculations. We further outline how this approach has been combined with the LCINS description of the electronic structure (cf. Section 9.2.3) to provide significant insight into the impact of N on the nature of the CB dispersion and DOS in realistic, disordered GaN$_x$As$_{1-x}$ alloys.

In Section 9.5.2 we review the theory of electron transport in GaN$_x$As$_{1-x}$ alloys. We begin with simple analytical considerations, which demonstrate that N incorporation places intrinsic limitations on the electron mobility. This is followed by a discussion of resonant scattering of electrons by N-related localized states, highlighting the important role played by N–N pairs and larger clusters of N atoms in determining the transport properties. We review refinements in the theory of carrier transport in dilute nitride alloys—encompassing models based on the Boltzmann transport equation (BTE) as well as Monte Carlo techniques—and show how these approaches (in conjunction with the detailed description of the CB structure described in Sections 9.2 and 9.5.1) have facilitated detailed analysis of electron transport in GaN$_x$As$_{1-x}$, and have provided quantitative agreement with experimental measurements.

9.5.1 The Anderson Impurity Model: Self-Consistent Green's Function Method

That the BAC model provides an unphysical description of the CB structure and DOS can be seen clearly by considering the BAC band structure described in Section 9.2. As the wave vector k increases, the energy $E_{-}(k)$ of the lower-energy band increases asymptotically toward the N-resonant state energy E_{N} (cf. Equation 9.2). As a result, the DOS calculated using the BAC model increases strongly with increasing energy above the alloy CB edge $E_{-}(0)$, leading to a singularity in the DOS at energy $E = E_{\mathrm{N}}$. This clearly unphysical result is associated with the fact that the BAC model intrinsically miscounts the number of N-related states per unit volume in an (assumed infinite) crystal [122,123]. Another problematic feature of the BAC model from the perspective of transport calculations is that it predicts the existence of a gap in the DOS in the energy range $E_{\mathrm{N}} < E < E_{+}(0)$, even at ultra-dilute N compositions. The presence of such a gap in the DOS is counter to the results of experimental measurements carried out on $\mathrm{GaN}_x\mathrm{As}_{1-x}$ materials, which indicate that the CB DOS is continuous and gapless at energies close to E_{N} [124,125].

These features of the BAC model, which were not an issue in the calculation of the properties of semiconductor lasers (due to the fact that for analysis of the optical properties we are interested primarily in carriers lying close in energy to the CB and VB edges), have significant implications for calculations of the carrier transport in $\mathrm{GaN}_x\mathrm{As}_{1-x}$ and related alloys. For example, the singularity in the DOS at $E = E_{\mathrm{N}}$ is associated with a breakdown in **k**-selection, suggesting that carriers can acquire extremely large crystal momenta without increasing their energy. Additionally, the presence of a gap in the DOS suggests that electrons in $\mathrm{GaN}_x\mathrm{As}_{1-x}$ are not capable of accelerating out of the lower E_{-} band, even under the influence of large applied electric fields. Given these shortcomings of the BAC model, it is clear that a quantitative understanding of carrier scattering and transport, particularly under intermediate and high field conditions, must begin with a detailed analysis of the nature of the CB structure and DOS. Such an analysis has been carried out by several authors, who used Green's function techniques to solve the AIM [122,123,126–129]. We summarize these results here and demonstrate by comparison with the results of large-scale supercell diagonalizations of the AIM that an accurate description of the CB structure and DOS, particularly at energies close to N-related localized states, requires a self-consistent approach.

The AIM for dilute nitride alloys describes a set of N-related localized states $|j\rangle$ interacting with the extended CB states $|\mathbf{k}\rangle$ of the unperturbed host matrix semiconductor via the Hamiltonian $\widehat{H} = \widehat{H}_0 + \widehat{V}$. Here, $\widehat{H}_0$ is a sum of two terms, describing the energies of the extended and localized states [126,129,130]

$$\widehat{H}_0 = \sum_{\mathbf{k}} E_{\mathbf{k}} \, |\mathbf{k}\rangle\langle\mathbf{k}| + \sum_{j} E_j \, |j\rangle\langle j| \,, \tag{9.21}$$

where $E_{\mathbf{k}}$ and E_j are, respectively, the energy of the unperturbed host matrix CB at wave vector **k**, and of the N-related localized state $|j\rangle$. $\widehat{V}$ describes the interaction between the extended and localized states [126,129,130]

$$\widehat{V} = \frac{1}{\sqrt{N_c}} \sum_{\mathbf{k},j} \left(V_{\mathbf{k},j} \, e^{i\mathbf{k}\cdot\mathbf{r}_j} \, |\mathbf{k}\rangle\langle j| + V_{\mathbf{k},j}^{*} \, e^{-i\mathbf{k}\cdot\mathbf{r}_j} \, |j\rangle\langle\mathbf{k}| \right), \tag{9.22}$$

where $V_{\mathbf{k},j}$ is a matrix element describing the hybridization of the extended and localized states, $\mathbf{r}_j$ is the position of the substitutional N impurity associated with the localized state $|j\rangle$, and N_c is the total number of primitive unit cells in the system. It is generally assumed that the hybridization parameters $V_{\mathbf{k},j}$ depend only weakly on the wave vector **k**, so that we can reasonably approximate them as being independent of **k**—that is, $V_{\mathbf{k},j} \approx V_j = \frac{\beta_j}{\sqrt{N_c}}$ [126].

In the case that an N-related localized state $|j\rangle$ lies energetically within the host matrix CB, its interaction with an extended host matrix CB state $|\mathbf{k}\rangle$ results in the creation of a quasi-localized resonance having a finite lifetime [131]. Mathematically, this N-derived, quasi-localized alloy state can be described in terms of a complex-valued energy broadening ΔE_j, corresponding to an eigenstate time dependence described by the phase factor $\exp(\frac{i(E_j+\Delta E_j)t}{\hbar}) = \exp(\frac{i(E_j+\mathrm{Re}\{\Delta E_j\})t}{\hbar}) \times \exp(-\frac{\mathrm{Im}\{\Delta E_j\}t}{\hbar})$. From this, we see that the real and imaginary parts of the complex energy broadening ΔE_j respectively describe an energy shift and homogeneous broadening, with the lifetime of the N-derived quasilocalized resonance given by $\tau_j \sim \frac{\hbar}{\mathrm{Im}\{\Delta E_j\}}$ [128].

Exact extended energy eigenstates of the AIM Hamiltonian $\widehat{H}$ can be found via the poles of the Green's function [126]

$$G_{\mathbf{kk}}(E) = \left(E - E_{\mathbf{k}} - \frac{1}{N_c} \sum_j \frac{|V_j|^2}{E - E_j - \Delta E_j(E)} \right)^{-1}, \tag{9.23}$$

where the complex energy-dependent broadening associated with each localized state $|j\rangle$ is given by

$$\Delta E_j(E) = \frac{|V_j|^2}{N_c} \sum_{\mathbf{k}} G_{\mathbf{kk}}(E), \tag{9.24}$$

the real and imaginary parts of which then describe, respectively, the energy shift and broadening of the N-related localized states, brought about by their interaction with the extended host matrix CB states $|\mathbf{k}\rangle$ with which they are resonant [129]. The alloy CB DOS projected to extended (Bloch) states, $D(E)$, can be calculated directly from the imaginary part of the extended state Green's function, Equation 9.23, as [122,126]

$$D(E) = -\frac{1}{\pi} \sum_{\mathbf{k}} \mathrm{Im}\{G_{\mathbf{kk}}(E)\}. \tag{9.25}$$

In practice, Equations 9.24 and 9.25 are evaluated by converting the sum into an integral over the continuous wave vector **k** so that, e.g. the DOS is given by

$$D(E) = -\frac{1}{\pi} \mathrm{Im}\left\{ \int_0^{\infty} G_{\mathbf{kk}}(E) D_0(E_{\mathbf{k}}) \mathrm{d}E_{\mathbf{k}} \right\}, \tag{9.26}$$

where $D_0(E) = \frac{(2m_0^*)^{2/3}\sqrt{E}}{4\pi^2\hbar^3}$ is the DOS of the unperturbed host matrix CB, (with the zero of energy taken at the host matrix CB edge) [123,129].

In general, Equations 9.23 and 9.24 do not admit an analytical solution due to the energy dependence of the complex localized state broadening $\Delta E_j(E)$. For this reason, several authors have introduced approximations for $\Delta E_j(E)$ in order to produce analytically tractable solutions to Equation 9.23. Typically, this consists of assuming $\Delta E_j(E)$ to be purely complex and independent of energy: $\Delta E_j(E) \approx i\Delta_j$ so that it described uniform (energy-independent) localized state broadening. For dilute nitride alloys this approximation was originally introduced by Wu et al., who used it to solve the AIM in the coherent potential approximation for a BAC-like two-band model in which all of the N-related localized states are assigned the same energy E_{N} [122]. The real and imaginary parts of the resultant complex energy eigenstates were then interpreted, respectively, in terms of a BAC-like band dispersion and an N-induced energy broadening of the band dispersion at each wave vector **k**. This approach was further developed by Vaughan and Ridley, who derived the AIM Green's function in the form given by Equation 9.23 and extended the treatment of the band structure to a three-band model including localized states associated with N–N pairs in addition to isolated N impurities [126]. A similar approach has also been developed by Vogiatzis and Rorison, who derived and applied the Matsubara Green's function to calculate the complex energy band structure and localized state lifetimes of the AIM eigenstates for the two- and three-band models both in bulk alloys and in (In)GaN$_x$As$_{1-x}$ QW heterostructures [127,128].

As we describe below, the inclusion of a purely imaginary, energy-independent energy broadening removes the issues associated with the BAC description of the DOS: the singularity at the N-resonant state energy E_N is removed, and the gap in the DOS between E_N and $E_+(0)$ is closed. However, while the closing of the gap in the DOS brought about by taking into account complex energy broadening of the N-related localized states in this manner is qualitatively in agreement with the results of experimental measurements, Seifikar et al. showed by comparison to large-scale supercell calculations that this approximation to the AIM Green's function introduces a number of errors into the description of the CB structure and DOS at energies close to localized states [123]. In order to overcome these issues, it is necessary to obtain a direct numerical solution of Equations 9.23 and 9.24 in order to fully determine the extended state Green's function, by computing the full energy-dependent broadening $\Delta E_j(E)$. These issues were identified by Seifikar et al. [123,129], who demonstrated the need for and developed a self-consistent approach to solving Equations 9.23 and 9.24. The need for a self-consistent Green's function (SCGF) approach was originally identified by Juaho and Wilkins, who studied the general problem of resonant scattering impurities interacting with the extended eigenstates of a parabolic band, a generic version of the problem encountered in dilute nitride and related alloys [132].

The SCGF calculation proceeds as follows: first, an initial guess is generated for the broadening function $\Delta E_j(E)$ using perturbation theory [123]. Second, this broadening is used to compute the energy-dependent extended state Green's function $G_{\mathbf{kk}}(E)$ according to Equation 9.23. Third, this Green's function is used to recompute the energy-dependent broadening $\Delta E_j(E)$ according to Equation 9.24. Fourth, this new energy-dependent broadening is used to recompute $G_{\mathbf{kk}}(E)$. The third and fourth steps in this procedure are then iterated until $\Delta E_j(E)$ has converged throughout the energy range of interest. The converged complex energy broadening $\Delta E_j(E)$ can then be used to calculate the CB structure and DOS via Equations 9.23 and 9.24.

Having described the motivation for and implementation of the SCGF method, we now summarize the results of applying this approach to study the impact of N incorporation in GaN$_x$As$_{1-x}$ alloys for two models of the CB structure: (1) a BAC-like two-band model in which the N-related localized states are all assumed to have the same energy E_N, and (2) a multiband model based upon the LCINS approach outlined in Section 9.2.3, in which a full distribution of N-related localized states—associated with not only isolated N impurities, but also with N–N pairs and larger clusters of N atoms—are included in the sums in Equations 9.23 and 9.24. We note that the results we summarize here for the SCGF method correspond roughly to the strong scattering regime investigated by Juaho and Wilkins [132], highlighting that substitutional N impurities have a strongly perturbative effect on the CB structure.

We begin by analyzing the CB structure and DOS calculated using the SCGF method for a two-band model of the GaN$_x$As$_{1-x}$ CB. In the two-band model, all N-related localized states are assumed to have energy E_N and the composition dependence of the interaction matrix element is assumed to be of the form $V_j = \beta\sqrt{x}$. As a result, Equation 9.23 can be written for the two-band model as [129]

$$G_{\mathbf{kk}}(E) = \left(E - E_{\mathbf{k}} - \frac{\beta^2 x}{E - E_N - \Delta E_N(E)} \right)^{-1}, \tag{9.27}$$

where the complex energy broadening of the N-related localized states is given by

$$\Delta E_N(E) = \beta^2 x \sum_{\mathbf{k}} G_{\mathbf{kk}}(E). \tag{9.28}$$

The results of the self-consistent calculation of $G_{\mathbf{kk}}(E)$ and $\Delta E_N(E)$ using Equations 9.27 and 9.28 are shown in Figures 9.9a and 9.9b for GaN$_{0.002}$As$_{0.998}$ ($x = 0.2\%$; solid lines). The zero of energy is taken at the CB edge of the unperturbed GaAs host matrix. The initial, energy-independent estimates for the real and imaginary parts of $\Delta E_N(E)$, generated using second-order perturbation theory as outlined in Ref. [123], are

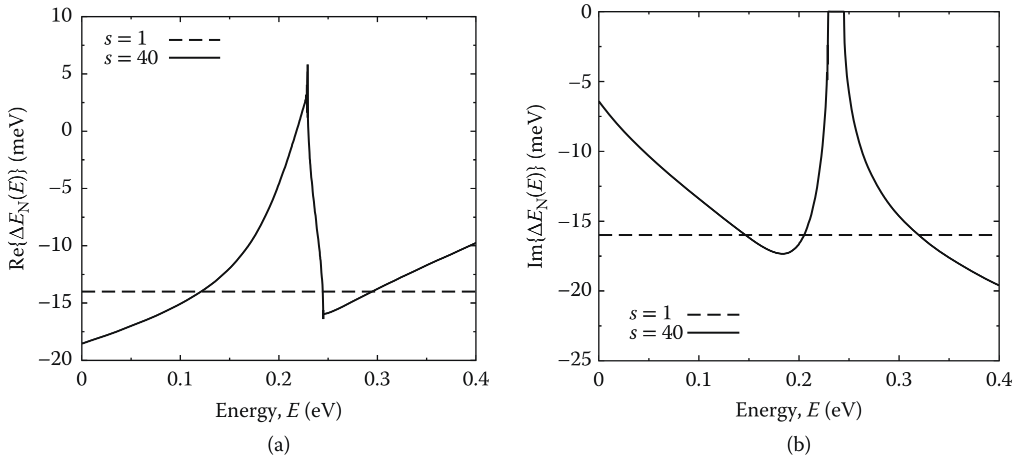
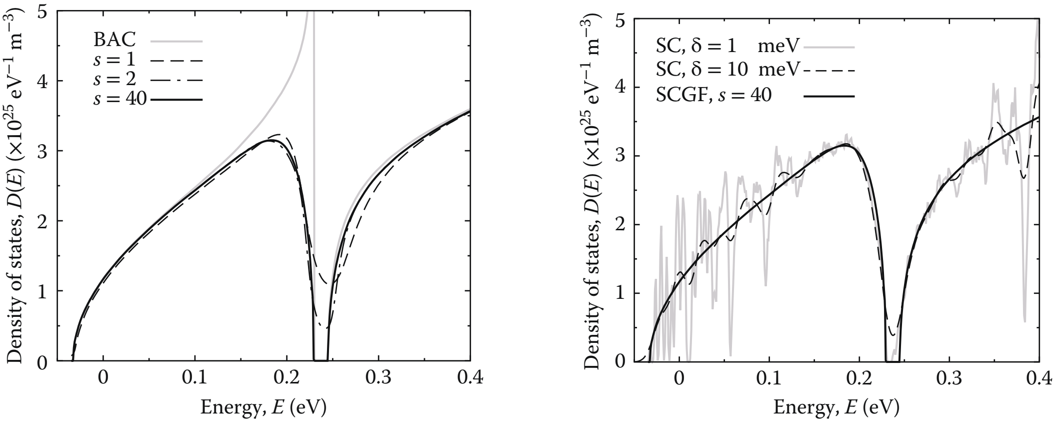


FIGURE 9.9 (a) Real and (b) imaginary parts Re{$\Delta E_N(E)$} and Im{$\Delta E_N(E)$} of the complex energy-dependent N-localized state broadening $\Delta E_N(E)$, calculated using the SCGF method for a two-band model of the $GaN_{0.002}As_{0.998}$ ($x = 0.2\%$) CB structure in which all N-related localized states are taken to have the same energy (cf. Equations 9.27 and 9.28). Dashed lines show the initial energy-independent values of Re{$\Delta E_N(E)$} and Im{$\Delta E_N(E)$} at the first iteration $s = 1$ of the SCGF calculation, computed using perturbation theory. Solid lines show the converged Re{$\Delta E_N(E)$} and Im{$\Delta E_N(E)$} after $s = 40$ iterations. (Adapted from M. Seifikar et al., *J. Phys. Condens. Matter*, 26, 365502, 2014.)

shown in each case using a dashed line. It is found that approximately 40 iterations ($s = 40$) are required to ensure that $\Delta E_N(E)$ is fully converged. However, we note that convergence is much more rapid at energies away from the N-resonant state energy E_N, with as few as two iterations required to ensure convergence. Comparing the converged real and imaginary parts of $\Delta E_N(E)$ to the constant initial values in Figure 9.9a and 9.9b highlights the strong energy dependence of the N-localized state resonant energy and broadening, which are neglected in non-self-consistent approaches. For example, the real part Re{$\Delta E_N(E)$} of $\Delta E_N(E)$ has a value of approximately −20 meV at the alloy band edge, from which it increases monotonically with increasing energy, changing sign and reaching a value of approximately +5 meV at $E = E_N$. Re{$\Delta E_N(E)$}, then decreases strongly in the energy range $E_N < E < E_+(0)$, again changing sign, before once again increasing monotonically with energy for $E > E_+(0)$. It follows from Equations 9.25 and 9.28 that the imaginary part Im{$\Delta E_N(E)$} of $\Delta E_N(E)$ is proportional to the DOS. We note that the converged self-consistent calculation indicates that Im{$\Delta E_N(E)$}, and hence the DOS, vanishes in the energy range $E_N < E < E_+(0)$. This confirms, as described above, that the DOS calculated using the SCGF method for a two-band model of the GaN_xAs_{1-x} CB contains a gap at energies just above the N-resonant state energy [129].

The opening of this gap in the DOS can be observed in Figure 9.10a, which shows the DOS calculated using the converged $\Delta E_N(E)$ of Equation 9.28 in Equation 9.27 for (1) the first three iterations of the self-consistent calculation ($s = 1$, 2, and 3), and (2) the fortieth iteration, $s = 40$, by which $\Delta E_N(E)$ has fully converged. The first iteration for the DOS (dashed black line) corresponds to the standard, nonself-consistent Green's function approach [122,126,128], with the broadening calculated using perturbation theory [123]. For reference, we have also included the DOS calculated using the BAC model (solid gray line), which contains the aforementioned singularity at the N-resonant state energy $E_N = 0.23$ eV. For $s = 1$ there is no singularity in the DOS at E_N and no gap in the DOS above E_N. As the self-consistent calculation progresses, the rapid opening up of a gap in the DOS is observed after two ($s = 2$) iterations (depicted using a dash-dotted black line). Comparing these calculations highlights the role of the energy dependence of the complex localized state broadening $\Delta E_N(E)$: the converged DOS is well described by the $s = 1$ (energy independent) broadening at energies away from E_N, while iterating to obtain self-consistency results in large changes to the calculated DOS at energies close to E_N.



(a) (b)

FIGURE 9.10 (a) DOS calculated for $GaN_{0.002}As_{0.998}$ ($x = 0.2\%$) using the SCGF method for a two-band model of the CB structure in which all N-related localized states are taken to have the same energy (cf. Equations 9.27 and 9.28). Dashed and dash-dotted black lines show the calculated DOS after $s = 1$ and 2 iterations of the SCGF calculation. The solid black line shows the converged DOS, obtained after $s = 40$ iterations of the SCGF calculation. The solid gray line shows the DOS calculated using the simple two-band BAC model of the CB structure (cf. Section 9.2.2), which is singular at the N-localized state energy $E_N = 0.23$ eV. The opening up of a gap in the DOS above E_N is observed as the SCGF calculation converges, which is not present in the standard, nonself-consistent ($s = 1$) Green's function calculation. (b) Converged DOS calculated using the SCGF model (solid black line, as in (a)), compared to the DOS calculated via a large-scale supercell (SC) diagonalization of the AIM Hamiltonian for the same two-band model of the CB structure. Solid gray and dashed black lines show the SC DOS calculated by assuming in the numerical calculations that the contribution of each $GaN_{0.002}As_{0.998}$ CB state to the DOS is described by a Gaussian lineshape having a standard deviation $\delta = 1$ and 10 meV, respectively. The zero of energy is taken in each case at the unperturbed GaAs CB edge. (Adapted from M. Seifikar et al., *J. Phys. Condens. Matter*, 26, 365502, 2014.)

In order to investigate the importance of including self-consistent energy-dependent broadening to describe the DOS close in energy to N-related localized states, Seifikar et al. performed a direct diagonalization of the AIM Hamiltonian using a large-scale supercell model [123]. Figure 9.10b compares the converged DOS calculated using the SCGF method (solid black line) to that calculated using the supercell method for $GaN_{0.002}As_{0.998}$. The DOS calculated using the supercell approach is shown for Gaussian broadenings $\delta = 1$ and 10 meV associated with each supercell eigenstate (solid gray and dashed black lines, respectively). It was demonstrated in Ref. [123] that assuming a large > 5 meV broadening for the calculation of the supercell DOS leads to results which are in good agreement with the standard, non–self-consistent Green's function method. However, this large broadening obscures the details that are visible in the supercell DOS calculated using a smaller 1 meV broadening, in which case there is a clear indication of a gap in the supercell DOS at energies above E_N. It can be further seen in this case that the DOS calculated using the SCGF and supercell models are in excellent agreement at energies close to E_N. Further detailed comparison of the SCGF and supercell models verifies that the SCGF approach describes the DOS projected to individual GaAs **k** states at energies close to E_N with a high degree of accuracy, elucidating detailed features of the CB structure which are not captured by the standard Green's function approach [129].

Having demonstrated that the SCGF model is capable of fully capturing the physics of the AIM, we now turn our attention to the presence of the gap in the DOS at energies just above E_N. Both the SCGF and supercell models confirm that the presence of this gap is an intrinsic feature of the two-band description of the GaN_xAs_{1-x} CB structure, suggesting that one must go beyond this simple approach in order to obtain a description of the DOS which is in agreement with experiment—i.e. for which there is no gap in the DOS. Seifikar et al. extended the SCGF model to take into account the full distribution of N-related states present in a disordered GaN_xAs_{1-x} alloy using the LCINS approach outlined in Section 9.2.3. In practice, this again consists of solving Equations 9.23 and 9.24 self-consistently, but where the sum in Equation 9.23 runs over all N-related states in the LCINS distribution, and where the broadening described by Equation 9.24 is evaluated directly for each individual N-related localized state $|j\rangle$.

The DOS calculated using this SCGF-LCINS approach is shown in Figure 9.11a (solid line) for $GaN_{0.0036}As_{0.9964}$ ($x = 0.36\%$), where it is compared to the results of a nonself-consistent LCINS Green's function calculation in which a purely imaginary energy-independent broadening $\Delta E_j = i\Delta_N$, with $\Delta_N = 1$ meV, is assumed for all N-related states (dashed line). We note that there are several qualitative differences between the DOS calculated using the SCGF-LCINS and two-band SCGF approaches [129]. First, and most importantly, we note that the inclusion of a full distribution of N-related localized states in the SCGF calculation leads to a closing of the gap in the DOS. This confirms that (1) the presence of a gap in the DOS in the converged SCGF calculation shown in Figure 9.10a (solid line) is an intrinsic property of simple two-band description of the CB structure, and (2) the lack of a gap in the measured DOS in real GaN_xAs_{1-x} materials [124,125] is associated with N-related alloy disorder, which plays a strong role in determining the details of the CB structure (cf. Section 9.2). Second, it is possible to directly identify features associated with distinct classes of N-related localized states in the DOS [129] and optical absorption spectra [133], on the basis of the understanding of these states developed through the LCINS theory. For example, the large dip in the DOS between 0.2 and 0.3 eV originates from hybridization of GaAs extended states with localized states associated with isolated N impurities, while the shoulder in the DOS lying just above the alloy CB edge in energy is associated with N–N pair-related localized states. The quantitative differences between the DOS calculated using these two approaches are significant at energies close to those of the N-related localized states associated with isolated N impurities and N–N pairs, once again highlighting

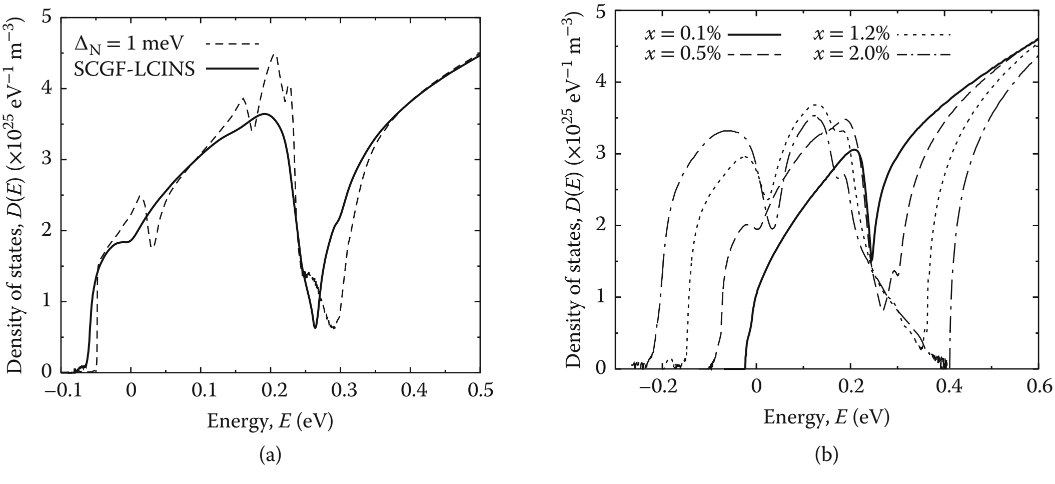


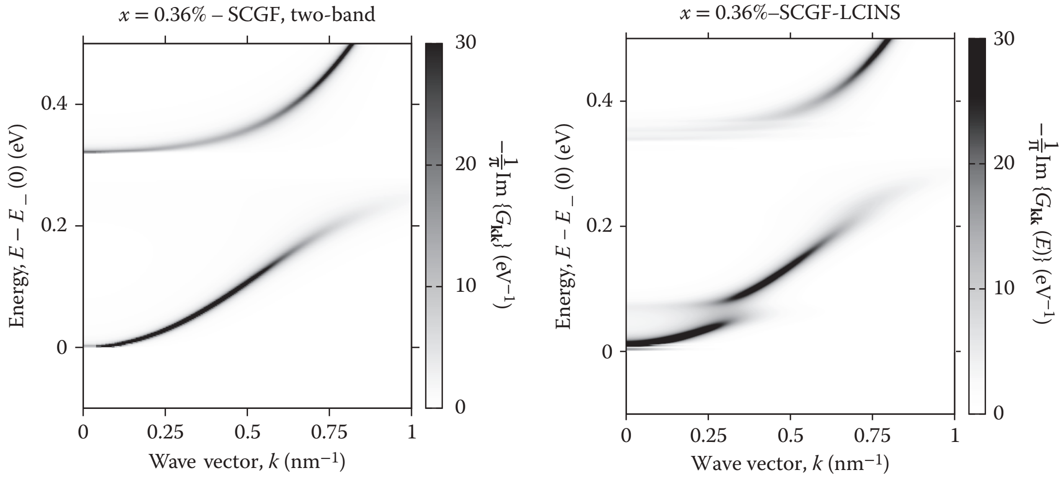
FIGURE 9.11 (a) DOS calculated for $GaN_{0.0036}As_{0.9964}$ ($x = 0.36\%$) using the standard, nonself-consistent Green's function method (dashed line)—assuming an energy-independent N-localized state broadening $\Delta E_j = i\Delta_N$ for all localized states, with $\Delta_N = 1$ meV—compared to that calculated using the SCGF method (solid line). In both cases, the CB structure is described by incorporating the full LCINS distribution of N-related localized states (cf. Section 9.2.3 and Equations 9.23 and 9.24). Significant quantitative differences exist between the DOS calculated using these approaches, particularly at energies close to N-related localized states, emphasizing the importance of the energy dependence of ΔE_j (see text). (b) DOS calculated using the full SCGF-LCINS model for GaN_xAs_{1-x} alloys having N compositions $x = 0.1\%$, 0.5%, 1.2%, and 2.0% (solid, dashed, dotted, and dash-dotted lines, respectively). The zero of energy is taken in each case at the unperturbed GaAs CB edge. (Adapted from M. Seifikar et al., *J. Phys. Condens. Matter*, 26, 365502, 2014.)

the importance of the energy-dependent complex localized state broadening in determining the details of the CB structure.

Figure 9.11b shows the DOS calculated using the SCGF-LCINS method for GaN_xAs_{1-x} alloys having N compositions $x = 0.1\%$, 0.5%, 1.2%, and 2.0% (shown using solid, dashed, dotted, and dash-dotted lines, respectively). Despite the detailed features present at each composition due to the effects of N-related alloy disorder on the CB structure, several general trends in the calculated evolution of the DOS with increasing N composition can be understood qualitatively in terms of the BAC description of the band structure. First, the reduction in energy of the alloy CB edge with increasing x is clearly apparent. Second, the dip in the DOS at energies above E_N increases in depth and width as x increases. This is generally consistent with the BAC model, in which the energy $E_-(0)$ ($E_+(0)$) of the lower (higher) energy BAC band at $k = 0$ decreases (increases) with increasing x. In disordered alloys this trend is clearly apparent as the energy at which the increase in the DOS above E_N occurs increases with increasing x, from approximately 0.23 eV in $GaN_{0.001}As_{0.999}$ to 0.41 eV in $GaN_{0.02}As_{0.98}$. This upward shift in energy of the higher-energy N-derived bands leads to an overall reduction of the DOS at energies above E_N, suggesting that a gap is present in the DOS of real, disordered GaN_xAs_{1-x} alloys for N compositions $x > 2\%$. This is to be contrasted with the results of the two-band calculations summarized above where, in the absence of alloy disorder, a gap in the DOS was predicted even at ultra-dilute N compositions $x \sim 0.1\%$. Features in the DOS which are similar to those associated with isolated N impurities are clearly visible at energies close to the unperturbed GaAs CB edge (0 eV), reflecting the impact of hybridization between the extended GaAs CB states and localized states associated with N–N pairs.

The SCGF method can also be used to calculate the broadened GaN_xAs_{1-x} CB dispersion described by the AIM. This is achieved by using the alloy CB eigenstates of the AIM Hamiltonian to compute the DOS projected to individual extended GaAs CB states $|\mathbf{k}\rangle$—i.e. by neglecting the integration (sum) over wave vectors in Equation 9.25 and instead computing the projected DOS separately at each wave vector $\mathbf{k}$ as $-\frac{1}{\pi}\text{Im}\{G_{\mathbf{kk}}(E)\}$, using the converged complex broadening functions $\Delta E_j(E)$ [129]. The broadened CB structures computed for $GaN_{0.0036}As_{0.9964}$ ($x = 0.36\%$) in this manner are shown in Figure 9.12a and 9.12b. Figure 9.12a shows the CB structure calculated using the two-band model described by Equations 9.27 and 9.28, in which all N-related localized states $|j\rangle$ are assumed to have the same energy E_N. The SCGF CB structure calculated using the two-band model corresponds well to that calculated using the two-band BAC model [129] and demonstrates (1) the strong N-induced energy broadening of the CB states, (2) the breakdown in the extended nature of the states in the lowest-energy alloy CB with increasing $k = |\mathbf{k}|$, as evidenced by the strong reduction in the projected DOS at large wave vectors, and (3) the presence of a gap in the CB DOS—that is, a range of energies at and above E_N for which the projected DOS vanishes at all $\mathbf{k}$ (cf. Figure 9.10a).

Figure 9.12b shows the CB structure calculated using the full SCGF-LCINS method, which takes into account the full LCINS distribution of N-related localized states. Comparing Figure 9.12a and 9.12b we note two important qualitative differences between the two-band and LCINS descriptions of the CB structure. First, the strong impact of localized states associated with N–N pairs can be observed at energies ~50 meV above the alloy CB edge (cf. Figure 9.11a)—hybridization of GaAs extended states $|\mathbf{k}\rangle$ with N–N pair-related localized states leads to a strong breakdown of the CB dispersion at energies close to those of the latter states, as evidenced by a strong reduction in the projected DOS. Second, the strong broadening of the CB structure associated with N-related disorder in the LCINS model leads to a closing of the gap in the DOS that is present in the two-band model (cf. Figure 9.11a): the lower and upper CBs are then linked by an energy range across which the project DOS is nonzero, thereby providing a pathway for scattering of electrons to higher-energy CB states (which has important consequences for the electron transport in intermediate and high applied fields). We again note that the CB dispersion obtained from the BAC model provides a reasonable description of the main features of the CB structure, but emphasize that the SCGF method provides much deeper insight into the nature of the GaN_xAs_{1-x} CB eigenstates, particularly given the strong impact of N-related alloy disorder [129].



(a)

(b)

FIGURE 9.12 CB structure of $GaN_{0.0036}As_{0.9964}$ ($x = 0.36\%$) calculated using the SCGF method—by projecting the alloy CB states to individual GaAs extended CB states, to obtain the alloy DOS projected to extended host matrix **k** states—$-\frac{1}{\pi}\mathrm{Im}\{G_{\mathbf{kk}}(E)\}$—for (a) a two-band model in which all N-related states have the same energy E_N and (b) the full LCINS distribution of N-related localized states (cf. Section 9.2.3). The two-band SCGF CB structure in (a) corresponds well to the band dispersion calculated using the two-band BAC model (cf. Equation 9.2). The SCGF-LCINS CB structure in (b) shows enhanced N-related broadening associated with alloy disorder, and a strong deviation from the BAC band dispersion due to hybridization of the GaAs CB edge with localized states associated with N–N pairs (lying approximately 50 meV above the alloy CB edge in energy). The zero of energy is taken at the $GaN_{0.0036}As_{0.9964}$ CB edge energy $E\_(0)$, calculated using the two-band BAC model. (Adapted from From M. Seifikar et al., *J. Phys. Condens. Matter*, 26, 365502, 2014.)

In summary, we have reviewed how the shortcomings of the BAC approach to modeling the CB structure and DOS in GaN_xAs_{1-x} and related alloys can be overcome by using a Green's function approach to solve the AIM in the presence of N impurities. In particular, we emphasized that the key physics missing from the BAC model is associated with (1) the complex energy-dependent broadening of N-derived quasilocalized states, consideration of which is necessary to remove unphysical aspects of the description of the DOS based on the BAC model and, (2) N-related alloy disorder, which is accurately captured using the LCINS method and introduces important qualitative and quantitative differences compared to calculations of the CB structure and DOS undertaken using a simplified two-band model. The detailed features of the CB structure and DOS revealed through a combined LCINS and SCGF analysis are of importance for understanding carrier transport in dilute nitride alloys, as they quantitatively capture the strong intrinsic and disorder-induced energy broadening of localized states in real materials, as well as the breakdown in Bloch character associated with the strong hybridization of the extended host matrix CB states with localized impurities.

We note that the theory outlined above is generally applicable to analyzing the band structure of highly mismatched semiconductor alloys in which highly electronegative isovalent impurities strongly perturb the CB, for example in dilute oxide [134,135] or dilute carbide [136,137] alloys. It is further expected that similar approaches could be applied to investigate highly mismatched alloys containing strongly electropositive isovalent impurities—for example, dilute bismide alloys, which will be discussed in the next chapter). However, in dilute bismides and related alloys the strong perturbation of the electronic structure at energies close to the VB edge, combined with the T_2 (as opposed to A_1) symmetry of the localized states associated with isolated impurities, necessitates that spin-orbit coupling be taken into account, which significantly complicates the description of the band structure and which would hence require that the SCGF method outlined above be modified in order to account for the more complicated nature of the p-like VB eigenstates. Having reviewed these detailed aspects of the impact of N incorporation on the CB structure and DOS, we now turn our attention to the theory of electron transport in GaN$_x$As$_{1-x}$ alloys, and review how these same features play a key role in determining the transport properties of real materials.

9.5.2 Electron Transport in GaN$_x$As$_{1-x}$ Alloys

Initial experimental measurements on dilute nitride alloys revealed that N incorporation has a significant impact on electron transport. While this is to be expected in light of the strong N-induced modifications to the CB structure discussed throughout this chapter, development of sufficiently accurate theoretical models to describe the measured transport properties of GaN$_x$As$_{1-x}$ has motivated detailed investigations of the physics of dilute nitride alloys which have significantly enhanced the fundamental understanding of this unusual class of semiconductor materials. Given the importance of electron transport in the operation of many of the devices described in Section 9.4, developing quantitative theoretical models describing the impact of N incorporation on electron transport represents a task which not only seeks to elucidate further aspects of the physics of N-containing alloys, but which is also of significant importance from the perspective of practical applications. Here, we review the development of the theory of electron transport in GaN$_x$As$_{1-x}$ alloys, beginning from general considerations that provide fundamental insight into the impact of N incorporation and analytical expressions describing important qualitative trends in the electron mobility, before considering more detailed models and numerical calculations that quantitatively describe the results of experimental measurements.

Experimental measurements have identified two general trends which must be described by theoretical models of the electron transport in GaN$_x$As$_{1-x}$. First, early experiments on GaN$_x$As$_{1-x}$ materials revealed that N incorporation results in a strong reduction of the electron mobility μ, from ~8500 cm^2 V^{-1} s^{-1} at $x = 0$ (GaAs) to <2000 cm^2 V^{-1} s^{-1}, even at ultra-dilute N compositions $x \sim 0.1\%$ [138,139]. Second, experimental measurements have suggested that while negative differential velocity[†] (NDV) is present in ultra-dilute GaN$_x$As$_{1-x}$ alloys [140–142], incorporation of N leads to a significant weakening of this effect compared to that present in GaAs. Additional measurements have demonstrated that NDV in GaN$_x$As$_{1-x}$ is strongly suppressed by incorporation of N at slightly higher compositions $x \sim 1\%$ [143], with NDV then being absent in GaN$_x$As$_{1-x}$ alloys except for a limited, ultradilute range of N compositions.

These experimental studies clearly demonstrated that N incorporation has a significant impact on the electron transport in GaN$_x$As$_{1-x}$ alloys, which must be understood in detail in order to predict and optimize the transport properties of dilute nitride materials and devices. That the electron mobility in dilute nitride alloys is intrinsically limited by N incorporation was explicitly demonstrated by Fahy and O'Reilly, who used an S-matrix approach to derive the scattering cross section σ of N impurities in the ultra-dilute limit [144]. Within the context of the BAC model they obtained the following expression connecting σ and the dependence of the CB edge energy E_- on the N composition x

$$\sigma = \frac{\pi a_0^6}{4}\left(\frac{m_e^*}{2\pi\hbar^2}\right)^2\left(\frac{\mathrm{d}E_-}{\mathrm{d}x}\right)^2, \tag{9.29}$$

[†] The drift velocity v_d of electrons in semiconductors tends to initially increase with the strength F of an applied electric field, displaying conventional ohmic behavior. As F increases it is possible for electrons to be accelerated to energies sufficient to facilitate scattering from high-mobility Γ valley states to states having lower mobilities in higher energy-satellite valleys. This intervalley scattering of electrons results in the existence of a range of F for which further increases in F result in a decrease of v_d—that is, the differential velocity $\frac{dv_d}{dF}$ is negative. In terms of the flow of electrical current through the material, this results in a negative (differential) electrical resistance for a given range of F, which is of use for a variety of applications. For example, NDV arising from intervalley scattering of electrons from the Γ to L valleys facilitates the operation of GaAs-based Gunn diodes.

where m_e^* is the electron effective mass, a_0 is the host matrix lattice constant, and $\frac{\mathrm{d}E_-}{\mathrm{d}x}$ is the derivative of the BAC model CB edge energy with respect to the N composition. One significant piece of information can immediately be gleaned from Equation 9.29: the fact that σ is directly proportional to the square of $\frac{\mathrm{d}E_-}{\mathrm{d}x}$ means that the large, composition-dependent bowing of the CB edge energy present in $\mathrm{GaN}_x\mathrm{As}_{1-x}$ and related alloys ensures that the cross section for electron scattering by N impurities is large. For example, a straightforward evaluation of Equation 9.29 for $\mathrm{GaN}_x\mathrm{As}_{1-x}$ gives an estimated electron scattering cross section $\sigma \approx 0.3\ \mathrm{nm}^2$ for an isolated N impurity in GaAs [144].

It is noteworthy that Equation 9.29 was derived independently of the fact that the scattering impurities are N atoms. This implies that the connection between σ and the dependence of the CB edge energy on the impurity concentration holds generally, regardless of the nature of the localized perturbation or defect which acts as a scattering center for electrons. Fahy and O'Reilly further demonstrated that the Born approximation (which is commonly used to analyze alloy scattering in conventional semiconductor alloys) underestimates σ, and hence the scattering rate due to N impurities, by roughly two orders of magnitude [144]. This stems from the fact that the Born approximation is inadequate for the calculation of *resonant* scattering, such as that associated with substitutional N impurities in GaAs.

Despite the fact that the above expression for σ is strictly valid only in the ultra-dilute limit, it can be extrapolated to higher N compositions with some degree of success. On the basis of Equation 9.29, Fahy and O'Reilly derived the following expression for the electron mobility μ at temperature T and N composition x [144]

$$\mu = \frac{e}{\pi a_0^3 x} \frac{1}{\sqrt{3 m_e^* k_\mathrm{B} T}} \left(\frac{2\pi\hbar^2}{m_e^*} \right)^2 \left(\frac{\mathrm{d}E_-}{\mathrm{d}x} \right)^{-2} . \tag{9.30}$$

Despite the fact that this simple treatment takes into account scattering due only to isolated N impurities, Equation 9.30 importantly describes the placement of an intrinsic limit on the electron mobility in dilute nitride alloys due to the presence of resonant scattering by N impurities. Interestingly, we deduce from Equation 9.30 that the strong reduction of the $\mathrm{GaN}_x\mathrm{As}_{1-x}$ CB edge energy E_- (and hence band gap) with increasing N composition—a major motivating factor for the use of dilute nitride alloys to develop GaAs-based photonic devices—plays a direct role in limiting the electron mobility, since μ is inversely proportional to the square of $\frac{\mathrm{d}E_-}{\mathrm{d}x}$. Using Equation 9.30 to calculate the variation of μ as a function of N composition suggests that the electron mobility is limited to values $<1000\ \mathrm{cm}^2\ \mathrm{V}^{-1}\ \mathrm{s}^{-1}$ at $x = 1\%$, an estimation which is consistent with the available experimental data which explicitly confirms the notion of intrinsically limited electron mobility in $\mathrm{GaN}_x\mathrm{As}_{1-x}$ and related alloys [144].

Following this initial analysis of μ based on the BAC model, in Ref. [131] Fahy et al. took the first steps toward developing a quantitative theory of the impact of N incorporation on the electron mobility by (1) explicitly deriving the scattering rates associated with N-related localized states, (2) evaluating the total scattering rate at a given N composition x using the full LCINS distribution of N-related localized states (cf. Section 9.2.3), and (3) using this scattering rate in conjunction with the BTE to derive an expression for μ which describes the impact of N-related scattering on the electron mobility.

In the case that an N-related localized state $|j\rangle$ lies energetically within the host matrix CB, we recall that the interaction of this localized state with an extended host matrix CB state results in its broadening into a quasi-localized resonance having a finite lifetime, described mathematically by a complex energy whose imaginary part is related to the lifetime as $\tau_j \sim \frac{\hbar}{\Delta_j}$ [128]. Under the assumption that the energy broadening associated with this resonance can be described using a Lorentzian lineshape [128], Fahy et al. used Fermi's golden rule to derive an expression for the scattering rate $R_j(E)$ between extended CB states $|\mathbf{k}\rangle$ and $|\mathbf{k}'\rangle$ due to the N-related localized state $|j\rangle$ [131]

$$R_j(E) = \frac{2\pi}{\hbar} \frac{|V_j|^4}{(E - E_j')^2 + \Delta_j^2} D_0(E) . \tag{9.31}$$

There are two important points to note regarding Equation 9.31. First, the scattering rate is isotropic—that is, it does not depend on $|\mathbf{k}-\mathbf{k}'|$. Second, Equation 9.31 is an analytical continuation of the expression describing the rate of scattering from a nonresonant localized state, which amounts to replacing the localized state energy E_j by the complex energy $E_j' + i\Delta_j$ to encapsulate the energy shift and broadening of $|j\rangle$ (cf. Section 9.5.1).

In order to calculate the total scattering rate $R_N(E)$ due to N impurities Fahy et al. employed the full LCINS distribution of N-related localized states $|j\rangle$ in the independent scattering approximation. The independent scattering approximation assumes that the total scattering rate due to randomly placed localized states is equal to the sum of the scattering rates associated with each individual localized state—that is, $R_N(E) = \sum_j R_j(E)$ [131]. This approximation is justified in practice because (1) by construction, the LCINS basis of N-related localized states interact only with the extended states of the unperturbed CB and not with one another, and (2) due to the low concentrations of large clusters of N atoms at dilute N compositions, it can reasonably be assumed that N-related localized states are not influenced by statistical correlations between their occurrence and location in a randomly disordered alloy.

Figure 9.13 shows the total energy-dependent resonant scattering rate $R_N(E)$ due to N-related localized states in GaN_xAs_{1-x}, computed using the full LCINS distribution (cf. Section 9.2.3) in the independent scattering approximation, for N compositions $x = 0.1\%$, 0.36%, 0.5%, 1.2%, and 2.0% (shown using solid, dashed, dotted, dash-dotted, and short-dash-gapped lines, respectively). Examining $R_N(E)$ in light of Equation 9.31 and the distribution of N impurity states described by the LCINS distributions of Section 9.2.3 we note a crucial trend: the N-related scattering is strongly enhanced for CB electrons having energies close to those of the N-related states in a given alloy. As we will see below, this is borne out by detailed theoretical calculations which are in quantitative agreement with experimental measurements of μ. This general trend is clearly evident for scattering by localized states associated with isolated N atoms, by the presence of a peak in $R_N(E)$ lying approximately 200 meV above the CB edge energy at $x = 0.1\%$, which increases in magnitude and shifts to higher energies as x increases to 2% and more N-related resonant scattering centeres are introduced into the CB. An additional, smaller peak in $R_N(E)$—describing scattering by localized states associated with N–N pairs—can be seen to emerge at energies closer to the CB edge, and increases in magnitude as x increases and more N-N pairs form in a statistically random

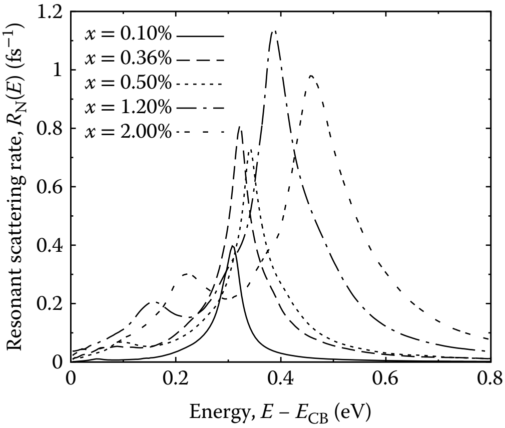


FIGURE 9.13 Total resonant scattering rate $R_N(E)$ associated with N-related localized states in GaN_xAs_{1-x}, calculated using the full LCINS distribution (cf. Section 9.2) in Equation 9.31 for N compositions $x = 0.1\%$, 0.36%, 0.5%, 1.2%, and 2.0% (solid, dashed, dotted, dash-dotted, and short-dash-gapped lines, respectively). The zero of energy is taken in each case at the alloy CB edge energy E_{CB}. A large peak in $R_N(E)$, associated with scattering by isolated N impurities, can be seen in each case. A second smaller peak in $R_N(E)$, associated with scattering by N–N pairs, can also be seen at lower energy. (Adapted from S. Fahy et al., *Phys. Rev. B*, 74, 035203, 2006.)

alloy. This demonstrates, as we see below, that while the number of N–N pairs at dilute N compositions is relatively small, localized states associated with N–N pairs can be expected to have a significant impact on the transport of CB electrons even at low energies. The strong impact of the energy dependence of N-resonant scattering on the transport properties can also be understood qualitatively in terms of the mean free path for electrons, $\lambda(E) = \frac{v_d}{R_{\mathrm{N}}(E)}$, which in applied fields can be as low as 5 nm for certain combinations of N compositions and electron energies [131].

As a point of technical significance, we emphasize that the complex energy broadening $i\Delta_j$ of the localized states in Equation 9.31 is strictly independent of energy. Ideally, calculations of the transport properties of electrons in dilute nitride alloys should incorporate complex energy-dependent broadening of the N-related localized states, as described by the SCGF model of Section 9.5.1 (cf. Equation 9.24). However, these aspects of the band structure are not straightforward to incorporate into conventional models of carrier transport. As a result, published calculations have to date employed energy-independent broadening in the computation of the scattering rates and simplified models of the CB structure which, despite the fact that they omit some of the details described in Section 9.5.1, provide reliable and quantitative descriptions of the measured trends in the electron mobility and NDV in $\mathrm{GaN}_x\mathrm{As}_{1-x}$ alloys. Following an approach of this type, Fahy et al. used the total N-related scattering rate $R_{\mathrm{N}}(E)$ in conjunction with the BTE to derive an expression for μ under the assumptions that (1) the CB dispersion is parabolic, with $E_{\mathbf{k}} = \frac{\hbar^2|\mathbf{k}|^2}{2m_e^*}$, and (2) the thermal (energy) distribution of electrons in the CB can be described using a Boltzmann (Fermi–Dirac) distribution in the non-degenerate (degenerate) regime [131]. Starting from the BTE in the relaxation time approximation, Fahy et al. calculated the current density J, at carrier density n, associated with electron drift in the presence of an applied electric field of strength F. This was then used to compute the electron mobility as $\mu = \frac{J}{enF}$, resulting in the following expression for μ at temperature T [131]

$$\mu = \frac{4e}{3\pi a_0^3}\sqrt{\frac{2}{\pi m_e^* k_{\mathrm{B}}T}}\left(\frac{2\pi\hbar^2}{m_e^*}\right)^2 \int_0^\infty \frac{\epsilon^3\, e^{-\epsilon^2}}{M(\epsilon^2 k_{\mathrm{B}}T)}\, d\epsilon\,, \tag{9.32}$$

where the zero of energy has been taken at the CB minimum, $\epsilon = \frac{\hbar^2k^2}{2m_e^*k_{\mathrm{B}}T}$ is a dimensionless measure of the energy of a CB electron, defined as the ratio of the kinetic energy to the thermal energy, and the function $M(E)$ is closely related to the total scattering rate $R_{\mathrm{N}}(E)$ due to N-related localized states [131]

$$M(E) = \frac{1}{N_c}\sum_j \frac{\beta_j^4}{(E-E_j')^2 + \Delta_j^2}\,. \tag{9.33}$$

In addition to highlighting the strong influence of N-resonant scattering on μ, Equation 9.32 also reveals the important role played by the difference in energy between N-related localized states and the Fermi level in determining the mobility. The important role played by the Fermi level in this context can lead to unusual trends such as, e.g. an increase in μ with increasing x depending upon the carrier density n. Additionally, the dependence of the Fermi level on T leads to further variations in μ, due to the change in energy between the temperature-dependent extended CB states and the N-related localized states (which are relatively insensitive to changes in temperature). This aspect of the calculation of μ raises interesting points related to the interpretation of measurements of μ, particularly in doped materials, emphasizing the highly unusual nature of the impact of N incorporation on the transport properties. A detailed analysis of this aspect of the calculation of μ is beyond the scope of the current discussion—we refer interested readers to Refs. [131] and [145] for a discussion of this issue relating to bulk $\mathrm{GaN}_x\mathrm{As}_{1-x}$ alloys, as well as in QWs and gated QW heterostructures.

In order to ascertain the predictive capability of this more sophisticated model of μ, it is instructive to compare the results obtained using Equation 9.32 to experiment. Measurements of μ in $\mathrm{GaN}_x\mathrm{As}_{1-x}$

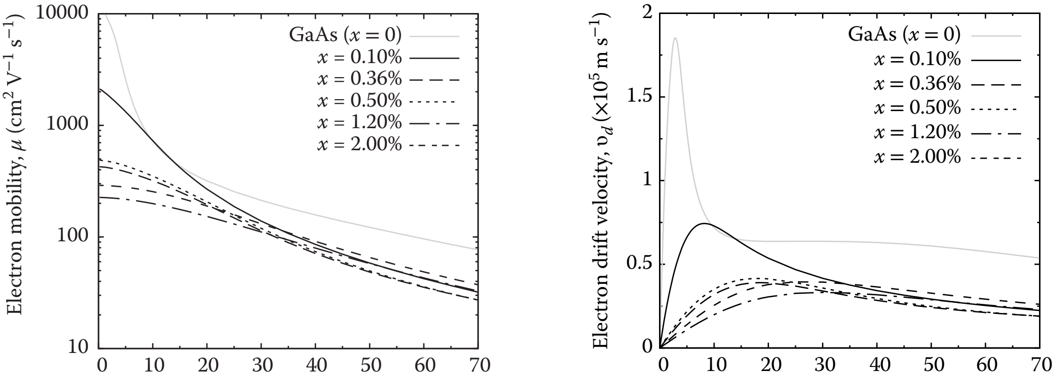
at ultra-dilute N compositions $x < 0.4\%$ have been undertaken by a number of groups, both in bulk [146] and QW samples [138,147]. For example, Young et al. measured extremely low values of 262 and 187 $\text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$ for μ in bulk-like thin films having respective N compositions $x = 0.1\%$ and 0.4% and relatively high carrier densities $n \sim 5 \times 10^{18}\ \text{cm}^{-3}$ [146]. Using Equation 9.32, the calculated mobilities at these compositions and carrier densities are 2200 and 385 $\text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$ for $x = 0.1\%$ and 0.4%, respectively [131]. We note that the calculated value at $x = 0.1\%$ overestimates μ by close to an order of magnitude. While the calculated value of μ at $x = 0.4\%$ is much closer to experiment, the theoretical value obtained using Equation 9.32 still represents an overestimation of the mobility. This trend is relatively consistent across bulk and QW samples, as well as across different temperatures and doping levels. For example, a value $\mu = 3000\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$ for a QW having $x = 0.12\%$ and an areal carrier density $\sim 1 \times 10^{12}\ \text{cm}^{-2}$—calculated using the equivalent expression to Equation 9.32 in two dimensions [131]—overestimates the measured value $\mu \sim 600\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$ [138] by a factor of five.

From these comparisons to experimental measurements, it is clear that while N-related scattering plays the dominant role in limiting μ in $\text{GaN}_x\text{As}_{1-x}$, the tendency of Equation 9.32 to overestimate μ highlights that additional scattering mechanisms also play an important role in the electron transport. From a theoretical perspective, this mandates that the models presented above be refined in order to better capture the physics involved in quantitatively determining μ in $\text{GaN}_x\text{As}_{1-x}$. In practice, this is achieved by abandoning the assumed carrier distribution function(s) used to derive Equation 9.32 and instead solving the BTE numerically in order to directly compute the influence of additional scattering mechanisms on the drift current density J. An initial analysis of this type was undertaken by Vaughan and Ridley, who in Ref. [148] developed a ladder method to solve the BTE for a nonparabolic CB. Using the BAC description of the CB structure as the basis for their solution of the BTE, Vaughan and Ridley calculated the dependence of μ in $\text{GaN}_x\text{As}_{1-x}$ on x and T in a model including polar optical, acoustic phonon, and piezoelectric scattering, as well as scattering by ionized, neutral, and N impurities. Despite the level of detail included in this model, the conclusions reached by Vaughan and Ridley are qualitatively similar to those described by Fahy and O'Reilly in Ref. [144] on the basis of Equation 9.30—i.e. that the room temperature mobility in $\text{GaN}_x\text{As}_{1-x}$ is limited to values $<2000\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$ by scattering associated with isolated N impurities. Quantitatively, the analysis of Vaughan and Ridley suggests a slight reduction of the intrinsic "ceiling" on μ, with the calculated value $\mu \sim 800\ \text{cm}^2\ \text{V}^{-1}\ \text{s}^{-1}$ at $x = 1\%$ being approximately 80% of that estimated using Equation 9.30. However, regardless of its detailed treatment of a range of scattering mechanisms, this BTE-based model still significantly overestimates μ compared to experimental measurements. This strongly reinforces that the description of the CB structure by the BAC model, which in this context describes N-resonant scattering associated with isolated N impurities only, is insufficient to account for the significant reduction in μ observed in real $\text{GaN}_x\text{As}_{1-x}$ alloys.

In light of these insights, it is clear that a quantitative analysis of the transport properties of $\text{GaN}_x\text{As}_{1-x}$ alloys requires a refined approach which, in addition to resonant scattering by isolated N impurities, also takes into account resonant scattering associated with N pair and cluster states, as well as the influence of additional scattering mechanisms. Such a model was developed in Ref. [149] by Seifikar et al. who solved the BTE numerically in the relaxation time approximation and in the presence of a range of scattering mechanisms including polar optical and acoustic phonon scattering, intervalley scattering between the Γ and L valleys, and resonant scattering by the full LCINS distribution of N-related localized states. Comparing the calculated carrier distributions for the Γ and L valleys as a function of the strength of F of an applied electric field, Seifikar et al. first showed that resonant scattering by N-localized states strongly inhibits acceleration of electrons to higher energies. As a result, intervalley scattering at fixed F is strongly suppressed in $\text{GaN}_x\text{As}_{1-x}$ compared to GaAs. For example, intervalley scattering in room temperature GaAs results in approximately 80% of electrons occupying L valley states at $F = 10\ \text{kV cm}^{-1}$, while at $x = 0.1\%$ (2.0%) a much stronger applied field $F = 30\ \text{kV cm}^{-1}$ ($45\ \text{kV cm}^{-1}$) is required to bring about the same level of electron transfer from Γ to L valley states. That this reduction of the intervalley scattering is brought about by N incorporation was confirmed by a calculation of the average energy of Γ electrons as a function of F, demonstrating explicitly that N-resonant scattering erodes the initial strong increase in electron energy with F calculated for GaAs. The analysis of Seifikar et al. further showed that the strong nonparabolicity of the lowest CB in GaN_xAs_{1-x} brings about significant increases in the rates of polar optical and acoustic phonon scattering, particularly at energies close to N-related states, suggesting in particular that scattering by polar optical phonons plays an important (but not dominant) role in determining the transport properties of GaN_xAs_{1-x} alloys [149].

Using the carrier distribution functions obtained by numerical solution of the BTE to compute J, Seifikar et al. calculated the variation of μ $(= \frac{J}{eFn})$ and v_d $(= \frac{J}{en})$ with F for a series of GaN_xAs_{1-x} alloys having N compositions $x = 0.1 - 2.0\%$. The results of these calculations are summarized in Figures 9.14a and 9.14b. Figure 9.14a shows the calculated variation of μ with F in GaAs (solid gray line) and in GaN_xAs_{1-x} for $x = 0.1\%$, 0.36%, 0.5%, 1.2% and 2.0% (solid, dashed, dotted, dash-dotted, and short-dash-gapped lines, respectively) [150,151]. The calculated low-field mobility $\mu = 2000$ cm^2 V^{-1} s^{-1} at $x = 0.1\%$ is relatively close to the measured value of 1200–1400 cm^2 V^{-1} s^{-1} [139], while the calculated value $\mu = 410$ cm^2 V^{-1} s^{-1} at $x = 0.4\%$ is closer to the measured value of 280–360 cm^2 V^{-1} s^{-1} [139]. Given that this model overestimates μ in GaAs—due in part to the fact that the model neglects CB nonparabolicity in the absence of N—these comparisons to experiment suggest that including resonant scattering by the full LCINS distribution of N-localized states correctly describes the impact of N incorporation on the electron transport. That this is the case can also be seen by considering the calculated variation of v_d with F, and the corresponding implications for the presence of NDV in GaN_xAs_{1-x}.

Figure 9.14b shows the calculated variation of v_d with F in GaAs (solid gray line) and in GaN_xAs_{1-x} for $x = 0.1\%$, 0.36%, 0.5%, 1.2% and 2.0% (solid, dashed, dotted, dash-dotted, and short-dash-gapped lines, respectively). These calculations suggest that incorporating N results in a strong reduction of the maximum value $v_{d,\max}$ of v_d, as well as an increase in the field strength F required to reach $v_{d,\max}$. For



(a) (b)

FIGURE 9.14 (a) Variation of the electron mobility μ in GaN_xAs_{1-x} as a function of the strength F of an applied electric field, calculated using the total N-related resonant scattering rate of Equation 9.31 in a numerical solution of the BTE, for N compositions $x = 0.1\%$, 0.36%, 0.5%, 1.2%, and 2.0% (solid, dashed, dotted, dash-dotted, and short-dash-gapped lines, respectively). The calculated variation of μ with F for GaAs is shown for reference (solid gray line). A reduction of μ by approximately an order of magnitude is calculated even at the ultra-dilute N composition $x = 0.1\%$, reflecting the strong impact of N-resonant scattering on the electron transport. (b) Variation of the electron drift velocity v_d in GaN_xAs_{1-x} with F, calculated using the total N-related resonant scattering rate of Equation 9.31 in a numerical solution of the BTE, for N compositions $x = 0.1\%$, 0.36%, 0.5%, 1.2%, and 2.0% (solid, dashed, dotted, dash-dotted, and short-dash-gapped lines, respectively). The calculated variation of v_d with F for GaAs is shown for reference (solid gray line). N incorporation leads to a strong reduction in the maximum value $v_{d,\max}$ of v_d. We also note a calculated strong reduction in NDV compared to GaAs, and that appreciable NDV is present only for ultra-dilute N compositions $x < 0.4\%$. (Adapted from M. Seifikar et al., *Phys. Rev. B*, 84, 165216, 2011.)

example, the calculated maximum value $v_{d,\max}$~0.5×10^5 m s^{-1} at $x = 1.2\%$ is approximately one-quarter of that calculated for GaAs, while the field strength required to achieve $v_{d,\max}$ increases from $F = 4$ kV cm^{-1} in GaAs to ~30 kV cm^{-1} at $x = 1.2\%$. Additionally, while the variation of v_d with F displays strong NDV above $F = 4$ kV cm^{-1} in GaAs, at $x = 1.2\%$ there is no evidence of NDV, with the calculated $v_d(F)$ saturating for applied fields $F > 30$ kV cm^{-1} [149]. Examining the calculated variation of v_d with F for the other N compositions shown in Figure 9.14b demonstrates that (1) N incorporation leads to a weakening of the NDV effect, as evidenced by a strong reduction in the magnitude of $\frac{dv_d}{dF}$ above $v_{d,\max}$, and (2) NDV is absent except at ultra-dilute N compositions $x < 0.5\%$. We note that these trends are in good agreement with those observed in experiment [140–143] and that the increasing occurrence of N pairs and clusters with increasing x leads to enhanced resonant scattering by N impurities which suppresses NDV by simultaneously limiting electron acceleration and reducing intervalley scattering from Γ to L states [149].

Calculations of the transport properties of GaN$_x$As$_{1-x}$ have also been undertaken by Vogiatzis and Rorison, who used a single-electron Monte Carlo model to analyze μ and v_d as a function of F in the ultra-dilute regime [150,151]. In Ref. [150] they considered a model of the transport including resonant scattering by localized states associated with isolated N impurities only, but which also included polar optical and acoustic phonon scattering, as well as intervalley scattering between Γ and L states. Qualitatively, the conclusions reached on the basis of this detailed analysis are similar to those of Seifikar et al.—i.e. that (1) N-related scattering limits the acceleration of electrons in the presence of an applied electric field, leading to a strong reduction of μ and $v_{d,\max}$, (2) intervalley scattering mediated by polar optical phonons is reduced compared to that present in GaAs, and (3) NDV is present in ultra-dilute GaN$_x$As$_{1-x}$ alloys having $x < 0.4\%$, but with the magnitude of $\frac{dv_d}{dF}$ above $v_{d,\max}$ strongly reduced compared to that in GaAs. Quantitatively, the analysis of Vogiatzis and Rorison suggested that the presence and character of NDV in the ultra-dilute regime is governed by the interplay of polar optical scattering with N-related resonant scattering of electrons to higher-energy N-derived CB states. This is in contrast to the nature of NDV in GaAs, which is understood to arise primarily due to intervalley scattering of electrons from Γ to L valley states. This Monte Carlo analysis was further developed in Ref. [151], in which Vogiatzis and Rorison extended their treatment to include resonant scattering due to localized states associated with N–N pairs and selected larger clusters of N atoms. Using this approach, it was again confirmed that N pairs and clusters play a dominant role in limiting μ and $v_{d,\max}$ in GaN$_x$As$_{1-x}$. However, while the conclusions of these Monte Carlo analyses are qualitatively similar to those arising from studies based on the BTE, it is noteworthy that the electron mobilities calculated by Vogiatzis and Rorison using this extended model are in excellent quantitative agreement with experiment. For example, the calculated values $\mu = 1296$, 738, and 344 cm^2 V^{-1} s^{-1} at $x = 0.1\%$, 0.2%, and 0.4% [151] correspond very well with the measured values $\mu \sim 1200$–1400, 700, and 280–360 cm^2 V^{-1} s^{-1} [139]. These calculations highlight the utility of the Monte Carlo approach for the calculation of the transport properties of semiconductors, and demonstrate the ability of this approach to provide a predictive platform for the design and optimization of carrier transport in dilute nitride materials and devices.

Comparison of the results described above—obtained using the BTE [149] and Monte Carlo [150,151] models—to calculations in which N-related scattering associated only with isolated N impurities is taken into account demonstrates that scattering of electrons by N pair and cluster states (lying close in energy to the CB edge) plays an important role in bringing about the measured strong reduction in μ and suppression of NDV in GaN$_x$As$_{1-x}$. These analyses reinforce the need for a detailed treatment of N-related localized states and their role in electron scattering beyond the BAC model, and further emphasize the utility of the LCINS description of the CB structure to predict the properties of real dilute nitride alloys, in which short-range disorder plays a crucial role (cf. Section 9.2).

To summarize, in this section we have presented an overview of the theory and simulation of electron transport in GaN$_x$As$_{1-x}$. The theoretical models reviewed in this section—beginning from simple descriptions which identify intrinsic limitations on the electron mobility in N-containing alloys and increasing in sophistication to provide quantitative descriptions of measured transport properties—constitute the most complete available treatment of carrier transport in dilute nitride alloys. The results of calculations based on these models not only provide quantitative insight into the complicated nature of electron transport in dilute nitride alloys, but also serve as rigorous tests of models describing the impact of N incorporation on the CB structure in $\text{GaN}_x\text{As}_{1-x}$ and related alloys. In particular, the sophisticated nature of the models required to understand measurements of basic transport properties in $\text{GaN}_x\text{As}_{1-x}$ alloys emphasizes the unusual characteristics of dilute nitride alloys, and highlights the challenges that N-containing and related classes of highly mismatched alloys present on a theoretical level.

9.6 Summary and Conclusions

In this chapter, we have provided an introduction to the theory and simulation of dilute nitride materials and devices, which we have framed within the context of recent and ongoing research on $(\text{In})\text{GaN}_x\text{As}_{1-x}$ and related N-containing alloys. Beginning with an overview of the electronic properties of $\text{GaN}_x\text{As}_{1-x}$ and related alloys, we highlighted the potential to exploit the strong N-induced modifications of the band structure of conventional III–V semiconductors to develop a range of photonic and photovoltaic devices with enhanced performance and capabilities. In particular, we discussed that the extremely rapid reduction in band gap brought about by N incorporation offers the potential to greatly extend the wavelength range accessible via growth on GaAs substrates, which has stimulated significant interest in developing $(\text{In})\text{GaN}_x\text{As}_{1-x}$ and related alloys for applications in a range of GaAs-based photonic and photovoltaic technologies. Additionally, we highlighted that experimental measurements have demonstrated that the impurity-like behavior of substitutional N atoms leads to a reduction in the temperature sensitivity of the $(\text{In})\text{GaN}_x\text{As}_{1-x}$ band gap. This, combined with the enhanced carrier and optical confinement offered by GaAs heterostructures (compared to conventional InP-based structures), means that $(\text{In})\text{GaN}_x\text{As}_{1-x}$ alloys are particularly attractive for the development of GaAs-based 1.3- and 1.55-μm edge-emitting lasers and VCSELs, which have the potential to offer improved performance and new capabilities compared to the incumbent InP-based technologies in this wavelength range.

Beginning in Section 9.2, we highlighted the need for detailed atomistic models to provide a quantitative understanding of the strong impact of N incorporation on the band structure of (In)GaAs(P). We reviewed detailed analysis of the electronic structure of $\text{GaN}_x\text{As}_{1-x}$ and $\text{GaN}_x\text{P}_{1-x}$ alloys based on first principles calculations and on the EP method, which revealed that N acts as an isovalent impurity when incorporated into common III–V compounds and alloys in dilute concentrations and leads to a strong perturbation of the CB structure. We described how a simple two-band BAC model provides a simple phenomenological description of the main features of the $(\text{In})\text{GaN}_x\text{As}_{1-x}$ band structure, allowing for simple interpretations of a broad range of experimental measurements. We then reviewed that the TB method can be used to explicitly parameterize the BAC model, placing it on a quantitative footing, and summarized the results of the calculations that describe the impact of isolated substitutional N impurities on the CB structure of a range of III–V compounds. On this basis we outlined the derivation of an extended basis set 10-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian for $(\text{In})\text{GaN}_x\text{As}_{1-x}$ alloys, which directly accounts for the impact of N incorporation on the CB structure using a BAC approach. We provided a consistent set of band structure parameters for performing calculations on $(\text{In})\text{GaN}_x\text{As}_{1-x}$ alloys and heterostructures.

Following this, we discussed models of the dilute nitride band structure beyond the BAC model, focusing in particular on the TB-derived LCINS method, which is capable of quantitatively describing the highly unusual properties of dilute nitrides via direct treatment of the full range of N-related localized impurity states present in real, disordered alloys. The results of LCINS calculations were shown to be in excellent quantitative agreement with experimental measurements of the electron effective mass and g factor in $\text{GaN}_x\text{As}_{1-x}$ alloys, highlighting the predictive power of the approach and underscoring its use as a platform to facilitate detailed analysis of the electronic, optical, and transport properties of real N-containing materials.

Next, in Section 9.3, we provided an overview of the theory and modeling of 1.3-μm InGaNAs QW lasers grown on GaAs. We applied the 10-band **k·p** model to the study of InGaNAs/GaAs QW heterostructures and quantified the impact of N on the (1) CB structure and DOS, (2) material gain, and (3) threshold characteristics of a realistic device structure. We showed that while N incorporation compensates the large compressive strain associated with long-wavelength InGaAs/GaAs QWs and readily enables the emission wavelength to be extended to 1.3 μm, this is associated with a slight degradation of the gain characteristics resulting from (1) a reduction in the optical transition strengths due to N-induced hybridization, and (2) a reduction in the quasi-Fermi level separation at fixed injection arising from the N-induced increase of the DOS in the CB. However, these calculations also indicate that InGaNAs single QW laser structures designed to emit at 1.3 μm offer (1) enhanced carrier confinement, and (2) roughly equivalent differential gain than that calculated for optimized InP-based laser structures. This demonstrates that InGaNAs alloys can be used to achieve telecom-wavelength devices on GaAs having low leakage currents, and which offer impressive high-speed/bandwidth performance.

Following this, we considered the impact of N incorporation on the recombination pathways in 1.3 μm InGaNAs lasers. We outlined a method by which experimental measurements of the integrated SE from a QW device can be used to provide detailed insight into the loss mechanisms present in a semiconductor laser, by allowing direct extraction of the defect-related, radiative, and nonradiative (Auger) recombination coefficients. We reviewed a direct comparison between theory and experiment for the temperature dependence of the recombination coefficients, thereby demonstrating that a model based on the 10-band **k·p** Hamiltonian is capable of providing a quantitative description of the temperature dependent loss mechanisms in real InGaNAs QW lasers. Experimental measurements show that defect-related recombination accounts for approximately one-half of the threshold current density at room temperature in a 1.3-μm InGaNAs laser, similar to that observed in GaAs-based metamorphic laser structures. This suggests that there is significant potential for improvement in the performance of these devices, through refinement of the material growth in order to reduce the impact of N-related defects.

While good characteristics have been demonstrated by 1.3- and 1.55-μm dilute nitride semiconductor lasers, the advantages of using dilute nitrides—e.g. in GaAs-based long-wavelength VCSEL structures—have not yet been sufficiently pronounced to displace the well-established incumbent technologies based on InP. However, as outlined in this chapter, following a loss of momentum in the development of dilute nitride devices in the mid-2000s, more recently dilute nitrides have made a resurgence due to interest in a wider range of potential applications. In Section 9.4 we provided an overview of a number of alternative and emerging applications of dilute nitride alloys. We reviewed recent progress related to the development of (1) GaAs-based 1.55-μm lasers incorporating GaAs-based InGaNAs(Sb) heterostructures, (2) 1.3-μm SOAs, EAMs, and APDs based on (In)GaN$_x$As$_{1-x}$ alloys, and (3) GaNAsP alloys and heterostructures for monolithic integration of direct-gap III–V laser structures on Si. Through these summaries we demonstrated that several classes of dilute nitride devices display impressive characteristics and significant promise for practical applications, and also highlighted areas of ongoing research interest relating to the use of dilute nitride alloys in the development of novel GaAs- and Si-based photonics technologies.

In Section 9.5, we presented an overview of the theory and simulation of electron transport in dilute nitride alloys. We began with a detailed consideration of the CB structure and demonstrated that while a BAC-based approach is sufficient to calculate, for example, the electronic and optical properties to a level required to understand and predict the properties of dilute nitride semiconductor lasers, the BAC model has fundamental limitations which make it unsuitable to analyze carrier transport. We reviewed how, in order to overcome these limitations, the AIM has been adapted and applied to the study of the dilute nitride CB structure. We reviewed theoretical work based on solving the AIM for GaN$_x$As$_{1-x}$ alloys using Green's function methods, and showed how a self-consistent model provides a conceptually simple, physically rich, and quantitatively accurate description of the strongly perturbed CB structure. In light of the SCGF model, we highlighted the important role played by alloy disorder in determining the nature of the CB structure and DOS in real GaN$_x$As$_{1-x}$ alloys.

Next, we reviewed the development of theoretical models describing the impact of N incorporation on the transport of electrons in GaN_xAs_{1-x}. Beginning with general considerations, we highlighted a generic relationship between the electron scattering cross section and the strong N composition dependence of the CB edge energy, thereby demonstrating that N incorporation places intrinsic limitations on the electron mobility in dilute nitride alloys. We then reviewed the resonant scattering of CB electrons by N-related localized states, highlighting that (1) electrons having energies close to those of N-related states undergo particularly strong impurity scattering, and (2) while N-related scattering plays the dominant role in bringing about the low electron mobilities observed in experimental measurements, alternative scattering mechanisms must be considered in order to develop a quantitative understanding of the electron transport. On this basis, we reviewed detailed numerical calculations of the transport properties based on BTE and Monte Carlo approaches. These calculations again confirm that N-related scattering plays a dominant role in determining the transport properties of real GaN_xAs_{1-x} alloys, and indicate that scattering by N–N pairs and larger clusters of N atoms plays a crucial role in determining the transport properties as the N composition increases above the ultra-dilute regime. This highlights that approaches beyond the BAC model are required to quantitatively understand basic experimental results pertaining to electron transport in GaN_xAs_{1-x}. Overall, the calculations reviewed in Section 9.5 emphasize that (1) the observed low electron mobilities in GaN_xAs_{1-x} bulk materials and heterostructures are an intrinsic property of dilute nitride alloys, and (2) resonant scattering by N-related localized states plays a dominant role in limiting the electron mobility and suppressing NDV in GaN_xAs_{1-x}.

Overall, we conclude that the potential to develop photonic and photovoltaic materials and devices with enhanced characteristics and capabilities is certain to continue to stimulate interest in dilute nitride alloys in the coming years. The theoretical results and models described in this chapter form the basis of a consistent and accurate approach to the theory and simulation of dilute nitride alloys, which continues to find use in the interpretation of experimental studies, as well as in the design and optimization of new materials and structures for novel device applications.

Acknowledgments

The authors dedicate this chapter to the memory of Prof. Naci Balkan (University of Essex, UK), with whom they shared many enjoyable and fruitful years of collaboration. The writing of this chapter was supported by the Engineering and Physical Sciences Research Council, UK (project no. EP/K029665/1), and by Science Foundation Ireland (project no. 15/IA/3082). The authors acknowledge the contributions of the many colleagues with whom they have had the pleasure to collaborate on the study of dilute nitride alloys. In particular, the authors thank Profs. S. J. Sweeney and A. R. Adams (University of Surrey, UK), Prof. S. Tomić (University of Salford, UK), Prof. S. Fahy (University College Cork, Ireland), Prof. H. Riechert (Paul-Drude-Institut für Festkörperelektronik, Germany), and Dr. N. Vogiatzis (Semtech Corporation, UK) for many invaluable discussions on the physics and applications of dilute nitrides, and for their contributions to the research described in this chapter. C.A.B. thanks Ms. W. Xiong (University of Bristol, UK) for her assistance in producing several of the figures contained in this chapter.

References

1. P. S. Zory. *Quantum Well Lasers*. Boston, MA: Academic Press, 1993.
2. E. Kapon. *Semiconductor Lasers*, Vols. I & II. San Diego, CA: Academic Press, 1999.
3. C. A. Broderick, M. Usman, S. J. Sweeney, and E. P. O'Reilly. Band engineering in dilute nitride and bismide semiconductor lasers. *Semicond. Sci. Technol.*, 27:094011, 2012.
4. E. Söderberg, J. S. Gustavsson, P. Modh, A. Larsson, Z. Zhang, J. Berggren, and M. Hammar. High-temperature dynamics, high-speed modulation, and transmission experiments using 1.3-μm InGaAs single-mode VCSELs. *J. Lightwave Technol.*, 25:2791, 2007.

5. D. L. Huffaker, G. Park, Z. Zou, O. B. Shchekin, and D. G. Deppe. 1.3 μm room-temperature GaAs-based quantum-dot laser. *Appl. Phys. Lett.*, 73:2564, 1998.
6. V. M. Ustinov, N. A. Maleev, A. E. Zhukov, A. R. Kovsh, A. Yu. Egorov, A. V. Lunev, B. V. Volovik, I. L. Krestnikov, Yu. G. Musikhin, N. A. Bert, P. S. Kop'ev, Zh. I. Alferov, N. N. Ledentsov, and D. Bimberg. InAs/InGaAs quantum dot structures on GaAs substrates emitting at 1.3 μm. *Appl. Phys. Lett.*, 74:2815, 1999.
7. I. Tångring, S. M. Wang, M. Sadeghi, and A. Larsson. 1.27 μm metamorphic InGaAs quantum well lasers on GaAs substrates. *Electron. Lett.*, 42:691, 2006.
8. S. M. Wang, editor. *Lattice Engineering: Technology and Applications*. Singapore: Pan Stanford Publishing, 2012.
9. S. Bogusevschi, C. A. Broderick, and E. P. O'Reilly. Theory and optimization of 1.3 μm metamorphic quantum well lasers. *IEEE J. Quant. Electron.*, 52:2500111, 2016.
10. M. Kondow, K. Uomi, A. Niwa, T. Kitatani, S. Watahiki, and Y. Yazawa. Gas-source molecular beam epitaxy of GaN_xAs_{1-x} using a N radical as the N source. *Jpn. J. Appl. Phys.*, 33:L1056, 1996.
11. M. Kondow, T. Kitatani, S. Nakatsuka, M. C. Larson, K. Hakahara, Y. Yazawa, M. Okai, and K. Uomi. GaInNAs: A novel material for long-wavelength semiconductor lasers. *IEEE J. Sel. Top. Quantum Electron.*, 3:719, 1997.
12. E. P. O'Reilly, A. Lindsay, P. J. Klar, A. Polimeni, and M. Capizzi. Trends in the electronic structure of dilute nitride alloys. *Semicond. Sci. Technol.*, 24:033001, 2009.
13. D. A. Livshits, A. Y. Egorov, and H. Riechert. 8 W continuous wave operation of InGaAsN lasers at 1.3 μm. *Electron. Lett.*, 36:1381, 2000.
14. G. Steinle, H. Riechert, and A. Y. Egorov. Monolithic VCSEL with InGaAsN active region emitting at 1.28 μm and CW output power exceeding 500 μW at room temperature. *Electron. Lett.*, 37:632, 2001.
15. H. Riechert, A. Ramakrishnan, and G. Steinle. Development of InGaAsN-based 1.3 μm VCSELs. *Semicond. Sci. Technol.*, 17:892, 2002.
16. S. R. Bank, H. P. Bae, H. B. Yuen, M. A. Wistey, L. L. Goddard, and J. S. Harris. Room-temperature continuous wave 1.55 μm GaInNAsSb laser on GaAs. *Electron. Lett.*, 42:156, 2006.
17. S. R. Bank, M. A. Wistey, L. L. Goddard, H. B. Yuen, V. Lordi, and J. S. Harris. Low-threshold continuous-wave 1.5-μm GaInNAsSb lasers grown on GaAs. *IEEE J. Quant. Electron.*, 40:656, 2004.
18. S.-H. Wei and A. Zunger. Giant and composition-dependent optical bowing coefficient in GaAsN alloys. *Phys. Rev. Lett.*, 76:664, 1996.
19. J. Slater and G. Koster. Simplified LCAO method for the periodic potential problem. *Phys. Rev.*, 94:1498, 1954.
20. H. P. Hjalmarson, P. Vogl, D. J. Wolford, and J. D. Dow. Theory of substitutional deep traps in covalent semiconductors. *Phys. Rev. Lett.*, 44:810, 1980.
21. L. Bellaiche, S.-H. Wei, and A. Zunger. Localization and percolation in semiconductor alloys: GaAsN vs. GaAsP. *Phys. Rev. B*, 54:17568, 1996.
22. P. R. C. Kent and A. Zunger. Theory of electronic structure evolution in GaAsN and GaPN alloys. *Phys. Rev. B*, 64:115208, 2001.
23. J. Wu, W. Walukiewicz, K. M. Yu, J. W. Ager III, E. E. Haller, Y. G. Hong, H. P. Xin, and C. W. Tu. Band anticrossing in $GaP_{1-x}N_x$ alloys. *Phys. Rev. B*, 65:241303, 2002.
24. M. Güngerich, P. J. Klar, W. Heimbrodt, G. Weiser, J. F. Geisz, C. Harris, A. Lindsay, and E. P. O'Reilly. Experimental and theoretical investigation of the conduction band edge of GaN_xP_{1-x}. *Phys. Rev. B*, 74:241202, 2006.
25. C. Harris, A. Lindsay, and E. P. O'Reilly. Evolution of N defect states and optical transitions in ordered and disordered $GaP_{1-x}N_x$ alloys. *J. Phys. Condens. Matter*, 20:295211, 2008.
26. P. R. C. Kent and A. Zunger. Evolution of III–V nitride alloy electronic structure: The localized to delocalized transition. *Phys. Rev. Lett.*, 86:2613, 2001.
27. W. Shan, W. Walukiewicz, J. W. Ager III, E. E. Haller, J. F. Geisz, D. H. Friedman, J. M. Olson, and S. R. Kurtz. Band anticrossing in GaInNAs alloys. *Phys. Rev. Lett.*, 82:1221, 1999.

28. P. J. Klar, H. Grüning, W. Heimbrodt, J. Koch, F. Höhnsdorf, W. Stolz, P. M. A. Vicente, and J. Camassel. From N isoelectronic impurities to N-induced bands in the $GaAs_{1-x}N_x$ alloy. *Appl. Phys. Lett.*, 76:3439, 2000.
29. E. P. O'Reilly, A. Lindsay, S. Tomić, and M. Kamal-Saadi. Tight-binding and $\mathbf{k}\cdot\mathbf{p}$ models for the electronic structure of Ga(In)NAs and related alloys. *Semicond. Sci. Technol.*, 17:870, 2002.
30. A. Lindsay and E. P. O'Reilly. Theory of enhanced bandgap non-parabolicity in GaN_xAs_{1-x} and related alloys. *Solid State Commun.*, 112:443, 1999.
31. A. Lindsay and E. P. O'Reilly. Influence of nitrogen resonant states on the electronic structure of GaN_xAs_{1-x}. *Solid State Commun.*, 118:313, 2001.
32. A. Lindsay, S. Tomić, and E. P. O'Reilly. Derivation of a 10-band $\mathbf{k}\cdot\mathbf{p}$ model for dilute nitride semiconductors. *Solid State Commun.*, 47:443, 2003.
33. S. Tomić, E. P. O'Reilly, R. Fehse, S. J. Sweeney, A. R. Adams, A. D. Andreev, S. A. Choulis, T. J. C. Hosea, and H. Riechert. Theoretical and experimental analysis of 1.3-μm InGaNAs/GaAs lasers. *IEEE J. Sel. Top. Quantum Electron.*, 9:1228, 2003.
34. S. B. Healy and E. P. O'Reilly. Influence of electrostatic confinement on optical gain in GaInNAs quantum-well lasers. *IEEE J. Quant. Electron.*, 42:608, 2006.
35. C. A. Broderick, M. Usman, and E. P. O'Reilly. Derivation of 12- and 14-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonians for dilute bismide and bismide–nitride alloys. *Semicond. Sci. Technol.*, 28:125025, 2013.
36. I. A. Buyanova, G. Pozina, P. N. Hai, W. M. Chen, H. P. Xin, and C. W. Tu. Type I band alignment in the GaN_xAs_{1-x}/GaAs quantum wells. *Phys. Rev. B*, 64:033303, 2000.
37. P. N. Hai, W. M. Chen, I. A. Buyanova, H. P. Xin, and C. W. Tu. Direct determination of electron effective mass in GaNAs/GaAs quantum wells. *Appl. Phys. Lett.*, 77:1843, 2000.
38. F. Masia, A. Polimeni, G. B. H. von Högersthal, M. Bissiri, M. Capizzi, P. J. Klar, and W. Stolz. Early manifestation of localization effects in diluted Ga(AsN). *Appl. Phys. Lett.*, 82:4474, 2003.
39. J. D. Perkins, A. Mascarenhas, Y. Zhang, J. F. Geisz, D. J. Friedman, J. M. Olson, and S. R. Kurtz. Nitrogen-activated transitions, level repulsion, and band gap reduction in $GaAs_{1-x}N_x$ with $x < 0.03$. *Phys. Rev. Lett.*, 82:3312, 1999.
40. E. P. O'Reilly, A. Lindsay, and S. Fahy. Theory of the electronic structure of dilute nitride alloys: Beyond the band-anti-crossing model. *J. Phys. Condens. Matter*, 16:S3257, 2004.
41. G. Hendorfer and J. Schneider. g-factor and effective mass anisotropies in pseudomorphic strained layers. *Semicond. Sci. Technol.*, 6:595, 1991.
42. P. Le Jeune, D. Robart, X. Marie, T. Amand, M. Brousseau, J. Barrau, V. Kalevich, and D. Rodichev. Anisotropy of the electron Landé g-factor in quantum wells. *Semicond. Sci. Technol.*, 12:380, 1997.
43. M. Oestreich, S. Hallstein, and W. W. Rühle. Spin quantum beats in semiconductors. *IEEE J. Sel. Top. Quantum Electron.*, 2:747, 1996.
44. A. Lindsay and E. P. O'Reilly. Unification of the band anticrossing and cluster-state models of dilute nitride semiconductor alloys. *Phys. Rev. Lett.*, 93:196402, 2004.
45. R. Fehse, S. Tomić, A. R. Adams, S. J. Sweeney, E. P. O'Reilly, A. D. Andreev, and H. Riechert. A quantitative study of radiative, Auger, and defect related recombination in 1.3-μm GaInNAs-based quantum-well lasers. *IEEE J. Sel. Top. Quantum Electron.*, 8:801, 2002.
46. M. Ehrhardt and T. Koprucki, editors. *Multiband Effective Mass Approximations: Advanced Mathematical Models and Numerical Techniques*. Berlin: Springer, 2014.
47. A. T. Meney, B. Gonul, and E. P. O'Reilly. Evaluation of various approximations used in the envelope-function method. *Phys. Rev. B*, 50:10893, 1994.
48. F. Szmulowicz. Derivation of a general expression for the momentum matrix elements within the envelope-function approximation. *Phys. Rev. B*, 51:1613, 1995.
49. C.-S. Chang, S. L. Chuang, J. R. Minch, W.-C. W. Fang Y. K. Chen, and T. Tanbun-Ek. Amplified spontaneous emission spectroscopy in strained quantum well lasers. *IEEE J. Sel. Top. Quantum Electron.*, 1:1100, 1995.

50. J. Hader, S. W. Koch, and J. V. Moloney. Microscopic theory of gain and spontaneous emission in GaInNAs laser material. *Solid State Electron.*, 47:513, 2003.
51. J. Hader, J. V. Moloney, and S. W. Koch. Microscopic evaluation of spontaneous emission- and Auger-processes in semiconductor lasers. *IEEE J. Quant. Electron.*, 41:1217, 2005.
52. S. W. Koch and W. W. Chow. *Semiconductor–Laser Fundamentals: Physics of the Gain Materials*. Berlin: Springer, 1999.
53. M. Hofmann, A. Wagner, C. Ellmers, C. Schlichenmeier, S. Schäfer, F. Höhnsdorf, J. Koch, W. Stolz, S. W. Koch, W. W. Rühle, J. Hader, J. V. Moloney, E. P. O'Reilly, B. Borchert, A. Y. Egorov, and H. Riechert. Gain spectra of (GaIn)(NAs) laser diodes for the 1.3-µm-wavelength regime. *Appl. Phys. Lett.*, 78:3009, 2001.
54. S. Tomić and E. P. O'Reilly. Gain characteristics of ideal dilute nitride quantum well lasers. *Physica E*, 13:1102, 2002.
55. E. P. O'Reilly and A. R. Adams. Band-structure engineering in strained semiconductor lasers. *IEEE J. Quant. Electron.*, 30:366, 1994.
56. A. R. Adams. Strained-layer quantum-well lasers. *IEEE J. Sel. Top. Quant. Electron.*, 17:1364, 2011.
57. S. Tomić and E. P. O'Reilly. Optimization of material parameters in 1.3µm InGaAsN–GaAs lasers. *IEEE Photon. Tech. Lett.*, 15:6, 2003.
58. P. J. Klar, H. Grüning, W. Heimbrodt, G. Weiser, J. Koch, K. Volz, W. Stolz, S. W. Koch, S. Tomić, S. A. Choulis, T. J. C. Hosea, E. P. O'Reilly, M. Hofmann, J. Hader, and J. V. Moloney. Interband transitions of quantum wells and device structures containing Ga(N,As) and (Ga,In)(N,As). *Semicond. Sci. Technol.*, 17:830, 2002.
59. E. P. O'Reilly. Valence band engineering in strained-layer structures. *Semicond. Sci. Technol.*, 4:121, 1989.
60. N. Tansu, N. J. Kirsch, and L. J. Mawst. Low-threshold-current-density 1300-nm dilute-nitride quantum well lasers. *Appl. Phys. Lett.*, 81:2523, 2002.
61. N. Tansu and L. J. Mawst. Low-threshold strain-compensated InGaAs(N) ($\lambda = 1.19\text{-}1.31$) quantum-well lasers. *IEEE Photon. Technol. Lett.*, 14:444, 2002.
62. N. Tansu, J.-Y. Yeh, and L. J. Mawst. Low-threshold 1317-nm InGaAsN quantum-well lasers with GaAsN barriers. *Appl. Phys. Lett.*, 83:2512, 2003.
63. N. Tansu, J.-Y. Yeh, and L. J. Mawst. Experimental evidence of carrier leakage in InGaAsN quantum-well lasers. *Appl. Phys. Lett.*, 83:2112, 2003.
64. S. Seki, H. Oohashi, H. Sugiura, T. Hirono, and K. Yokoyama. Study on the dominant mechanisms for the temperature sensitivity of threshold current in 1.3-µm InP-based strained-layer quantum-well lasers. *IEEE J. Quant. Electron.*, 32:1478, 1996.
65. J. W. Pan and J. I. Chyi. Theoretical study of the temperature dependence of 1.3-µm AlGaInAs-InP multiple-quantum-well lasers. *IEEE J. Quant. Electron.*, 32:2133, 1996.
66. J. Hader, S. W. Koch, J. V. Moloney, and E. P. O'Reilly. Gain in 1.3 µm materials: InGaNAs and InGaPAs semiconductor quantum-well lasers. *Appl. Phys. Lett.*, 77:630, 2000.
67. A. F. Phillips, S. J. Sweeney, A. R. Adams, and P. J. A. Thijs. The temperature dependence of 1.3- and 1.5-µm compressively strained InGaAs(P) MQW semiconductor lasers. *IEEE J. Sel. Top. Quantum Electron.*, 5:401, 1999.
68. T. R. Chen, B. Chang, L. C. Chiu, K. L. Yu, S. Margalit, and A. Yariv. Carrier leakage and temperature dependence of InGaAsP lasers. *IEEE J. Sel. Top. Quantum Electron.*, 43:217, 1983.
69. J. Barrau, T. Amand, M. Brousseau, R. J. Simes, and L. Goldstein. Induced electrostatic confinement of the electron gas in tensile strained InGaAs/InGaAsP quantum well lasers. *J. Appl. Phys.*, 8:801, 2002.
70. A. R. Adams, M. Asada, Y. Suematsu, and S. Arai. The temperature dependence of the efficiency and threshold current of $\mathrm{In}_{1-x}\mathrm{Ga}_x\mathrm{As}_y\mathrm{P}_{1-y}$ lasers related to intervalence band absorption. *Jpn. J. Appl. Phys.*, 19:L621, 1980.
71. N. K. Dutta and R. J. Nelson. The case for Auger recombination in $\mathrm{In}_{1-x}\mathrm{Ga}_x\mathrm{As}_y\mathrm{P}_{1-y}$. *Appl. Phys. Lett.*, 38:407, 1981.

72. A. D. Andreev and E. P. O'Reilly. Theoretical study of Auger recombination in a GaInNAs 1.3 μm quantum well laser structure. *Appl. Phys. Lett.*, 84:1826, 2004.
73. T. Higashi, S. J. Sweeney, A. F. Phillips, A. R. Adams, E. P. O'Reilly, T. Uchida, and T. Fujii. Experimental analysis of temperature dependence in 1.3μm AlGaInAs-InP strained MQW lasers. *IEEE J. Sel. Top. Quantum Electron.*, 5:413, 1999.
74. S. J. Sweeney. Novel experimental techniques for semiconductor laser characterisation and optimisation. *Phys. Scripta*, T114:152, 2004.
75. A. Haug. Relations between the T_0 values of bulk and quantum well GaAs. *Appl. Phys. B*, 44:151, 1987.
76. G. P. Agrawal and N. K. Dutta. *Long-Wavelength Semiconductor Lasers*. New York, NY: Van Nostrand Reinhold, 1986.
77. J. M. Pikal, C. S. Menoni, H. Temkin, P. Thiagarajan, and G. Y. Robinson. Carrier lifetime and recombination in long-wavelength quantum-well lasers. *IEEE J. Sel. Top. Quantum Electron.*, 5:613, 1999.
78. E. P. O'Reilly, G. Jones, M. Silver, and A. R. Adams. Determination of gain and loss mechanisms in semiconductor lasers using pressure techniques. *Phys. Status Solidi B*, 198:363, 1996.
79. S. J. Sweeney, R. Fehse, A. R. Adams, and H. Riechert. Intrinsic temperature sensitivities of 1.3 μm GaInNAs/GaAs-, InGaAsP/InP- and AlGaInAs/InP-based semiconductor lasers. In *Proceedings of the 16th Annual Meeting of the IEEE Lasers and Electro-Optics Society*, 2003.
80. Y. Q. Wei, M. Sadeghi, S. M. Wang, P. Modh, and A. Larsson. High performance 1.28 μm GaInNAs double quantum well lasers. *Electron. Lett.*, 41:1328, 2005.
81. S. M. Wang, G. Adolfsson, H. Zhao, Y. Q. Wei, J. S. Gustavsson, M. Sadeghi, and A. Larsson. High performance 1.3 μm GaInNAs quantum well lasers on GaAs. *Proc. SPIE*, 6909:690905, 2008.
82. J. S. Gustavsson, Y. Q. Wei, M. Sadeghi, S. M. Wang, and A. Larsson. 10 Gbit/s modulation of 1.3 μm GaInNAs lasers up to 110 °C. *Electron. Lett.*, 42:925, 2006.
83. Y. Q. Wei, J. S. Gustavsson, Å. Haglund, P. Modh, M. Sadeghi, S. M. Wang, and A. Larsson. High-frequency modulation and bandwidth limitations of GaInNAs double-quantum-well lasers. *Appl. Phys. Lett.*, 88:051103, 2006.
84. E. P. O'Reilly, N. Balkan, I. A. Buyanova, X. Marie, and H. Riechert. Dilute nitride and related mismatched semiconductor alloys. *IEE Proc. Optoelectron.*, 151:245, 2004.
85. G. Jaschke, R. Averbeck, L. Geelhaar, and H. Riechert. Low threshold InGaAsN/GaAs lasers beyond 1500 nm. *J. Cryst. Growth*, 278:224, 2005.
86. X. Yang, M. J. Jurkovic, J. B. Heroux, and W. I. Wang. Molecular beam epitaxy growth of InGaAsN:Sb-GaAs quantum wells for long-wavelength semiconductor lasers. *Appl. Phys. Lett.*, 75:178, 1999.
87. S. R. Bank, H. Bae, L. L. Goddard, H. B. Yuen, M. A. Wistey, R. Kudraweic, and J. S. Harris. Recent progress on 1.55-μm dilute-nitride lasers. *IEEE J. Quant. Electron.*, 43:773, 2007.
88. V.-M. Korpijärvi, E. L. Kantola, T. Leinonen, R. Isoaho, and M. Guina. Monolithic GaInNAsSb/GaAs VECSEL operating at 1550 nm. *IEEE J. Sel. Top. Quant. Electron.*, 21:1700705, 2015.
89. M. Silver, E. P. O'Reilly, and A. R. Adams. Determination of the wavelength dependence of Auger recombination in long-wavelength quantum-well semiconductor lasers using hydrostatic pressure. *IEEE J. Quant. Electron.*, 33:1557, 1997.
90. S. R. Bank, L. L. Goddard, M. A. Wistey, H. B. Yuen, and J. S. Harris. On the temperature sensitivity of 1.5-μm GaInNAsSb lasers. *IEEE J. Sel. Top. Quant. Electron.*, 11:1089, 2005.
91. K. Uesugi, I. Suemune, T. Hasegawa, T. Akutagawa, and T. Nakamura. Temperature dependence of band gap energies of GaAsN alloys. *Appl. Phys. Lett.*, 76:1285, 2000.
92. A. Polimeni, M. Capizzi, M. Geddo, M. Fischer, M. Reinhardt, and A. Forchel. Effect of nitrogen on the temperature dependence of the energy gap in $In_xGa_{1-x}As_{1-y}N_y$/GaAs single quantum wells. *Phys. Rev. B*, 63:195320, 2001.
93. I. Suemune, K. Uesugi, and W. Walukiewicz. Role of nitrogen in the reduced temperature dependence of band-gap energy in GaNAs. *Appl. Phys. Lett.*, 77:3021, 2000.

94. D. Alexandropoulos, M. J. Adams, Z. Hatzopoulos, and D. Syvridis. Proposed scheme for polarization insensitive GaInNAs-based semiconductor optical amplifiers. *IEEE J. Quant. Electron.*, 41:817, 2005.
95. X. Sun, N. Vogiatzis, and J. M. Rorison. Theoretical study on dilute nitride 1.3 μm quantum well semiconductor optical amplifiers: Incorporation of N compositional fluctuations. *IEEE J. Quant. Electron.*, 49:811, 2013.
96. J. Hashimoto, K. Koyama, T. Katsuyama, Y. Iguchi, T. Yamada, S. Takagishi, M. Ito, and A. Ishida. 1.3 μm travelling-wave GaInNAs semiconductor optical amplifier. In *Proceedings of Optical Amplifiers and Their Applications*, p. WB3, 2003.
97. J. Pozo, N. Vogiatzis, O. Ansell, P. J. Heard, J. M. Rorison, P. Tuomisto, J. Konttinen, M. Saarinen, C. Peng, J. Viheriälä, T. Leinonen, and M. Pessa. Fabrication and characterization of GaInNAs/GaAs semiconductor optical amplifiers. *Proc. SPIE*, 6997:69970C, 2008.
98. M. Dumitrescu, M. Wolf, K. Schulz, Y. Q. Wei, G. Adolfsson, J. Gustavsson, J. Bengtsson, M. Sadeghi, S. M. Wang, A. Larsson, J. Lim, E. Larkins, P. Melanen, P. Uusimaa, and M. Pessa. 10 Gb/s uncooled dilute-nitride optical transmitters operating at 1.3 μm. In *Proceedings of the Optical Fiber Communication Conference*, p. OWJ7, 2009.
99. D. Fitsios, G. Giannoulis, N. Iliadis, V.-M. Korpijärvi, J. Viheriälä, A. Laakso, S. Dris, M. Spyropoulou, H. Avramopoulos, G. T. Kanellos, N. Pleros, and M. Guina. High gain 1.3-μm GaInNAs SOA with fast-gain dynamics and enhanced temperature stability. *Proc. SPIE*, 8982:898208, 2014.
100. D. Fitsios, G. Giannoulis, V.-M. Korpijärvi, J. Viheriälä, A. Laakso, N. Iliadis, S. Dris, M. Spyropoulou, H. Avramopoulos, G. T. Kanellos, N. Pleros, and M. Guina. High-gain 1.3 μm GaInNAs semiconductor optical amplifier with enhanced temperature stability for all-optical signal processing at 10 Gb/s. *Appl. Optics*, 54:47, 2015.
101. D. Alexandropoulos and M. J. Adams. GaInNAs-based vertical cavity semiconductor optical amplifiers. *J. Phys. Condens. Matter*, 16:S3345, 2004.
102. A. H. Clark, S. Calvez, N. Laurand, R. Macaluso, H. D. Sun, M. D. Dawson, T. Jouhti, J. Kontinnen, and M. Pessa. 1.3 μm GaInNAs optically-pumped vertical cavity semiconductor optical amplifier. *IEEE J. Quant. Electron.*, 9:878, 2004.
103. N. Laurand, S. Calvez, M. D. Dawson, A. C. Bryce, T. Jouhti, J. Konttinen, and M. Pessa. Performance comparison of GaInNAs vertical-cavity semiconductor optical amplifiers. *IEEE J. Quant. Electron.*, 41:642, 2005.
104. N. Laurand, S. Calvez, M. D. Dawson, and A. E. Kelly. Index and gain dynamics of optically pumped GaInNAs vertical-cavity semiconductor optical amplifiers. *Appl. Phys. Lett.*, 87:231115, 2005.
105. N. Laurand, S. Calvez, M. D. Dawson, and A. E. Kelly. Slow-light in a vertical-cavity semiconductor optical amplifier. *Opt. Express*, 9:6858, 2006.
106. F. A. I. Chaqmaqchee and N. Balkan. Gain studies of 1.3-μm dilute nitride HELLISH-VCSOA for optical communications. *Nanoscale Res. Lett.*, 7:526, 2012.
107. S. B. Lisesivdin, N. A. Khan, S. Mazzucato, N. Balkan, M. J. Adams, V.-M. Korpijärvi, M. Guina, G. Mezosi, and M. Sorel. Optical gain in 1.3μm electrically driven dilute nitride VCSOAs. *Nanoscale Res. Lett.*, 9:22, 2014.
108. T. Piwonski, J. Pulka, G. Madden, G. Huyet, J. Houlihan, J. Pozo, N. Vogiatzis, P. Ivanov, J. M. Rorison, P. J. Barrios, and J. A. Gupta. Ultrafast gain and refractive index dynamics in GaInNAsSb semiconductor optical amplifiers. *J. Appl. Phys.*, 106:083104, 2009.
109. G. Giannoulis, V.-M. Korpijärvi, N. Iliadis, J. Mäkelä, J. Viheriälä, D. Apostolopoulos, M. Guina, and H. Avramopoulos. Bringing high-performance GaInNAsSb/GaAs SOAs to true data applications. *IEEE Photon. Tech. Lett.*, 27:1691, 2015.
110. G. Giannoulis, N. Iliadis, D. Apostolopoulos, P. Bakopoulos, H. Avramopoulos, V.-M. Korpijärvi, J. Mäkelä, J. Viheriälä, and M. Guina. 1.55-μm dilute nitride SOAs with low temperature sensitivity

for coolerless on-chip operation. In *Proceedings of the IEEE International Conference on Electronics, Circuits, and Systems*, p. 677, 2015.

111. V. Lordi, H. B. Yuen, S. R. Bank, and J. S. Harris. Quantum-confined Stark effect of GaInNAs(Sb) quantum wells at 1300–1600 nm. *Appl. Phys. Lett.*, 85:902, 2004.
112. T. Fujisawa, M. Arai, and F. Kano. Many-body design of highly strained GaInNAs electroabsorption modulators on GaInNAs ternary substrates. *J. Appl. Phys.*, 107:093107, 2010.
113. J. Hashimoto, K. Koyama, T. Ishizuka, Y. Tsuji, K. Fujii, T. Yamada, and T. Katsuyama. Electroabsorption effect of GaInNAs in waveguiding structure. *Jpn. J. Appl. Phys.*, 48:122403, 2009.
114. A. R. Adams. Band structure engineering to control impact ionisation and related high-field processes. *Electron. Lett.*, 40:1086, 2004.
115. S. L. Tan, W. M. Soong, J. E. Green, M. J. Steer, S. Zhang, L. J. J. Tan, J. S. Ng, I. P. Marko, S. J. Sweeney, A. R. Adams, J. Allam, and J. P. R. David. Experimental evaluation of impact ionization in dilute nitride GaInNAs diodes. *Appl. Phys. Lett.*, 103:102101, 2013.
116. G. Read, I. P. Marko, N. Hossain, and S. J. Sweeney. Physical properties and characteristics of III–V lasers on silicon. *IEEE J. Sel. Top. Quantum Electron.*, 21:1502208, 2015.
117. B. Kunert, K. Volz, and W. Stolz. Properties of laser applications of the GaP-based (GaNAsP)-material system for integration to Si substrates. *Springer Ser. Mat. Sci.*, 105:317, 2008.
118. N. Hossain, T. J. C. Hosea, S. J. Sweeney, S. Liebich, M. Zimprich, K. Volz, B. Kunert, and W. Stolz. Band structure properties of novel $B_xGa_{1-x}P$ alloys for silicon integration. *J. Appl. Phys.*, 110:063101, 2011.
119. S. Liebich, M. Zimprich, A. Beyer, C. Lange, D. J. Franzbach, S. Chatterjee, N. Hossain, S. J. Sweeney, K. Volz, B. Kunert, and W. Stolz. Laser operation of Ga(NAsP) lattice-matched to (001) silicon substrate. *Appl. Phys. Lett.*, 99:071109, 2011.
120. J. Chamings, A. R. Adams, S. J. Sweeney, B. Kunert, K. Volz, and W. Stolz. Temperature dependence and physical properties of Ga(NAsP)/GaP semiconductor lasers. *Appl. Phys. Lett.*, 93:201108, 2008.
121. J. Chamings, S. Ahmed, A. R. Adams, S. J. Sweeney, V. A. Odnoblyudov, C. W. Tu, B. Kunert, and W. Stolz. Band anti-crossing and carrier recombination in dilute nitride phosphide based lasers and light emitting diodes. *Phys. Status Solidi B*, 246:527, 2009.
122. J. Wu, W. Walukiewicz, and E. E. Haller. Band structure of highly mismatched semiconductor alloys: Coherent potential approximation. *Phys. Rev. B*, 65:233210, 2001.
123. M. Seifikar, E. P. O'Reilly, and S. Fahy. Analysis of band-anticrossing model in GaNAs near localised states. *Phys. Status Solidi B*, 248:1176, 2011.
124. L. Ivanova, H. Eisele, M. P. Vaughan, Ph. Ebert, A. Lenz, R. Timm, O. Schumann, L. Geelhaar, M. Dähne, S. Fahy, H. Riechert, and E. P. O'Reilly. Direct measurement and analysis of the conduction band density of states in diluted $GaAs_{1-x}N_x$ alloys. *Phys. Rev. B*, 82:161201, 2010.
125. M. P. Vaughan, S. Fahy, E. P. O'Reilly, L. Ivanova, H. Eisele, and M. Dähne. Modelling and direct measurement of the density of states in GaAsN. *Phys. Status Solidi B*, 248:1167, 2011.
126. M. P. Vaughan and B. K. Ridley. Electron-nitrogen scattering in dilute nitrides. *Phys. Rev. B*, 75:195205, 2007.
127. N. Vogiatzis and J. M. Rorison. Single impurity Anderson model and band anti-crossing in the $Ga_{1-x}In_xN_yAs_{1-y}$ material system. *Phys. Status Solidi A*, 205:120, 2008.
128. N. Vogiatzis and J. M. Rorison. Density of states of dilute nitride systems: Calculation of lifetime broadening. *J. Phys. Condens. Matter*, 21:255801, 2009.
129. M. Seifikar, E. P. O'Reilly, and S. Fahy. Self-consistent Green's function method for dilute nitride conduction band structure. *J. Phys. Condens. Matter*, 26:365502, 2014.
130. P. W. Anderson. Localized magnetic states in metals. *Phys. Rev.*, 124:41, 1961.
131. S. Fahy, A. Lindsay, H. Ouerdane, and E. P. O'Reilly. Alloy scattering of n-type carriers in GaN_xAs_{1-x}. *Phys. Rev. B*, 74:035203, 2006.
132. A. P. Jauho and Wilkins J. W. Dilute resonant scatterers in a parabolic band: Density of states as a function of scattering strength. *Phys. Rev. B.*, 28:4628, 1983.

133. M. Seifikar, E. P. O'Reilly, and S. Fahy. Optical absorption of dilute nitride alloys using self-consistent Green's function method. *Nanoscale Res. Lett.*, 9:51, 2014.
134. M. Welna, R. Kudraweic, Y. Nabetani, and W. Walukiewicz. Band anticrossing in ZnOSe highly mismatched alloy. *Appl. Phys. Express*, 7:071202, 2014.
135. W. Welna, R. Kudraweic, Y. Nabetani, T. Tanaka, M. Jaquez, O. D. Dubon, K. M. Yu, and W. Walukiewicz. Effects of a semiconductor matrix on the band anticrossing in dilute group II-VI oxides. *Semicond. Sci. Technol.*, 30:085018, 2015.
136. C. A. Stephenson, W. A. O'Brien, M. Qi, M. Penninger, W. F. Schneider, and M. A. Wistey. Band anticrossing in dilute germanium carbides using hybrid density functionals. *J. Electron. Mater.*, 45:2121, 2016.
137. C. A. Stephenson, W. A. O'Brien, M. Penninger, W. F. Schneider, M. Gillet-Kunnath, J. Zajicek, K. M. Yu, R. Kudraweic, R. A. Stillwell, and M. A. Wistey. Band structure of germanium carbides for direct bandgap silicon photonics. *J. Appl. Phys.*, 120:053102, 2016.
138. R. Mouillet, L. de Vaulchier, E. Delporte, Y. Guldner, L. Travers, and J.-C. Harmand. Role of nitrogen in the mobility drop of electrons in modulation-doped GaAsN/AlGaAs heterostructures. *Solid State Commun.*, 126:333, 2003.
139. A. Patanè, G. Allison, L. Eaves, N. V. Kozlova, Q. D. Zhuang, A. Krier, M. Hopkinson, and G. Hill. Electron coherence length and mobility in highly-mismatched III–V-N alloys. *Appl. Phys. Lett.*, 93:252106, 2008.
140. A. Patanè, A. Ignatov, D. Fowler, O. Makarovsky, L. Eaves, L. Geelhaar, and H. Riechert. Hot-electrons and negative differential conductance in $GaAs_{1-x}N_x$. *Phys. Rev. B*, 72:033312, 2005.
141. A. Ignatov, A. Patanè, O. Makarovsky, and L. Eaves. Terahertz response of hot electrons in dilute nitride Ga(AsN) alloys. *Appl. Phys. Lett.*, 88:032107, 2006.
142. G. Allison, S. Spasov, A. Patanè, L. Eaves, A. Ignatov, D. K. Maude, M. Hopkinson, and R. Airey. Magnetophonon oscillations in the negative differential conductance of dilute nitride $GaAs_{1-x}N_x$ submicron diodes. *Phys. Rev. B*, 75:115325, 2007.
143. Y. Sun, M. P. Vaughan, A. Agarwal, M. Yilmaz, B. Ulug, A. Ulug, N. Balkan, M. Sopanen, O. Reentilä, M. Mattila, C. Fontaine, and A. Arnoult. Inhibition of negative differential resistance in modulation-doped n-type $Ga_xIn_{1-x}N_yAs_{1-y}$/GaAs quantum wells. *Phys. Rev. B*, 75:205316, 2007.
144. S. Fahy and E. P. O'Reilly. Intrinsic limits on electron mobility in dilute nitride semiconductors. *Appl. Phys. Lett.*, 83:3731, 2003.
145. J. Buckeridge and S. Fahy. Mobility in gated GaN_xAs_{1-x} heterostructures as a probe of nitrogen-related electronic states. *Phys. Rev. B*, 84:144120, 2011.
146. D. L. Young, J. F. Geisz, and T. J. Coutts. Nitrogen-induced decrease of the electron effective mass in $GaAs_{1-x}N_x$ thin films measured by thermomagnetic transport phenomena. *Appl. Phys. Lett.*, 82:1236, 2003.
147. D. Fowler, O. Makarovsky, A. Patanè, L. Eaves, L. Geelhaar, and H. Riechert. Electron conduction in two-dimensional $GaAs_{1-y}N_y$ channels. *Phys. Rev. B*, 69:153305, 2003.
148. M. P. Vaughan and B. K. Ridley. Solution of the Boltzmann equation for calculating Hall mobility in GaN_xAs_{1-x}. *Phys. Rev. B*, 72:075211, 2005.
149. M. Seifikar, E. P. O'Reilly, and S. Fahy. Theory of intermediate- and high-field mobility in dilute nitride alloys. *Phys. Rev. B*, 84:165216, 2011.
150. N. Vogiatzis and J. M. Rorison. Negative differential velocity in ultradilute $GaAs_{1-x}N_x$ alloys. *J. Appl. Phys.*, 109:083720, 2011.
151. N. Vogiatzis and J. M. Rorison. Electron transport in bulk GaAsN. *Phys. Status Solidi B*, 248:1183, 2011.

10

Dilute Bismide Alloys

10.1 Introduction 313
10.2 Electronic and Optical Properties of $GaBi_xAs_{1-x}$ Alloys: Atomistic Theory 316
Impact of Bi Incorporation on the Electronic Structure: Atomistic Theory • Comparison of Atomistic Theory with Experimental Measurements • Outlook for Atomistic Theory
10.3 Modeling Dilute Bismide Band Structure: The **k·p** Method 326
Valence Band-Anticrossing • 12-Band **k·p** Hamiltonian for Dilute Bismide Alloys • Calculation of Band Offsets in Bi-Containing Quantum Wells
10.4 $GaBi_xAs_{1-x}$/(Al)GaAs Quantum Well Lasers 333
Theoretical Model for Dilute Bismide Quantum Well Lasers • Impact of Bi Incorporation on the Band Structure and Gain of Quantum Wells • Optimization of $GaBi_xAs_{1-x}$/(Al)GaAs Laser Structures • Gain Characteristics of $GaBi_xAs_{1-x}$ Quantum Wells at $\lambda = 1.55$ μm • Theory versus Experiment for First-Generation $GaBi_xAs_{1-x}$/(Al)GaAs Lasers
10.5 Emerging Directions in Dilute Bismide Research 345
$In_yGa_{1-y}Bi_xAs_{1-x}$/InP Alloys • $GaBi_xN_yAs_{1-x-y}$/GaAs Alloys • $GaBi_xAs/GaN_yAs_{1-y}$ Type-II Quantum Wells on GaAs • Further Directions: Quaternary Alloys, Type-II Structures, and Narrow-Gap Materials
10.6 Summary and Conclusions 354

Christopher A. Broderick

Igor P. Marko

Eoin P. O′Reilly

and

Stephen J. Sweeney

10.1 Introduction

Dilute bismide alloys are III–V semiconductor alloys containing small fractions of substitutional bismuth (Bi) atoms. Similar to the incorporation of nitrogen (N) to form the dilute nitride alloys described in Chapter 9, dilute bismide alloys are characterized by the fact that Bi acts as an isovalent impurity when incorporated into, e.g. (In)GaAs, to form the (In)$GaBi_xAs_{1-x}$ alloy. Similar to the case of dilute nitride alloys, it is the large differences in size (covalent radius) and chemical properties (electronegativity) between Bi atoms and the group-V atoms they replace that brings about this impurity-like behavior, which in practice means that incorporating Bi at dilute concentrations has a significant impact on the material properties. Dilute bismide alloys, like dilute nitrides, are considered to be highly mismatched semiconductor materials. However, while N incorporation primarily affects the conduction band (CB) structure of the material into which it is incorporated, Bi, being significantly larger and more electropositive than N, primarily affects the valence band (VB). Since several general aspects of the physics of dilute bismide alloys are qualitatively similar to those of dilute nitrides, dilute bismide alloys can be considered a naturally complementary material system to the dilute nitrides.

The dilute bismide alloy $GaBi_xAs_{1-x}$ has several novel electronic properties. For example, substitution of As by Bi causes a rapid reduction in the band gap (E_g) with increasing Bi composition x, by up to 90 meV per % Bi at low x [1–6]. This strong Bi-induced decrease in E_g is accompanied by a strong increase in the VB spin-orbit-splitting energy (Δ_{SO}) [5–7], and the changes in both E_g and Δ_{SO} are characterized by strong, composition-dependent bowing, much like the band gap reduction in GaN_xAs_{1-x} [5,8]. These unusual material properties—among others, which we discuss in Section 10.2—have prompted significant interest in the development of dilute bismide alloys for a range of practical applications, including in semiconductor lasers [9–15] and photodiodes [16–19], as well as photovoltaics [20–23], spintronics [7,24,25], and thermoelectrics [26].

The theoretical modeling of dilute bismide alloys is significantly complicated by the impurity-like behavior of the Bi atoms in the alloy, and the resultant strong perturbation of the electronic structure of the host matrix semiconductor. The presence of alloy disorder due to the formation of pairs and larger clusters of Bi atoms in $GaBi_xAs_{1-x}$ has a particularly significant impact on the electronic structure, meaning that any serious attempt to quantify the consequences of Bi incorporation on the material properties must encompass an atomistic viewpoint. In Section 10.2, we provide an overview of the theory of the unusual properties of dilute bismide alloys, as revealed by experimental measurements, and discuss the various atomistic approaches that have been developed to understand the impact of Bi incorporation on the electronic structure. In particular, we focus on an atomistic tight-binding (TB) model we have developed to study the electronic structure of dilute bismide alloys of (In)GaAs(P) and compare the results of calculations undertaken using this approach to a series of experimental measurements of the electronic [5], optical [6,27,28], and spin properties [29] of $GaBi_xAs_{1-x}$ alloys.

While atomistic models provide significant insight into the properties of Bi-containing alloys, they are too computationally expensive to apply directly on the length scales required to model the properties of optoelectronic devices. This motivates the development of continuum models which are suited to calculating the properties of realistically sized Bi-containing device heterostructures. Through a series of detailed atomistic calculations, we review that the main features of the band structure of $GaBi_xAs_{1-x}$ and related alloys can be understood in terms of a band-anticrossing interaction, much like that considered for the dilute nitrides in Chapter 9. However, for $GaBi_xAs_{1-x}$ alloys the situation is the converse of that in GaN_xAs_{1-x}. In GaN_xAs_{1-x}, the band-anticrossing model describes the N composition-dependent coupling between the extended states of the host matrix (GaAs) CB edge and highly localized N-related impurity states which are resonant with the CB. In $GaBi_xAs_{1-x}$, atomistic calculations reveal that a Bi composition-dependent band-anticrossing interaction occurs between the extended states of the host matrix VB edge and highly localized Bi-related impurity states which are resonant with the VB.

Taking into account the p-like nature of the GaAs VB edge eigenstates, as well as the presence of strong spin-orbit-coupling in the VB, it is clear that even the simplest description of the $GaBi_xAs_{1-x}$ VB structure will be more complicated than that of the GaN_xAs_{1-x} CB structure. It is to this issue that we turn our attention in Section 10.3, where we review how atomistic calculations can be used to directly derive and parameterize valence band-anticrossing (VBAC) models for dilute bismide alloys, thereby removing the parametric ambiguity with which phenomenological models of this type are typically associated. Following this approach, we review the derivation of a 12-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian for the band structure of $(In)GaBi_xAs_{1-x}$ alloys directly from atomistic supercell calculations, and present a general method for calculating the band offsets in (001)-oriented pseudomorphically strained epilayers and quantum wells (QWs).

The practical application which has to date seen the largest research effort and most significant progress has been the development of $GaBi_xAs_{1-x}$ alloys for applications in GaAs-based semiconductor QW lasers [10–12,15,30]. A critical issue with existing InP-based 1.3- and 1.55-μm QW lasers is that their threshold currents and optical (cavity) losses tend to increase strongly with increasing temperature, due largely to a combination of two intrinsic loss mechanisms: non-radiative Auger recombination [31,32] and intervalence band absorption (IVBA) [33]. $GaBi_xAs_{1-x}$ alloys having Bi compositions $x > 10\%$ have been demonstrated to have a band structure in which $\Delta_{SO} > E_g$, while also possessing a band gap $E_g \approx 0.8$ eV

(1.55 μm) [5,6]. It has therefore been proposed that this large VB spin-orbit-splitting energy could lead to the suppression of the dominant "conduction-hole-spin-hole" (CHSH) Auger recombination process in 1.55-μm semiconductor lasers [9,10,12] in which the energy and crystal momentum of an electron–hole pair recombining across the band gap excite a hole from the VB edge to the spin-split-off (SO) band, as depicted schematically in Figure 10.1.

The CHSH Auger process is a loss mechanism which accounts for the majority of the current at threshold in conventional InP-based devices, and which also strongly degrades the stability and efficiency of the device operation above room temperature [31]. Suppression of the CHSH Auger recombination pathway is therefore expected to bring about highly efficient, temperature-stable laser operation, opening the route to uncooled GaAs-based telecom lasers with significantly reduced power consumption and providing large energy savings compared to the InP-based technologies currently deployed in optical communication networks [9–12]. Using dilute bismide alloys as a route to realizing long-wavelength semiconductor lasers on GaAs substrates should also enable the growth of monolithic telecom-wavelength vertical-cavity surface-emitting lasers (VCSELs), thereby bringing the benefits of (Al)GaAs-based distributed Bragg reflectors (DBRs) to 1.55 μm. This is appealing from a technological perspective, because of the possibility of taking advantage of the enhanced carrier and optical confinement offered by (Al)GaAs-based heterostructures, as well as the potential to monolithically integrate long-wavelength semiconductor lasers with high-speed GaAs-based microelectronics.

Considerable progress has been made in developing epitaxial growth of $GaBi_xAs_{1-x}$ alloys, leading to the demonstration of an optically pumped bulk-like laser by Tominaga et al. in 2010 [34]. Room temperature electrically pumped lasing from a $GaBi_xAs_{1-x}$ QW laser was first demonstrated at $x = 2.1\%$ by Ludewig et al. in 2013 [30], and the Bi composition in such laser structures has since extended up to

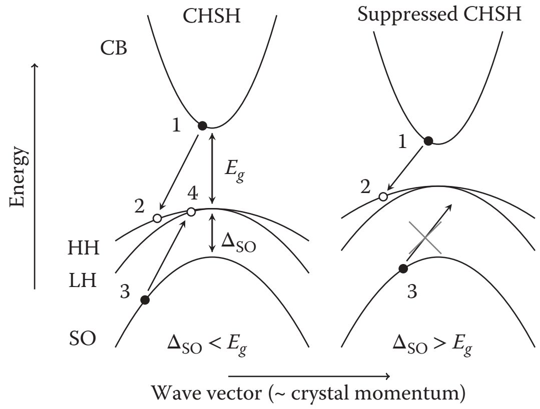


FIGURE 10.1 Left: Schematic illustration of the hot hole-producing CHSH Auger recombination process, in which a CB edge electron (1) recombines with a VB edge hole (2), with the energy released then exciting a VB edge hole (4) to the SO band (3), such that energy and crystal momentum ($\hbar\mathbf{k}$) are conserved. The hot hole generated by this process dissipates its excess energy and crystal momentum primarily via scattering within the lattice, which generates phonons, increasing the lattice temperature and generating significant waste heat. The presence of strong, temperature-dependent CHSH Auger recombination in the materials forming the active region of InP-based 1.3- and 1.55-μm semiconductor lasers means that external cooling equipment is required to extract Auger-generated waste heat, in order to maintain operational stability, but at the cost of significantly increasing power consumption. Right: Schematic illustration of suppressed CHSH Auger recombination, achieved for a band structure in which the spin-orbit-splitting energy SO exceeds the band gap: $\Delta_{SO} > E_g$. In this case, the CHSH process is forbidden by conservation of energy, since the energy made available by an electron–hole pair recombining across the band gap is insufficient to excite a VB edge hole to the SO band. The suppression of IVBA proceeds in a similar manner: for $\Delta_{SO} > E_g$ the energy of a photon emitted by an electron–hole pair recombining across the band gap is insufficient to promote (by optical absorption) an electron from the SO band to a hole state at the VB edge, thereby suppressing photon reabsorption in the active region.

$x = 6.5\%$ by Butkutė et al. [35,36]. However, Bi compositions $x > 10\%$ are required to obtain a band structure having $\Delta_{SO} > E_g$ (and hence suppression of Auger recombination and IVBA processes involving the SO band) in $GaBi_xAs_{1-x}$ alloys. Achieving high-quality $GaBi_xAs_{1-x}$ layers with $x > 10\%$ is challenging due to the difficulty associated with epitaxial growth of Bi-containing alloys (see, e.g., Refs. [37,38] and references therein). While significant progress has been made in developing $GaBi_xAs_{1-x}$ materials and devices over the last 5 years, efforts are ongoing to grow heterostructures with sufficiently high Bi compositions to demonstrate lasing at 1.55 μm, as well as suppression of Auger recombination and IVBA, in this new class of semiconductor alloys.

Given the significant potential of dilute bismide alloys for applications in next-generation semiconductor lasers, our focus throughout much of this chapter is on reviewing the theory of dilute bismide lasers. In Section 10.4 we provide an overview of the theory of GaAs-based dilute bismide QW lasers, using a theoretical model based on the 12-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian derived in Section 10.3. Using this approach, we review (1) the impact of Bi on the band structure, density of states (DOS), and optical gain, (2) pathways toward device optimization at low Bi compositions, (3) trends in the expected device performance as a function of Bi composition, and (4) the potential of $GaBi_xAs_{1-x}$ alloys for the development of highly efficient and temperature-stable 1.55-μm lasers. We also review direct comparisons of theoretical calculations of the (1) threshold characteristics, (2) spontaneous emission (SE), and (3) optical gain, to experimental measurements performed on first-generation $GaBi_xAs_{1-x}$ QW laser devices [15,39].

While the majority of research on dilute bismides has to date centered on $GaBi_xAs_{1-x}$ alloys and laser applications, the potential of alternative Bi-containing alloys for a range of practical applications has begun to drive an increasing diversity of research into related dilute bismide semiconductors. It is to these emerging directions in dilute bismide research that we turn our attention in Section 10.5. We discuss the significant potential offered by the quaternary alloys $In_yGa_{1-y}Bi_xAs_{1-x}$ and $GaBi_xN_yAs_{1-x-y}$—for respective applications in InP-based mid-infrared photonics and multijunction solar cells—due to the broad flexibility of their band structures, associated with the ability to engineer the band gap, spin-orbit-splitting energy, strain, and band offsets over wide ranges [12,13,40–45]. We also review the recent development of $GaBi_xAs_{1-x}/GaN_yAs_{1-y}$ type-II QWs, and demonstrate that this novel class of GaAs-based, strain-balanced heterostructures has great potential to deliver emission/absorption across a broad range of near- and mid-infrared wavelengths on, making them of interest for applications in photonics and photovoltaics [46]. Finally, we provide a brief overview of emerging interest in narrow-gap Bi-containing alloys, which are generating strong interest for applications in mid-infrared photonics, as well as in spintronics.

Overall, this chapter should equip the reader with a comprehensive overview of the theory and simulation of dilute bismide alloys and provide an up-to-date overview of ongoing research on this emerging class of semiconductor materials.

10.2 Electronic and Optical Properties of $GaBi_xAs_{1-x}$ Alloys: Atomistic Theory

In this section, we discuss some of the general properties of dilute bismide alloys, the understanding of which has been strongly informed by theoretical interpretations of experimental measurements. Since Bi acts as an isovalent impurity in (In)GaAs(P) and strongly perturbs the electronic structure, alloy disorder, which is inevitably present in real materials, plays an important role in determining the material properties. This, combined with the associated breakdown of the virtual crystal approximation in Bi-containing alloys, means that atomistic calculations are required in order to provide a quantitative understanding of the impact of Bi incorporation on the electronic structure. As such, we present our discussion in this section within the context of atomistic electronic structure calculations. In Section 10.2.1, we trace how such analyses have contributed strongly to developing the current understanding of this unusual class of semiconductor alloys. Next, in Section 10.2.2, we present and compare to experiment the results of atomistic calculations of a range of material properties for realistic, disordered materials. In Section 10.2.3 we provide a brief outlook for the atomistic theory of dilute bismide alloys, highlighting the tasks and challenges that must be completed and overcome to improve the fundamental understanding of dilute bismides and related emerging classes of highly mismatched semiconductor alloys.

Since $GaBi_xAs_{1-x}$ is the material which has been the subject of most of the research effort on dilute bismide alloys to date, it is to this material that we devote our attention in this section. However, the understanding of the impact of Bi incorporation on the properties of $GaBi_xAs_{1-x}$ alloys has been found to transfer to a range of related materials, so that the theory and material trends we outline in this section can be understood to apply in a general sense to other Bi-containing III–V semiconductor alloys.

10.2.1 Impact of Bi Incorporation on the Electronic Structure: Atomistic Theory

The first atomistic study of the impact of Bi incorporation on the electronic structure of GaAs was undertaken by Janotti et al. [47]. Using a first-principles approach based on density functional theory (DFT), Janotti et al. analyzed the electronic structure of $GaBi_xAs_{1-x}$ alloys, as well as the impact of co-alloying Bi and N in GaAs to form the $GaBi_xN_yAs_{1-x-y}$ quaternary alloy (cf. Section 10.5.2). In Ref. [47] the first detailed calculations for the III–V compound GaBi were also presented. In addition to calculating structural parameters such as the lattice constant and bulk modulus, Janotti et al. presented a calculation of the band structure of bulk GaBi in the zincblende phase. This calculation confirmed the presence of an extremely large spin-orbit-splitting energy in the VB, which first principles calculations indicate to be >2 eV [48]. Furthermore, an unusual (topological) band ordering was identified at the Γ-point in GaBi. As a result of the extremely large spin-orbit coupling, the Γ_{6c} CB edge states were calculated to lie lower in energy than the Γ_{8v} VB edge states, but higher in energy than the Γ_{7v} SO band edge states. As such, GaBi is predicted to be a metallic compound, with its negative band gap $E_g = E(\Gamma_{6c}) - E(\Gamma_{8v}) \simeq -1.5eV$. While the analysis of Janotti et al. correctly suggested that Bi incorporation leads to a reduction of E_g with increasing x in $GaBi_xAs_{1-x}$, the calculated band gap bowing was approximately one-sixth of that calculated for GaN_xAs_{1-x} using the same approach. However, experimental measurements and subsequent theoretical analyses have revealed that replacing 1% of the As atoms in GaAs by Bi (N) causes a reduction of E_g of up to 90 (150) meV, meaning that while the impact of Bi incorporation is not as strong as that of N incorporation, the effects are broadly comparable in magnitude.

The first major advancement in the understanding of the electronic structure of $GaBi_xAs_{1-x}$ came from the work of Zhang et al. [49]. Using first-principles pseudopotential calculations to investigate the electronic structure as a function of Bi composition in a series of ordered $GaBi_xAs_{1-x}$ supercells, Zhang et al. provided a number of key insights into the impact of Bi incorporation on the GaAs electronic structure and also highlighted the differences between Bi and N incorporation in terms of their relative impact on the electronic structure. First, it was calculated that Bi incorporation causes a significant decrease (increase) in E_g (Δ_{SO}). The calculated E_g data for ordered alloy supercells were shown to be in good agreement with experiment for Bi compositions up to $x = 3.125\%$, the largest dilute composition which could be reliably investigated using the small calculational supercells employed. However, no comment on the precise nature of the strong Bi-induced band gap bowing was provided, and no *direct* attempt was made to elucidate its origin. Second, it was demonstrated that a substitutional Bi atom in GaAs forms impurity states which are (1) highly localized about the Bi atomic site, and (2) resonant with the extended states of the GaAs host matrix VB—i.e. lie below the GaAs VB edge in energy. Third, it was demonstrated that application of hydrostatic pressure decreases the energy of the Bi-related resonant states. As such, while the N-related states in the GaN_xAs_{1-x} CB can be pushed into the band gap under the influence of high hydrostatic pressure [50], the Bi-related states in $GaBi_xAs_{1-x}$ are expected to move deeper into the VB with increasing pressure.

The presence of spin-orbit coupling splits the resonant states associated with an isolated Bi impurity into a doublet (lying close in energy to the top of the GaAs VB) and a singlet (lying deep within the VB, due to the large spin-orbit coupling). Zhang et al. focused on the higher-energy Bi-related doublet, $|\psi_{Bi}\rangle$, and investigated the character and evolution of the these states as a function of Bi composition. Following this approach, it was found that the states $|\psi_{\text{Bi}}\rangle$ are formed of a linear combination of host matrix VB states—having wave vectors located along the Δ, Λ, and Σ directions in the Brillouin zone—which are folded back to Γ. This suggests that the Bi-related states, which lie energetically within the VB, form at the expense of a host matrix VB state. This is in contrast to the case of N incorporation in GaAs, in which N is understood to contribute an additional state to the CB. It was shown using a series of ordered $\text{GaBi}_x\text{As}_{1-x}$ supercells containing a single substitutional Bi impurity that the calculated energy E_{Bi} of the doublet states $|\psi_{\text{Bi}}\rangle$ increases strongly with increasing supercell size (decreasing Bi composition), with the trend stabilizing in large supercells containing >2000 atoms where the impurity limit is approached. A significant number of host matrix states with large wave vectors are then folded back to Γ in these large supercells, and are hence available to contribute to the supercell representation of the states $|\psi_{\text{Bi}}\rangle$. As a result, a more accurate representation of the impact of Bi incorporation is obtained in larger supercells. In the impurity limit, Zhang et al. estimated that an isolated substitutional Bi impurity produces resonant doublet states that lie approximately 80 meV below the GaAs VB edge in energy—i.e. lying directly between the GaAs VB and SO band edge states in energy.

While this analysis has played a significant role in informing the current understanding of the fundamental aspects of the $\text{GaBi}_x\text{As}_{1-x}$ electronic structure, the charge-patching method used for larger supercells by Zhang et al. is restricted in the size and type of alloys to which it can be applied. For example, such an approach is not well-suited to analyze the properties of randomly disordered alloys at low composition, since it relies on patching the converged charge densities from small supercells into larger supercells: finite-size effects associated with the periodic boundary conditions employed in the underlying small supercell calculation(s) are implicitly retained, which introduces some degree of spurious Bi-Bi interactions in the large supercell calculation(s). In order to overcome these limitations Usman et al. [5] developed a semi-empirical atomistic TB model, based on the sp^3s^* basis set originally introduced by Vogl et al. [51]. Since it employs a localized basis of atomic orbitals, the TB method is well-suited to study the impact of localized impurities on the electronic structure (as we highlighted and reviewed in our discussion of dilute nitride alloys in Chapter 9). Compared to related atomistic approaches, the TB method is computationally cheap and can be used to efficiently study the electronic structure of large alloy supercells containing upward of thousands of atoms. As described in Chapter 9, the TB method has previously been applied with much success to the study of $\text{GaN}_x\text{As}_{1-x}$ alloys, where it has provided quantitative predictions of several important material properties in real, disordered alloys (see, e.g., Refs. [50], [52] and [53]).

Using the TB method, in Ref. [5], Usman et al. significantly broadened the scope of the analysis undertaken by Zhang et al. to include detailed investigation of (1) coupling between localized Bi-related states and the extended states of the host matrix VB, (2) the nature of the localized impurity states associated with the formation of pairs and larger clusters of Bi atoms in a realistic, disordered alloy, (3) the impact of alloy disorder on the overall electronic structure, (4) trends in the electronic structure across an extended range of Bi compositions, and (5) the nature of the VB structure of $\text{GaBi}_x\text{P}_{1-x}$ alloys. The TB method has been highly successful in elucidating the impact of Bi on the electronic and optical properties of $\text{GaBi}_x\text{As}_{1-x}$ alloys. For example, the calculated evolution of the band gap, spin-orbit-splitting energy, and electron effective g factor [5,29] are in quantitative agreement with experiment across the full composition range for which experimental data are available (cf. Section 10.2.2). Further investigations based on the TB method have also quantified the origin and consequences of the strong band gap bowing in Bi-containing alloys [42,54], as well as the effect of intrinsic alloy disorder on the optical properties of bulk and QW $\text{GaBi}_x\text{As}_{1-x}$ [27,28].

Before describing the TB analysis, it is worthy of note that the prediction by Zhang et al. that Bi forms a resonant state in GaAs, energetically within the VB, is contradictory to the conclusions of Janotti et al. in Ref. [47], whose analysis suggested that a substitutional Bi impurity in GaAs forms a bound state lying approximately 100 meV above the GaAs VB edge in energy. This result of the analysis of Janotti et al. is surprising. It has been known since the 1960s that a substitutional Bi impurity in GaP produces a bound state lying approximately 100 meV above the GaP VB edge in energy [55]. Since the difference in energy between the 3*p* valence orbitals of P and the 6*p* valence orbitals of Bi is larger than that between the latter and the energy of the 4*p* valence orbitals of As, it is expected that Bi is less likely to form a bound impurity state in GaAs than in GaP, with the energy of the Bi-related impurity states in GaAs expected to be lower than those in GaP (relative to the host matrix VB edge) [5]. That Bi forms a bound state in $\text{GaBi}_x\text{As}_{1-x}$ is therefore highly unlikely, given the known energy of the Bi-related states in GaP:Bi and the fact that the corresponding states in GaAs:Bi are expected to be at lower energy based on well-established chemical trends.

It has been proposed [4,54] that a band-anticrossing approach, similar to that for the $\text{GaN}_x\text{As}_{1-x}$ CB, can be applied to describe the VB structure in $\text{GaBi}_x\text{As}_{1-x}$ alloys. However, this interpretation of the VB structure has generated to a degree of controversy [54,56]. In order to confirm the presence of a Bi-induced VBAC interaction, Usman et al. used the TB method to investigate the validity of this approach for the VB structure of ordered $\text{GaBi}_x\text{P}_{1-x}$ and $\text{GaBi}_x\text{As}_{1-x}$ alloys by directly analyzing the Bi-related doublet states $|\psi_{\text{Bi}}\rangle$ [5]. For the case of a substitutional Bi atom in GaP the analysis and interpretation are straightforward, since the Bi-related states lie energetically within the GaP band gap. By explicitly constructing the states $|\psi_{\text{Bi}}\rangle$ in a series of ordered $\text{GaBi}_x\text{P}_{1-x}$ supercells (cf. Section 10.3.1), and by directly computing the strength with which these states couple to the GaP VB edge as a function of Bi composition, it was found that (1) in the impurity limit, an isolated, substitutional Bi impurity in GaP forms bound states which are highly localized about the Bi atomic site and lie approximately 120 meV above the GaP VB edge in energy, and (2) the coupling of the states $|\psi_{\text{Bi}}\rangle$ to the GaP VB edge states is precisely of the form $V_{\text{Bi}} = \beta\sqrt{x}$, as in $\text{GaN}_x\text{As}_{1-x}$ and $\text{GaN}_x\text{P}_{1-x}$ alloys [5]. The calculated Bi bound state energy for GaP:Bi in the impurity limit is in good agreement with the experimentally measured value of Ref. [55], and the TB analysis explicitly confirms the presence of a VBAC interaction in ordered $\text{GaBi}_x\text{P}_{1-x}$ alloys [5].

In order to examine the validity of the VBAC model for $\text{GaBi}_x\text{As}_{1-x}$, Usman et al. repeated the impurity state analysis described above for $\text{GaBi}_x\text{P}_{1-x}$ [5,54], and reached broadly similar conclusions to those of Zhang et al. [49]: (1) a substitutional Bi atom in $\text{GaBi}_x\text{As}_{1-x}$ forms a set of impurity states that are resonant with the extended states of the GaAs VB, (2) these Bi-related states are highly localized about the Bi atomic site, and (3) the calculated energy E_{Bi} of these Bi-related states varies strongly with supercell size (emphasizing the role of the finite-size effects described above), and converges in the impurity limit, a prediction which is qualitatively the same as, and quantitatively very close to, that of Zhang et al. The difference in the energies E_{Bi} calculated using the first principles pseudopotential and TB methods is likely related in part to the treatment of the supercells: the pseudopotential calculations rely on patching the converged charge density from a 64-atom supercell into larger supercells in order to reach the impurity limit, while the TB method treats all supercells directly. This analysis predicts that Bi forms resonant states in $\text{GaBi}_x\text{As}_{1-x}$ which lie approximately 180 meV below the GaAs VB edge in energy in the impurity limit. Usman et al. further found that the coupling between the Bi-related resonant states and the GaAs VB edge states in $\text{GaBi}_x\text{As}_{1-x}$ varies with Bi composition as $V_{\text{Bi}} = \beta\sqrt{x}$ in ordered alloys, explicitly confirming that the alloy VB edge states can be described in terms of a VBAC interaction. Recent experimental studies have provided direct evidence of the presence of Bi-related resonant states lying energetically within the $\text{GaBi}_x\text{As}_{1-x}$ VB [57,58]. These detailed measurements support the general conclusions of the theoretical calculations of Refs. [5], [49] and [54], and provide direct experimental support for this interpretation of the VB structure of $\text{GaBi}_x\text{As}_{1-x}$ alloys.

On the basis of DFT calculations, Deng et al. [56] suggested that the VBAC approach breaks down at higher Bi compositions, where alloy disorder effects become important and lead to strong modifications of the VB structure resulting from interactions between neighboring Bi atoms. Detailed analysis of this "band broadening" concept by Virkkala et al. [59] showed that a model Hamiltonian considering Bi–Bi interactions is indeed capable of producing a significant band gap with increasing Bi composition in disordered $\text{GaBi}_x\text{As}_{1-x}$ alloys. However, this approach has two significant shortcomings. First, the predicted decrease in the $\text{GaBi}_x\text{As}_{1-x}$ band gap is linear in x, meaning that the strong, composition-dependent bowing of the band gap observed in experimental measurements is not captured by the band broadening interpretation of the band structure. Second, the proposed model requires extraction of supercell- and composition-dependent parameters from DFT calculations in order to define the alloy band gap, which limits predictive power. Furthermore, it has been demonstrated [54,60] that the conclusions of Deng et al. in Ref. [56] were based on a misinterpretation of the VB ordering in the small $GaBi_xAs_{1-x}$ supercells studied. As such, while the strict validity of the VBAC approach must necessarily break down in a disordered alloy, it is noteworthy that the predictions of the model are in quantitative agreement with experimental measurements of E_g and Δ_{SO} in $GaBi_xAs_{1-x}$ across the full range of compositions for which data are available (cf. Section 10.3.2). Detailed investigations suggest that the VBAC model describes well the band edge states in $GaBi_xAs_{1-x}$ alloys, but breaks down for states lying deeper in the VB, particularly in the vicinity of the SO band [29,61]. As we review in Section 10.4, this means that a VBAC approach provides a simple and accurate means by which to describe the properties of semiconductor lasers and related optoelectronic devices, where we are interested only in states lying energetically in the immediate vicinity of the CB and VB edges.

The understanding of the $GaBi_xAs_{1-x}$ electronic structure developed through atomistic calculations and experimental measurements then describes that (1) Bi introduces highly localized impurity states which are resonant with the GaAs VB, (2) the main features of the band edge states, and hence the alloy band structure, can be effectively described in terms of a VBAC interaction between localized Bi-related doublet states (lying between the GaAs VB and SO band edges in energy) and the extended states of the GaAs VB edge, (3) Bi-related singlet states, which lie much deeper (by several eV) within the VB, play a negligible role in determining the details of the electronic structure, and (4) the CB and SO bands in $GaBi_xAs_{1-x}$ are relatively unperturbed by Bi incorporation—their evolution with increasing Bi composition can be described in a conventional manner using, e.g. the virtual crystal approximation. Given the strong impact of alloy disorder on the electronic properties, interpretation of the details of the $GaBi_xAs_{1-x}$ band structure at higher Bi compositions then remains a somewhat open question: detailed and systematic analysis of the relative impact of As–Bi (VBAC-like) and Bi–Bi (band broadening) interactions on the band structure at higher compositions—building upon that already present in the literature—is required to provide definitive insight. However, as we describe throughout this chapter, describing the impact of Bi incorporation via a simple VBAC approach can be used to analyze the VB edge states in $(In)GaBi_xAs_{1-x}$ alloys with a degree of accuracy that is sufficient to describe and predict the optical properties of semiconductor lasers and related devices.

Since our focus in this chapter is on developing theory with a view to the modeling of photonic devices, we have concentrated in this overview on the fundamental features of the dilute bismide electronic structure which are relevant for such applications (primarily the evolution of the band edge energies and character of the band edge eigenstates with Bi composition). We note there have been several additional studies which have either further explored the electronic structure, or which have focused on other aspects of the interesting and unusual properties of dilute bismide alloys. For example, further DFT-based calculations of the electronic structure of $GaBi_xAs_{1-x}$ and related Bi-containing alloys have been carried out by Mbarki et al. [62], and more recently by Polak et al. [63,64]. Calculations of the structural, elastic, and vibrational properties of III-Bi compounds have been undertaken by Ferhat et al. [65,66], and theoretical investigations of the thermo-dynamics and kinetics of the growth of $GaBi_xAs_{1-x}$ alloys have been carried out by Morgan et al. [67,68] and by Punkinnen et al. [69,70]. We direct the reader to the references provided for further information on these and related aspects of the physics and chemistry of Bi-containing compounds and alloys.

10.2.2 Comparison of Atomistic Theory with Experimental Measurements

Having outlined the general features of the electronic structure of $GaBi_xAs_{1-x}$ alloys, we now turn our attention to direct comparisons between theory and experiment for this material system. We reserve detailed discussion of the impact of isolated Bi impurities and quantitative analysis of the VBAC model until Section 10.3.1, and focus here on disordered alloys. The TB method described in Section 10.2.1 provides a particularly useful framework within which to analyze such systems, since it enables direct calculation of the electronic structure in large supercells containing several thousand atoms, thereby allowing alloy disorder effects *at dilute compositions* to be treated in a realistic manner. Theoretical analysis of the electronic structure of disordered GaBi$_x$As$_{1-x}$ alloys and QWs using the TB method—full details of which are beyond the scope of this chapter—can be found in Refs. [5,27–29]. Here, we provide a brief overview of this work, describing some general features of the electronic structure, and comparing the results of TB calculations on disordered alloys directly with experimental measurements of the band gap, spin-orbit-splitting energy, and electron effective g factor.

The electronic structure calculations we review in this section were performed on a series of disordered 4096-atom supercells, which are either free-standing (unstrained) or under compressive pseudomorphic strain. Beginning with a GaAs supercell, a disordered GaBi$_x$As$_{1-x}$ supercell was generated by replacing As atoms by Bi atoms at randomly chosen sites on the anion sublattice. In the free-standing supercells, all atomic positions were allowed to relax freely—using a valence force field model based on the Keating potential [71,72]—with no constraints applied to the atoms at the supercell boundaries. For pseudomorphically strained supercells, a macroscopic pseudomorphic strain was imposed by fixing the positions of the atoms on the exterior boundaries and relaxing the positions of the atoms in the interior of the supercell [29]. For the construction of the supercell TB Hamiltonian, the orbital energies at a given atomic site were computed depending on the overall neighbor environment. Local strain effects were taken into account by scaling the interatomic interaction matrix elements based on the relaxed nearest-neighbor bond length d as $\left(\frac{d_0}{d}\right)^{\eta}$, where d_0 is the equilibrium bond length in the equivalent binary compound and η is a dimensionless scaling parameter (the value of which depends on the type and symmetry of the interaction). The bond angle dependence of the interaction matrix elements was represented using the two-center integrals of Slater and Koster [73]. Full details of the TB model are described in Ref. [5].

In order to facilitate the TB analysis of the electronic structure of GaBi$_x$As$_{1-x}$ alloys, the fractional Γ character spectrum $G_\Gamma(E)$ serves as a useful tool to probe the nature and origin of the alloy eigenstates. $G_\Gamma(E)$ is calculated in general by projecting a specific choice of host matrix (in this case, GaAs) Γ-point eigenstates onto the full spectrum of zone-center eigenstates for a given alloy supercell. Specifically, using (0) and (1) to respectively denote the unperturbed host matrix and Bi-containing alloy supercell states, $G_\Gamma(E)$ is computed by projecting the unperturbed zone-center eigenstates $\{|\psi_l^{(0)}\rangle\}$ of a $2M$-atom Ga$_M$As$_M$ supercell onto the eigenstates $\{|\psi_k^{(1)}\rangle\}$ computed at Γ for a given Ga$_M$Bi$_L$As$_{M-L}$ alloy supercell

$$G_\Gamma(E) = \sum_k \sum_{l=1}^{g(E_l)} |\langle\psi_k^{(1)}|\psi_l^{(0)}\rangle|^2 \; T\left(E - E_k\right), \tag{10.1}$$

where $x = \frac{L}{M}$ is the Bi composition, E_k is the energy of the zone-center alloy supercell eigenstate $|\psi_k^{(1)}\rangle$, $g(E_l)$ is the degeneracy of the unperturbed host matrix eigenstate having energy E_l at Γ (so that $g(E_l) = 2, 4$, and 2 for the CB, VB, and SO band edge eigenstates, respectively), and where the “top hat” function $T\left(E - E_k\right)$ is defined so that $G_\Gamma(E_k)$ has a value of unity for a doubly (spin) degenerate host matrix Γ state at energy E_k. As has previously been demonstrated—in Refs. [5], [27] and [29]—computing $G_\Gamma(E)$ for a range of band edge states provides (1) a consistent approach with which to analyze the evolution of the electronic structure in both ordered and disordered alloy supercells, and (2) detailed insight into the nature of the zone-center eigenstates in a Bi-containing alloy, by quantifying their origin in terms of those of the unperturbed host matrix semiconductor.

Figure 10.2 shows the calculated distribution of the Γ character associated with the extended GaAs host matrix heavy-hole (HH), light-hole (LH), and SO band edge states over the full spectrum of zone-center alloy states, for a disordered 4096-atom Ga$_{2048}$Bi$_{82}$As$_{1966}$ ($x = 4\%$) supercell under compressive pseudomorphic strain [29]. The top, middle, and bottom panels of Figure 10.2 show, respectively, the $G_\Gamma^{\text{HH}}(E_{j,1})$, $G_\Gamma^{\text{LH}}(E_{j,1})$ and $G_\Gamma^{\text{SO}}(E_{j,1})$ spectra, calculated using the unperturbed GaAs host matrix HH, LH, and SO states

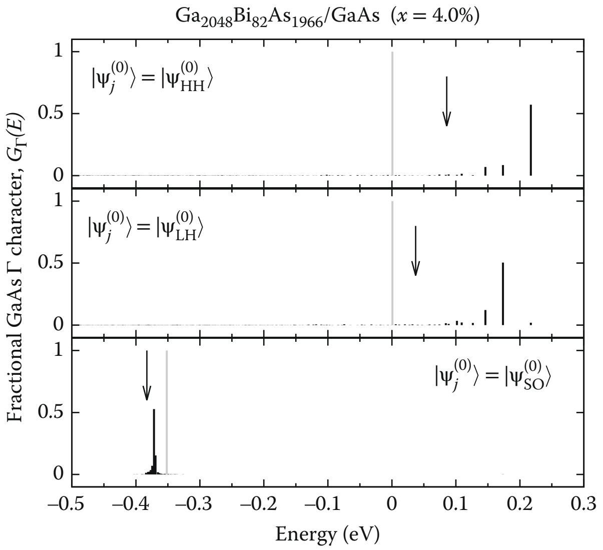


FIGURE 10.2 Black lines: Calculated distribution of the Γ character associated with the GaAs HH (top panel), LH (middle panel), and SO (bottom panel) band edge states over the full spectrum of alloy zone-center valence states in a disordered, 4096-atom $Ga_{2048}Bi_{82}As_{1966}$ ($x = 4\%$) supercell (cf. Equation 10.1). The supercell was placed under compressive pseudomorphic strain corresponding to epitaxial growth on a GaAs substrate, as described in the text. (Adapted from C.A. Broderick et al., *Phys. Rev. B*, 90, 195301, 2014.) Gray lines: Corresponding distribution of the Γ character in a Bi-free $Ga_{2048}As_{2048}$ supercell. The zero of energy is taken at the GaAs VB edge.

$|\psi_{l,0}\rangle = |\psi_{HH,0}\rangle$, $|\psi_{LH,0}\rangle$ and $|\psi_{SO,0}\rangle$ in Equation 10.1. For comparative purposes, the corresponding Γ character of the unperturbed $Ga_{2048}As_{2048}$ host matrix is also shown in Figure 10.2 using gray lines.

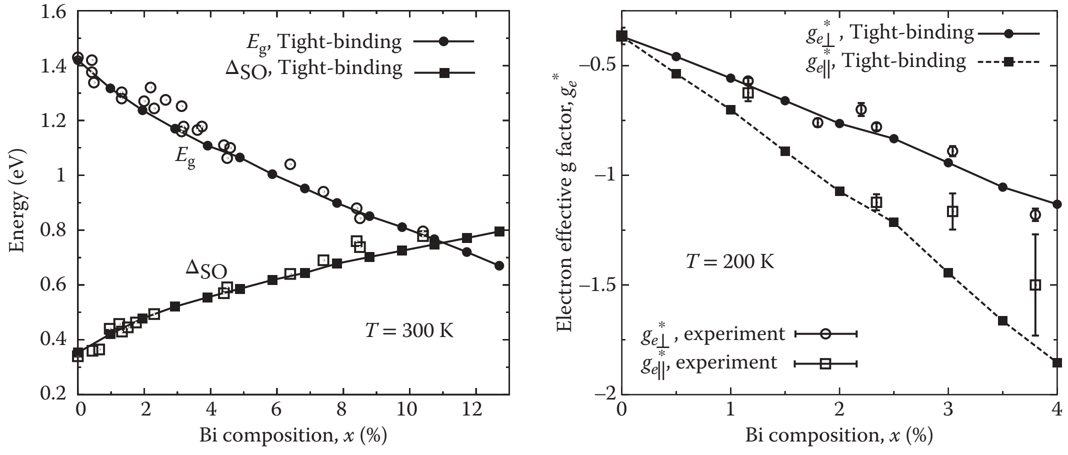
Comparing the computed spectra for the host matrix and Bi-containing supercells reveals several important effects of Bi incorporation on the VB structure. First, incorporation of Bi leads to strong hybridization between the GaAs VB edge states and a series of Bi-related states lying at lower energies, within the VB. Second, this hybridization acts to push to the alloy VB edge states upward in energy with increasing Bi composition. In the ordered supercell case, this behavior can be explained directly in terms of a VBAC interaction (cf. Section 10.3.1), and this general behavior is clearly observed in disordered $GaBi_xAs_{1-x}$ alloys across a broad composition range [5]. Third, we note that the Γ character associated with the GaAs HH and LH states is distributed over a large number of VB states in a disordered $GaBi_xAs_{1-x}$ supercell. There are primarily two reasons for this: (1) the localized impurity states associated with substitutional Bi atoms are resonant with a large density of host matrix VB states, which leads to strong energy broadening of the associated quasi-localized resonances (as described for N impurities in Chapter 9) and is reflected by the large number of alloy VB states having small Γ character [54], and (2) the presence of Bi–Bi pairs (in which a single Ga atom has two Bi nearest neighbors) and larger clusters of Bi atoms (having shared Ga nearest neighbors) leads to the formation of localized impurity states which more strongly bind holes—i.e. which lie close to and/or above the GaAs VB edge in energy [5]. This progression of localized Bi-related cluster states leads, in a disordered alloy, to the distribution of a significant fraction of the GaAs Γ character over a range of VB states lying close in energy to the alloy VB edge. This then quantifies quantifying the strong inhomogeneous broadening of the band edge features in $GaBi_xAs_{1-x}$ alloys observed in spectroscopic measurements in terms of the strong impact of Bi-related disorder on the nature of the $GaBi_xAs_{1-x}$ VB [27]. Fourth, in contrast to the character of the alloy VB edge, we note that the feature associated with the Γ character of the GaAs SO band edge states, visible in the bottom panel of Figure 10.2, remains relatively sharp in $GaBi_xAs_{1-x}$. This suggests that the impact of Bi incorporation and alloy disorder on the SO band is less perturbative than it is on the HH- and LH-like states at the VB edge. However, additional calculations indicate that this distribution of the SO-related Γ character persists as the Bi composition is increased, leading to strong energy broadening and a breakdown of the alloy SO band edge at higher Bi compositions [5]. This suggests a breakdown of a conventional band description of the SO states, which has been confirmed by further analysis [29,61]. Finally, as expected for an epilayer under compressive strain, we see that the highest-energy alloy VB state is predominantly HH-like, having >50% GaAs HH Γ character. We additionally note that both the LH and HH band edges in the alloy are strongly inhomogeneously broadened, which is evidenced by the large number of alloy VB states over which the GaAs VB edge Γ character is distributed [27].

Similar analysis of the conduction states suggests that the impact of Bi incorporation on the CB is generally weak, leading to a downward shift in energy and decrease in the CB-related Γ character that is consistent with alloying effects in conventional semiconductor materials. Overall, these calculated trends confirm that Bi incorporation almost exclusively and strongly perturbs the VB structure, while the CB structure is, by comparison, relatively unaffected [5].

On the basis of these supercell calculations, the variation of the band gap and spin-orbit-splitting energy as a function of Bi composition can be extracted for disordered $\text{GaBi}_x\text{As}_{1-x}$ alloys. These values were calculated directly for each supercell on the basis of the calculated $G_\Gamma(E)$ spectra: the band gap was computed as the difference in energy between the (largely GaAs-like) lowest-energy CB state and the alloy VB state having the largest HH or LH GaAs Γ character; the spin–orbit-splitting energy was then computed as the difference in energy between this (typically highest-energy) alloy VB state and a weighted average energy for the alloy SO band (obtained by averaging over the full set of zone-center alloy energies using the value of the SO Γ character as a weight for each contributing eigenstate) [5]. Given the important role played by alloy disorder in determining the details of the VB structure, for each Bi composition considered this analysis was undertaken for five supercells in which the Bi atoms were randomly distributed at different sites on the anion sublattice. The values of E_g and Δ_{SO} were then obtained for that Bi compostion as the average of those calculated using each of the different supercells considered. The results of these calculations are shown in Figure 10.3a, which compares the calculated variation of E_g and Δ_{SO} as a function of Bi composition x in $\text{GaBi}_x\text{As}_{1-x}$ to a range of experimental measurements. We note that the theoretical calculations are in excellent agreement with the experimental measurements, across the full composition range for which data are available. In particular, the TB model accurately captures the strong, composition-dependent bowing of E_g and Δ_{SO} demonstrating—in line with the discussion above—that the observed strong decrease (increase) of E_g (Δ_{SO}) with increasing x is primarily attributable to the impact of Bi incorporation on the VB structure, and results largely from the upward shift in energy of the alloy VB edge. The TB analysis confirms that this behavior originates due to hybridization of the GaAs VB edge states with a full distribution of Bi-related localized states in a disordered alloy [5,54].

The Landé effective g factor of conduction electrons (g_e^*) is a key parameter in semiconductor materials, describing the response of the electron spins to externally applied magnetic fields, as well as providing useful information regarding the symmetry of the electron and hole states at the Γ-point in the Brillouin zone. Therefore, in addition to providing information pertinent to evaluating the potential of specific materials for spintronic applications, analysis of the electron spin properties and g factor gives insight into the material band structure, providing valuable data for material characterization and modeling. In particular, g_e^* is extremely sensitive to the separation in energy between zone-center crystal eigenstates, as well as hybridization between zone-center eigenstates caused by, e.g. a reduction of symmetry due to strain [75] or quantum confinement [76]. As a result, calculations of g_e^* provide a stringent test of theoretical models of the material band structure [77].

Optical excitation of a semiconductor material with circularly polarized light is known to generate a spin-polarized population of electrons in the CB. These electrons can then relax by recombining with holes in the VB and emitting photons which must be circularly polarized in order to conserve angular momentum, with the nature of the circular polarization determined by the orientation of the electron spin. As such, Larmor precession of electron spins in the presence of an externally applied magnetic field leads to beating of the circular polarization of the photoluminescence (PL) generated by exciting a material sample with



(a) (b)

FIGURE 10.3 (a) Comparison between theory and experiment for the variation of the band gap (E_g) and spin–orbit-splitting energy (Δ_{SO}) as a function of Bi composition x in GaBi$_x$As$_{1-x}$ at 300 K. Experimental measurements of E_g are depicted by open circles. (The experimental data are taken from S. Tixier et al., *Appl. Phys. Lett.*, 82, 2245, 2003; S. Francoeur et al., *Appl. Phys. Lett.*, 82, 3874, 2003; W. Huang et al., *J. Appl. Phys.*, 98, 053505, 2005; K. Alberi et al., *Phys. Rev. B*, 75, 045203, 2007; Z. Batool et al., *J. Appl. Phys.*, 111, 113108, 2012; Yoshida et al., *Jpn. J. Appl. Phys.*, 42, 371, 2003.) Experimental measurements of Δ_{SO} are depicted by open squares. (The experimental data are taken from K. Alberi et al., *Phys. Rev. B*, 75, 045203, 2007; Z. Batool et al., *J. Appl. Phys.*, 111, 113108, 2012; B. Fluegel et al., *Phys. Rev. Lett.*, 97, 067205, 2006.). Closed circles (squares) show the variation of E_g (Δ_{SO}) calculated using the TB model for a series of disordered, 4096-atom supercells. (Adapted from M. Usman et al., *Phys. Rev. B*, 84, 245202, 2011; C.A. Broderick et al., *Semicond. Sci. Technol.*, 28, 125025, 2013.). (b) Comparison between theory and experiment for the variation of the transverse and longitudinal components $g_{\perp,e}^*$ and $g_{\parallel,e}^*$ of the electron effective g factor g_e^* as a function of Bi composition x in GaBi$_x$As$_{1-x}$ at 200 K. Closed circles (squares) show the variation of $g_{\perp,e}^*$ ($g_{\parallel,e}^*$) calculated using the TB model for a series of disordered, 4096-atom supercells placed under pseudomorphic strain as described in the text. Open circles (squares) show the variation of $g_{\perp,e}^*$ ($g_{\parallel,e}^*$) measured for a series of GaBi$_x$As$_{1-x}$/GaAs epitaxial layers using spin quantum beat spectroscopy. (Adapted from Broderick, C.A. et al., *Phys. Rev. B*, 90, 195301, 2014.)

circularly polarized light [78]. Using time-resolved PL spectroscopy, the Larmor frequency $\omega = \frac{g_e^* \mu_B}{\hbar} B$ of these "quantum beats" can be determined, allowing the relevant component of g_e^* to be determined directly for a given relative orientation of the material sample and externally applied magnetic field [76]. As a result of the long coherence times associated with the beating of the PL polarization, this experimental technique, known as spin quantum beat spectroscopy, enables highly accurate determination of the electron Larmor frequency, and hence the effective g factor g_e^* [77,78]. In order to investigate in further detail the impact of Bi incorporation on the electronic structure, as well as to provide insight into the impact of Bi incorporation on the spin properties of electrons, Broderick et al. [29] developed a TB-based model for the calculation of g_e^* in disordered GaBi$_x$As$_{1-x}$ alloys. Here, we provide a brief summary of the analysis of g_e^* in GaBi$_x$As$_{1-x}$, including comparison to experimental data, full details of which can be found in Ref. [29].

The hybridization between zone-center LH and SO states brought about by the tetragonal deformation and associated reduction of symmetry in pseudomorphically strained layers is known to play a key role in determining the anisotropy of g_e^* in conventional semiconductor alloys, leading to distinct transverse and longitudinal components $g_{\perp,e}^*$ and $g_{\parallel,e}^*$ of the g_e^* tensor [75]. Returning our attention to Figure 10.2, we begin by noting that the calculated distribution of GaAs SO character for this typical pseudomorphically strained GaBi$_x$As$_{1-x}$ supercell highlights that the strong inhomogeneous broadening of the LH-like alloy VB states brought about by alloy disorder minimizes this expected strain-induced LH–SO hybridization—i.e. there is no notable projection of the GaAs SO band edge character onto LH-like states close to the alloy

VB edge—providing the first hint that the unusual VB structure of Bi-containing alloys is likely to have a strong impact on g_e^*.

In the absence of strain, g_e^* is expected to be isotropic, with $g_{\perp,e}^* = g_{\parallel,e}^*$ [75]. Comparative calculations we have undertaken on free-standing (unstrained) supercells demonstrate strain-induced mixing between the LH and SO states in ordered supercells, enabling us to conclude that the strong inhibition of such mixing observed in Figure 10.2 arises due to the disorder-induced energy broadening of the GaAs LH character over a significant number of alloy VB states. In order to investigate the impact of Bi incorporation, pseudomorphic strain, and Bi-related alloy disorder on the isotropy of g_e^*, we have therefore first calculated $g_{\perp,e}^*$ and $g_{\parallel,e}^*$ as a function of Bi composition x in a series of free-standing supercells. These calculations show that (1) Bi incorporation leads to a rapid, monotonic increase in the magnitude of g_e^* with increasing x, and (2) the reduced local symmetry present in a disordered $GaBi_xAs_{1-x}$ alloy supercell has a negligible effect on the isotropy of g_e^* [29]. The latter result suggests that the anisotropy of g_e^* observed in experimental measurements performed on $GaBi_xAs_{1-x}$ epitaxial layers arises primarily due to the reduction in symmetry brought about by the impact of biaxial stress on the band structure, much like in conventional semiconductor materials, but that the LH–SO mixing to which this behavior is attributed in conventional semiconductor alloys [75] is not a significant factor determining the anisotropy of g_e^* in $GaBi_xAs_{1-x}$ alloys.

In order to determine the anisotropy of g_e^* in real, pseudomorphically strained epitaxial layers, as well as to compare the results of the theoretical calculations to experiment, Broderick et al. repeated this analysis for disordered $GaBi_xAs_{1-x}$ supercells under the application of a macroscopic strain corresponding to pseudomorphic growth on a GaAs substrate [29] (applied as described above). The results of this analysis are summarized in Figure 10.3b, which shows the calculated variation of $g_{\perp,e}^*$ (closed circles) and $g_{\parallel,e}^*$ (closed squares) with x, as well as experimental measurements of $g_{\perp,e}^*$ (open circles) and $g_{\parallel,e}^*$ (open squares) performed on a series of strained $GaBi_xAs_{1-x}$ epitaxial layers at a temperature of 200 K [29].

In the presence of pseudomorphic strain, we see that g_e^* is strongly anisotropic. As in the unstrained case, we note that the magnitudes of both $g_{\perp,e}^*$ and $g_{\parallel,e}^*$ increase strongly and monotonically with increasing x. We also calculate that (1) the magnitude of $g_{\parallel,e}^*$ exceeds that of $g_{\perp,e}^*$ for non-zero x, in agreement with experiment, and (2) the magnitude of $g_{\parallel,e}^*$ is close to the isotropic value of g_e^* calculated in the absence of strain [29], both of which are expected for compressively strained epitaxial layers [75]. Based on these comparative calculations, we conclude that while the overall symmetry of g_e^* is largely independent of the effects of alloy disorder, the impact of pseudomorphic strain is crucial for determining the experimentally observed anisotropy of g_e^* in $GaBi_xAs_{1-x}$ epitaxial layers. We note that the calculated enhancement of $g_{\perp,e}^*$ with increasing x is in excellent agreement with experiment, confirming not only the accuracy of the detailed picture of the $GaBi_xAs_{1-x}$ VB structure provided by the TB model, but also the strong impact of Bi incorporation on the spin properties of electrons in $GaBi_xAs_{1-x}$ alloys.

The effects of alloy disorder on the calculated variation of $g_{\perp,e}^*$ and $g_{\parallel,e}^*$ with x can be understood in more detail by considering the distribution of host matrix Γ character shown in Figure 10.2. In a disordered $GaBi_xAs_{1-x}$ alloy, the VB edge GaAs Γ character is distributed over a large number of alloy VB states, with the details of the distribution determined by the precise short-range disorder present in the alloy. The details of this distribution of the GaAs Γ character onto VB states lying below the alloy VB edge in energy plays a crucial role in determining the calculated variation of $g_{\perp,e}^*$ and $g_{\parallel,e}^*$ with x, since the contribution from a given alloy VB state to a specific component of g_e^* is weighted by its associated host matrix Γ character [29]. Specifically, mixing of GaAs VB edge Γ character into alloy VB states with greater separation in energy from the alloy CB edge results in contributions to g_e^* which act against the much stronger increase in the magnitudes of $g_{\perp,e}^*$ and $g_{\parallel,e}^*$ than would be expected based solely upon the measured variation of E_g and Δ_{SO} with x [24,29]. We recall from Section 10.2.1 that this breakdown of the simple VBAC picture of the VB structure in the vicinity of the SO band [29] suggests that detailed approaches are typically required to understand material properties involving eigenstates that are energetically remote from the CB and VB edges.

Overall, we conclude that the TB model developed in Ref. [5] provides a detailed and quantitative understanding of the electronic structure of $GaBi_xAs_{1-x}$ and related alloys. The variation of the band gap, spin–orbit-splitting energy, and electron effective g factor as a function of Bi composition calculated using the TB method are in good, quantitative agreement with experiment, and this theoretical approach has delivered significant insight into the impact of Bi incorporation on the $GaBi_xAs_{1-x}$ band structure.

10.2.3 Outlook for Atomistic Theory

Despite the detailed insight atomistic calculations have provided into the impact of Bi incorporation on the electronic and optical properties of $GaBi_xAs_{1-x}$ and related alloys, there is still significant potential to develop and apply atomistic methods and models to further analyze a range of fundamental material properties in this emerging class of III-V semiconductors.

For example, based on the observation that the VBAC picture of the band structure—which very accurately describes the evolution of E_g and Δ_{SO} with x (cf. Section 10.3), as well as the optical properties of dilute bismide semiconductor lasers (cf. Section 10.4)—breaks down for states energetically remote from the VB edge, it is possible to conclude that simple, phenomenological approaches are ill equipped to account for the impact of Bi incorporation and alloy disorder on a number of technologically relevant material properties. This is perhaps unsurprising, since the VBAC model treats only the main features of the band structure and does so in a phenomenological manner. However, this conclusion has significant implications for further theoretical analysis of dilute bismide alloys, implying, e.g. that simple models of the band structure are ill equipped to quantitatively describe key processes such as Auger recombination and IVBA, which involve a broad energy range of alloy VB states including, in particular, states lying close to the SO band edge in energy. As was described for dilute nitride alloys in Chapter 9, detailed and specialized theoretical approaches must therefore be developed in order to obtain a quantitative understanding of complicated processes and properties of this nature.

There have to date been no *direct* theoretical investigations on a fundamental level of the impact of Bi incorporation on a number of important material properties and processes including, e.g. (1) carrier lifetimes, (2) radiative and non-radiative recombination rates, (3) carrier transport, and (4) impact ionization. Detailed understandings of these and related properties must be developed in order to inform the development of Bi-containing alloys for applications in a range of novel semiconductor devices.

Previous theoretical analysis has highlighted the strong influence of Bi-related alloy disorder in determining the details of the electronic structure, meaning that further quantitative theoretical analysis of the properties of Bi-containing alloys is likely to require detailed atomistic treatments. This requirement presents the theoretical community with a series of interesting challenges, centered on developing the capability to describe—on an atomic scale—the impact of short-range disorder on a range of fundamental properties and processes in highly mismatched alloys, where dilute concentrations of impurities strongly perturb the electronic structure in a manner that is not compatible with many commonly employed approximations due to the prominence of alloy disorder effects in determining the material properties.

On this basis, further theoretical research on dilute bismides—perhaps informed by related research on dilute nitride alloys—is likely to deliver not only an improved understanding of this interesting class of semiconductors, but also insights and methods that can be applied to a broad range of emerging highly mismatched materials.

10.3 Modeling Dilute Bismide Band Structure: The k•p Method

Having presented an overview of atomistic electronic structure calculations for $GaBi_xAs_{1-x}$ alloys, in this section, we present a (continuum) **k·p** model of the $(In)GaBi_xAs_{1-x}$ band structure which is well suited to the simulation of the electronic and optical properties of dilute bismide materials and devices.

In Section 10.3.1, we outline how a VBAC approach can be used to provide a simple yet effective interpretation of the band structure of dilute bismide alloys and describe how detailed atomistic calculations have provided a means to deliver quantitative predictive power to this phenomenological picture of the electronic properties.

In Section 10.3.2 we review the use of TB supercell calculations to derive an extended basis set 12-band **k·p** Hamiltonian for (In)GaBi$_x$As$_{1-x}$ alloys. This rigorous analysis verifies that localized impurity states associated with isolated substitutional Bi atoms play the dominant role in determining the main features of the band structure close in energy to the VB edge. The result of this is—despite the fact that more sophisticated atomistic approaches are typically required to understand the detailed properties of these materials—a VBAC-based approach is sufficient to provide a understanding of the electronic and optical properties of dilute bismide devices such as QW lasers in which only quantitative optically active states lying close in energy to the CB and VB edges are of significant importance.

In Section 10.3.3 we present a general method—based on the 12-band **k·p** Hamiltonian—for calculating the band offsets in pseudomorphically strained (In)GaBi$_x$As$_{1-x}$ epitaxial layers and QWs, such as those which are attracting significant attention for the development of highly efficient GaAs- and InP-based semiconductor lasers. The theory outlined in this section lays the foundation for our analysis of dilute bismide QW lasers in Section 10.4.

10.3.1 Valence Band-Anticrossing

As described in Chapter 9, the strong, composition-dependent band gap bowing observed in the dilute nitride alloy (In)GaN$_x$As$_{1-x}$ is well explained in terms of a band-anticrossing interaction between two levels: one at energy E_{CB} associated with the extended CB edge states of the (In)GaAs host matrix, and the second at energy E_{N} associated with the highly localized N-related impurity states. In this simple model, the CB edge energy of the N-containing alloy is given by the lower eigenvalue of the two-band Hamiltonian [8]:

$$H = \begin{pmatrix} E_{\mathrm{N}} & V_{\mathrm{Nc}} \\ V_{\mathrm{Nc}} & E_{\mathrm{CB}} \end{pmatrix}, \tag{10.2}$$

where V_{Nc} is the N composition dependent matrix element describing the interaction between E_{CB} and E_{N}, which varies with N composition x as $V_{\mathrm{Nc}} = \beta\sqrt{x}$.

Bismuth, being the heaviest stable group-V element, is significantly larger and more electropositive than As. It is therefore expected that any Bi-related impurity levels in (In)GaAs should lie close to or below the VB edge in energy and, if present, a VBAC interaction will occur between the Bi-related impurity levels and the VB edge states of the host matrix (cf. Section 10.2.1). The presence of such an interaction has been proposed on the basis of experimental measurements [4], in order to account for the observed composition-dependent bowing of the (In)GaBi$_x$As$_{1-x}$ band gap. While detailed analysis of atomistic supercell calculations [5,54,60] directly confirms the validity of this approach to describing the band structure in the case of ordered alloys, we note that the VBAC model introduced in Ref. [4] is in practice strongly limited by parametric ambiguity which removes its predictive capability. For example, one typically has two parameters (the energy of the Bi-related states E_{Bi} and the VBAC coupling strength β) which must be fitted to a single piece of experimental data (the measured alloy band gap at a given Bi composition), meaning that there is insufficient information available with which to provide an unambiguous fit without resorting to limiting assumptions. For example, initial estimation of the energy of the localized states associated with an isolated substitutional Bi impurity in GaAs produced a value $E_{\mathrm{Bi}} = -0.4$ eV, which lies approximately 60 meV below the SO band edge [4]. However, first principles pseudopotential and large-supercell TB calculations confirm that these states lie above the SO band edge in energy (cf. Section 10.2.1), suggesting that simple estimates based on experimental data produce a qualitatively different description of the VB structure. Usman et al. overcame this limitation by using atomistic TB calculations to directly evaluate the

VBAC parameters, thereby producing a quantitative VBAC approach that provides a predictive capability which has been verified by comparison to a range of experimental data [15,39,54,79]. Here, we outline the TB analysis of the VBAC interaction, as a prelude to the derivation of a suitable **k·p** Hamiltonian with which to model the band structure of (In)GaBi$_x$As$_{1-x}$ alloys [54] and heterostructures [15,39,79].

To directly analyze the VBAC interaction, we consider the electronic structure of ordered GaBi$_x$As$_{1-x}$ and InBi$_x$As$_{1-x}$ alloys by inserting a single substitutional Bi atom in a series of cubic X$_M$Bi$_1$As$_{M-1}$ (X = Ga, In) supercells containing a total of $2M = 8N^3$ atoms, for $2 \leq N \leq 8$. In the VBAC model the Bi-hybridized (In)GaBi$_x$As$_{1-x}$ VB edge states are an admixture of the unperturbed GaAs VB edge states and the T_2 symmetric localized states associated with isolated Bi impurities. Writing the alloy VB edge states as a linear combination in this manner, we obtain an expression for the Bi-related (doublet) states that interact with the (In)GaAs VB edge [5,42,54]

$$|\psi_{\text{Bi},i}\rangle = \frac{|\psi_i^{(1)}\rangle - \sum_n |\psi_n^{(0)}\rangle\langle\psi_n^{(0)}|\psi_i^{(1)}\rangle}{\sqrt{1 - \sum_n |\langle\psi_n^{(0)}|\psi_i^{(1)}\rangle|^2}}, \tag{10.3}$$

where the indices n and i run respectively over the four-fold degenerate states of the (In)GaAs and (In)GaBi$_x$As$_{1-x}$ VB edge, respectively. The superscripts (0) and (1) refer respectively to the unperturbed host matrix and alloy VB edge eigenstates.

From Equation 10.3 we have calculated the energy of the Bi-related states, $E_{\text{Bi},i} = \langle\psi_{\text{Bi},i}|\hat{H}(x)|\psi_{\text{Bi},i}\rangle$, as well as the coupling strength $V_{\text{Bi},i}$ of the VBAC interaction between $|\psi_{\text{Bi},i}\rangle$ and the host matrix VB edge states, $V_{\text{Bi},i} = \langle\psi_{\text{Bi},i}|\hat{H}(x)|\psi_{\text{VB},i}^{(0)}\rangle$, where $\hat{H}(x)$ is the full TB Hamiltonian for the Bi-containing supercell and $|\psi_{\text{VB},i}^{(0)}\rangle$ is the unperturbed host matrix VB edge state with which $|\psi_{\text{Bi},i}\rangle$ interacts [5]. Doing this shows directly that (1) a substitutional Bi atom in Ga$_M$Bi$_1$As$_{M-1}$ or in In$_M$Bi$_1$As$_{M-1}$ forms a set of four-fold degenerate impurity states (at energy $E_{\text{Bi}} = E_{\text{Bi},i}$), which are highly localized about the Bi site, and (2) the strength of the VBAC interaction between the states $|\psi_{\text{Bi},i}\rangle$ and those of the unperturbed host matrix VB edge varies with Bi composition x as $V_{\text{Bi}} = \beta\sqrt{x}$.

Figure 10.4 shows the calculated values of the difference in energy between the Bi-related states and the unperturbed host matrix VB edge, $\Delta E_{\text{Bi}} = E_{\text{Bi}} - E_{\text{VB}}^{(0)}$, as well as the VBAC coupling parameter β, as a function of supercell size (alloy composition) for a series of ordered X$_M$Bi$_1$As$_{M-1}$ (X = Ga, In) supercells. We calculate a strong variation of ΔE_{Bi} and β with Bi composition at larger x, with the trends stabilizing as x approaches the impurity limit in the larger (>2000 atom) supercells [5]. In line with the first-principles pseudopotential calculations of Zhang et al. [49], we find that this behavior is attributable to the folding of host matrix VB states back to the Γ point as the supercell size is increased, which makes more zone-center supercell states having large **k** components available to construct $|\psi_{\text{Bi},i}\rangle$ in the larger supercells [54]. Therefore, we conclude that one should refer to large supercell calculations when examining the behavior of a substitutional Bi impurity in (In)GaAs alloys. We also note that $\Delta E_{\text{Bi}} < 0$ in both GaBi$_x$As$_{1-x}$ and InBi$_x$As$_{1-x}$, indicating that a substitutional Bi atom forms *resonant* states in (In)GaAs which lie below the unperturbed host matrix VB edge in energy. This is distinct from the case of a substitutional Bi impurity in GaP, where Bi forms *bound* states that lie energetically with the host matrix band gap [5,55].

10.3.2 12-Band k·p Hamiltonian for Dilute Bismide Alloys

In order to derive an appropriate **k·p** Hamiltonian for (In)GaBi$_x$As$_{1-x}$, we began with an eight-band model for the (In)GaAs host matrix [80], which employs a basis of spin-degenerate zone-center Bloch states for the lowest-energy CB, as well as the HH, LH, and SO VBs. By diagonalizing the (In)GaAs TB Hamiltonian at Γ, we obtain these basis states $|u_1\rangle, \ldots, |u_8\rangle$ directly. Then, we use this set of supercell basis states to construct an eight-band parameter set for (In)GaAs directly from the full TB calculations. This gives a

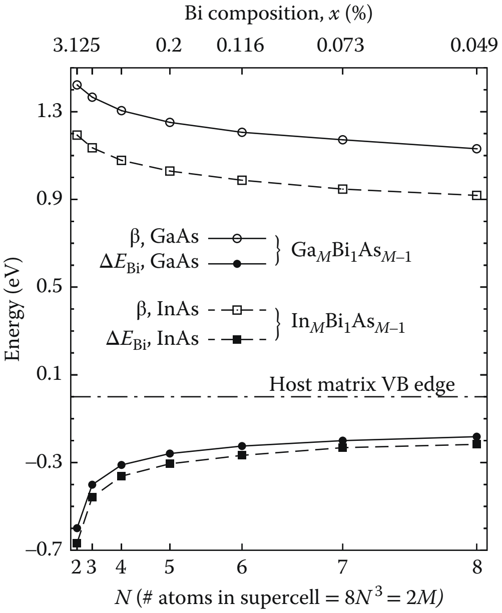


FIGURE 10.4 Calculated difference in energy between the Bi-related localized impurity states and host matrix VB edge, ΔE_{Bi} (closed symbols) and VBAC coupling parameter, β (open symbols) for a substitutional Bi impurity in a series of ordered, cubic $\mathrm{Ga}_M\mathrm{Bi}_1\mathrm{As}_{M-1}$ (circles) and $\mathrm{In}_M\mathrm{Bi}_1\mathrm{As}_{M-1}$ (squares) supercells. (Adapted from C.A. Broderick et al., *Semicond. Sci. Technol.*, 28, 125025, 2013; C.A. Broderick et al., *Phys. Stat. Sol. B*, 250, 773, 2013.) Each supercell contains a total of $2M = 8N^3$ atoms (where N is the number of cubic 8-atom unit cells along each of the x-, y- and z-directions), corresponding to Bi compositions ranging from $x = 0.049\%$ (4096 atoms) to $x = 3.125\%$ (64 atoms). The zero of energy is taken in each case at the host matrix VB edge and is denoted by a dash-dotted line at 0 eV.

parameterization of the eight-band Hamiltonian that reproduces the full TB band structure of an $\mathrm{X}_M\mathrm{As}_M$ (X = Ga, In) supercell with a high degree of accuracy in the vicinity of the zone center [42,54].

Having established this description of the host matrix band structure, we extend the eight-band basis set that describes the host matrix band structure by directly including the four Bi-related localized impurity (doublet) states described by Equation 10.3. The manner in which these Bi-related basis states couple to the host matrix VB edge states is then straightforward to include in the modified **k·p** Hamiltonian: the matrix elements between the localized Bi-related states and the extended VB edge states of the host matrix are precisely those evaluated in Section 10.3.1 as $V_{\mathrm{Bi}} = \beta\sqrt{x}$. In this manner, we arrive at a 12-band, extended basis set **k·p** Hamiltonian for (In)GaBi$_x$As$_{1-x}$ alloys in which the impact of Bi on the band structure is described primarily via the VBAC interaction. In addition to the VBAC-induced shift in the VB edge energies, there are also conventional (virtual crystal) contributions to each of the band edge energies which depend linearly upon the Bi composition x. We have again used the TB model to directly quantify these effects, by evaluating the matrix elements $\langle u_b|\widehat{H}(x)|u_b\rangle$, where $|u_b\rangle$ are the aforementioned basis states for the host matrix. Following this approach, we find firstly that the Bi-related localized states do not coupled directly to the extended CB edge states of the host matrix. Next, we calculate that the virtual crystal contributions to the variation of the (In)GaAs band edge energies with Bi composition x are given by: $E_{\mathrm{CB}}(x) = E_{\mathrm{CB}}^{(0)} - \alpha x$, $E_{\mathrm{HH}}(x) = E_{\mathrm{LH}}(x) = E_{\mathrm{VB}}^{(0)} + \kappa x$, and $E_{\mathrm{SO}}(x) = E_{\mathrm{SO}}^{(0)} - \gamma x$, for the CB, VB, and SO band edges, respectively (where the superscript (0) denotes the band edge energy of the unperturbed host matrix).

We note that the incorporation of Bi then introduces five band structure parameters in addition to those that describe the (In)GaAs host matrix, all of which are required to provide a reliable description of the $GaBi_xAs_{1-x}$ band structure using a VBAC-based approach. (As we see in Section 10.3.3, this number increases to seven in the presence of pseudomorphic strain.) These parameters are (1–3) the virtual crystal contributions to the energies of the CB (α), VB (κ), and SO (γ) band edge energies, (4) the energy of the Bi-related localized states relative to the (In)GaAs VB edge (ΔE_{Bi}), and (5) the VBAC coupling strength (β). This number of free parameters is too large to provide an unambiguous fit to the available experimental band structure data, thereby emphasizing the importance of atomistic calculations in informing not only the understanding of the fundamental properties of Bi-containing alloys, but also the development and parametrization of sufficiently detailed phenomenological models to enable the investigation of heterostructures and devices. The 12-band **k·p** Hamiltonian derived by Broderick et al. using this approach is presented in full in Ref. [54]; the Bi-related band structure parameters derived for $GaBi_xAs_{1-x}$ and $InBi_xAs_{1-x}$ alloys on the basis of the TB analysis are listed in Table 10.1. We note that while this 12-band model contains fewer bands than the 14-band model originally introduced by Alberi et al. in Ref. [4], it ultimately provides a more accurate description of the band structure since, in addition to VBAC interactions, the contributions of strain and virtual crystal effects on the composition dependence of the band edge energies have been explicitly taken into account. Additionally, the 12-band model is fully parametrized on the basis of atomistic supercell calculations, without making recourse to fits to experimental data, thereby providing genuine predictive capability.

Figure 10.5a compares the results of calculations of the $Ga_{32}Bi_1As_{31}$ ($x = 3.125\%$) band structure in the vicinity of the zone center using the 12-band **k·p** (solid lines) and full TB (closed circles) Hamiltonians. Note that since the TB calculation is carried out on a 64-atom supercell, the appropriate values of ΔE_{Bi} and β for the supercell must be used in the **k·p** calculation in order to directly compare the band structures calculated using the two methods (cf. Figure 10.4). We observe excellent agreement between the **k·p** and TB band structures. It can be seen that (1) the VBAC interaction has pushed the alloy VB edge upward in energy, and (2) the CB and SO band edge energies in the alloy are well described by the Bi-induced virtual crystal shifts described above. The VBAC interaction, which describes the strong composition-dependent band gap bowing, accounts for the majority of the observed band gap reduction with increasing x. We also note that the presence of the VBAC interaction is verified by the presence of TB bands that correspond to the lower energy Bi-related ($E_{-}^{LH/HH}$) bands in the **k·p** calculation. These calculations therefore confirm that the 12-band Hamiltonian provides an accurate description the unusual band structure of $(In)GaBi_xAs_{1-x}$ alloys [42,54].

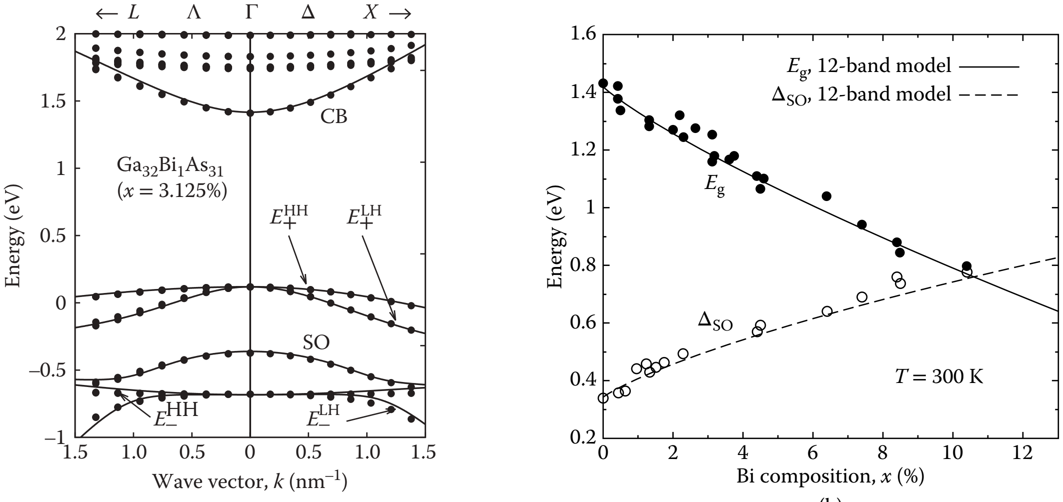
The main features of the band structure calculated using the 12-band Hamiltonian are in excellent agreement with those obtained from TB calculations performed on realistic disordered alloys across a wide range of Bi compositions, despite having been derived and parameterized for ordered alloys [54]. Figure 10.5b shows the variation of E_g (solid black line) and Δ_{SO} (dashed black line) as a function of Bi composition x at 300 K in $GaBi_xAs_{1-x}$, calculated using the 12-band **k·p** model. Comparing the results of these calculations to a range of experimental data from the literature, we see that the 12-band model produces a highly accurate description of the main features of the $GaBi_xAs_{1-x}$ band structure, and does so across a broad composition range. Overall, this analysis suggests that a suitably constructed and parametrized

TABLE 10.1 Bi-related Band Structure Parameters for the 12-band **k·p** Hamiltonian, Computed Directly Using TB Supercell Calculations for $GaBi_xAs_{1-x}$ and $InBi_xAs_{1-x}$ Alloys as Described in the Text

Alloy	ΔE_{Bi} (eV)	β (eV)	α (eV)	κ (eV)	γ (eV)
$GaBi_xAs_{1-x}$	−0.183	1.13	2.82	1.01	0.55
$InBi_xAs_{1-x}$	−0.217	0.92	2.60	1.03	0.02

Source: C.A. Broderick et al., *Semicond. Sci. Technol.*, 28, 125025, 2013.

The parameters ΔE_{Bi} and β specify the nature of the VBAC interaction, while α, κ, and γ, respectively, describe the Bi-induced virtual crystal (conventional alloy) shifts to the CB, VB, and SO band edge energies.



(a) (b)

FIGURE 10.5 (a) Calculated band structure of an ordered, cubic $Ga_{32}Bi_1As_{31}$ ($x = 3.125\%$, 64-atom) supercell, along the Λ- and Δ-directions, close to the center of the Brillouin zone, using the TB (closed circles) and 12-band **k·p** (solid lines) Hamiltonians. (Adapted from Broderick, C.A. et al., *Semicond. Sci. Technol.*, 28, 125025, 2013; C.A. Broderick et al., *Phys. Stat. Sol. B*, 250, 773, 2013.) The zero of energy is taken at the host matrix (GaAs) VB edge. (b) Comparison between theory and experiment for the variation of the band gap (E_g) and spin-orbit-splitting energy (Δ_{SO}) as a function of Bi composition x in $GaBi_xAs_{1-x}$ at 300 K. Experimental measurements of E_g are depicted by closed circles. (The experimental data are taken from S. Tixier et al., *Appl. Phys. Lett.*, 82, 2245, 2003; S. Francoeur et al., *Appl. Phys. Lett.*, 82, 3874, 2003; W. Huang et al., *J. Appl. Phys.*, 98, 053505, 2005; K. Alberi et al., *Phys. Rev. B*, 75, 045203, 2007; Z. Batool et al., *J. Appl. Phys.*, 111, 113108, 2012; J. Yoshida et al., *Jpn. J. Appl. Phys.*, 42, 371, 2003.) Experimental measurements of Δ_{SO} are depicted by open circles. (The experimental data are taken from K. Alberi et al., *Phys. Rev. B*, 75, 045203, 2007; Z. Batool et al., *J. Appl. Phys.*, 111, 113108, 2012; B. Fluegel et al., *Phys. Rev. Lett.*, 97, 067205, 2006.) Solid (dashed) lines show the variation of E_g (Δ_{SO}) calculated using the 12-band **k·p** model. (Adapted from C.A. Broderick et al., *Semicond. Sci. Technol.*, 28, 125025, 2013; C.A. Broderick et al., *Phys. Stat. Sol. B*, 250, 773, 2013; C.A. Broderick et al., *IEEE J. Sel. Top. Quantum Electron.*, 21, 1503313, 2015.)

VBAC model can be used to reliably describe the evolution of the main features of the band structure of (In)$GaBi_xAs_{1-x}$ alloys across a wide range of Bi compositions.

10.3.3 Calculation of Band Offsets in Bi-Containing Quantum Wells

Having considered the theoretical description of the bulk band structure of dilute bismide alloys, we turn our attention now to the calculation of the band offsets in pseudomorphically strained $GaBi_xAs_{1-x}$/(Al)GaAs QWs, as a prelude to our discussion of the electronic and optical properties of GaAs-based dilute bismide QW lasers in Section 10.4. Here, we outline the calculation of the band offsets at arbitrary Bi composition x using the 12-band **k·p** model described in Section 10.3.2. Full details of the analysis upon which this analysis of the QW band offsets is based can be found in Ref. [79].

In order to calculate the strained $GaBi_xAs_{1-x}$ band edge energies we consider the 12-band Hamiltonian including the Bir-Pikus matrix elements describing the impact of pseudomorphic strain on the band edge energies at Γ [81]. The steps in the calculation of the band offsets for a $GaBi_xAs_{1-x}$/GaAs QW are illustrated schematically in Figure 10.6. While we focus here on GaAs-based heterostructures, we note that this method for calculating the QW band offsets is generally applicable to alloys whose band structure can be described using the 12-band model, and so can readily be applied to Bi-containing alloys grown

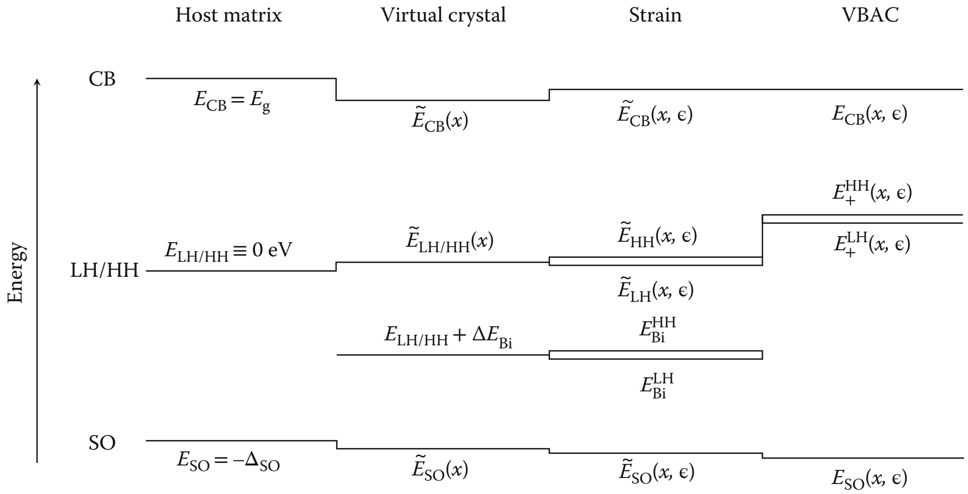

Host matrix
Virtual crystal
Strain
VBAC
CB
LH/HH
SO
Energy
$E_{\text{CB}} = E_{\text{g}}$
$\tilde{E}_{\text{CB}}(x)$
$\tilde{E}_{\text{CB}}(x, \epsilon)$
$E_{\text{CB}}(x, \epsilon)$
$E_{+}^{\text{HH}}(x, \epsilon)$
$E_{\text{LH/HH}} \equiv 0$ eV
$\tilde{E}_{\text{LH/HH}}(x)$
$\tilde{E}_{\text{HH}}(x, \epsilon)$
$E_{+}^{\text{LH}}(x, \epsilon)$
$\tilde{E}_{\text{LH}}(x, \epsilon)$
$E_{\text{LH/HH}} + \Delta E_{\text{Bi}}$
$E_{\text{Bi}}^{\text{HH}}$
$E_{\text{Bi}}^{\text{LH}}$
$E_{\text{SO}} = -\Delta_{\text{SO}}$
$\tilde{E}_{\text{SO}}(x)$
$\tilde{E}_{\text{SO}}(x, \epsilon)$
$E_{\text{SO}}(x, \epsilon)$

FIGURE 10.6 Schematic illustration of the calculation of the bulk GaBi$_x$As$_{1-x}$/GaAs CB, LH, HH, and SO band offsets, using the 12-band **k·p** model of Section 10.3.2 and including the effects of pseudomorphic strain on the band edge energies. (Adapted from Broderick, C.A. et al., *Semicond. Sci. Technol.*, 30, 094009, 2015; G.M.T. Chai et al., *Semicond. Sci. Technol.*, 30, 094015, 2015.) The zero of energy is taken at the VB edge of the unstrained GaAs host matrix. The full details of each step in the calculation are outlined in the text.

using alternative barriers and/or substrates. For barrier materials other than GaAs, the calculation follows straightforwardly from that presented here: one needs simply to calculate the (In)GaBi$_x$As$_{1-x}$ band edge energies as outlined below, with the barrier band edge energies then calculated with respect to the (In)GaAs host matrix in the usual way (using, e.g., the model solid theory [82]) in order to determine the band offsets.

The first step in the calculation is shown on the left side of Figure 10.6, in the portion labeled "Host matrix." Here, we begin with a GaAs host matrix (barrier material) and choose the zero of energy to lie at the unstrained VB edge. Second, in the section labeled "Virtual crystal," we include the conventional alloy contributions to the Bi-induced band edge energy shifts for a given Bi composition x as $\tilde{E}_{\text{CB}}(x) = E_{\text{CB}} - \alpha x \equiv E_{\text{g}} - \alpha x$, $\tilde{E}_{\text{LH/HH}}(x) = E_{\text{LH/HH}} + \kappa x \equiv \kappa x$, and $\tilde{E}_{\text{SO}}(x) = E_{\text{SO}} - \gamma x \equiv -\Delta_{\text{SO}} - \gamma x$, where, as described above, incorporation of Bi has introduced a set of degenerate resonant impurity levels at energy ΔE_{Bi} relative to the host matrix VB edge.

Third, in the section labeled "Strain," we calculate the strain-induced shifts to the band edge energies, thereby taking account of the fact that an (In)GaBi$_x$As$_{1-x}$ epilayer grown on a GaAs substrate will be under compressive pseudomorphic strain. Following the conventions of Krijn [83], we calculate these strain-induced shifts to the CB, HH, LH, and SO band edge energies as $\tilde{E}_{\text{CB}}(x, \epsilon) = \tilde{E}_{\text{CB}}(x) + \delta E_{\text{CB}}^{\text{hy}}$, $\tilde{E}_{\text{LH}}(x, \epsilon) = \tilde{E}_{\text{LH}}(x) + \delta E_{\text{VB}}^{\text{hy}} + \delta E_{\text{VB}}^{\text{ax}}$, $\tilde{E}_{\text{HH}}(x, \epsilon) = \tilde{E}_{\text{HH}}(x) + \delta E_{\text{VB}}^{\text{hy}} - \delta E_{\text{VB}}^{\text{ax}}$, and $\tilde{E}_{\text{SO}}(x, \epsilon) = \tilde{E}_{\text{SO}}(x) + \delta E_{\text{VB}}^{\text{hy}}$, where the energy shifts associated with the hydrostatic and axial components of the strain are $\delta E_{\text{CB}}^{\text{hy}} = a_c \left(\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz} \right)$, $\delta E_{\text{VB}}^{\text{hy}} = a_v \left(\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz} \right)$, and $\delta E_{\text{VB}}^{\text{ax}} = -\frac{b}{2} \left(\epsilon_{xx} + \epsilon_{yy} - 2\epsilon_{zz} \right)$ [81]. The non-zero components of the strain tensor given by $\epsilon_{xx} = \epsilon_{yy} = \frac{a_0 - a(x)}{a(x)}$ and $\epsilon_{zz} = -\frac{2C_{12}(x)}{C_{11}(x)}$ [83], where a_0 is the substrate lattice constant, and $a(x)$, $C_{11}(x)$, and $C_{12}(x)$ are the (In)GaBi$_x$As$_{1-x}$ lattice and elastic constants.

In order to obtain $a(x)$, $C_{11}(x)$, and $C_{12}(x)$ for (In)GaBi$_x$As$_{1-x}$ we interpolate linearly between those of the end-point binary compounds: we use the (Al)GaAs lattice and elastic constants recommended by Vurgaftman et al. [84], and for (In)GaBi we use those calculated *ab initio* by Ferhat and Zaoui [65]. Due to a lack of information in the literature regarding the band edge deformation potentials of (In)GaBi, we take a_c, a_v, and b for (In)GaBi$_x$As$_{1-x}$ to be equal to those of the (In)GaAs host matrix, which is expected to be a good approximation at the dilute Bi compositions under consideration.

Due to the p-like symmetry of the Bi-related impurity states, there is a lifting of the degeneracy of the LH- and HH-like impurity states in the presence of pseudomorphic strain, resulting from the tetragonal distortion of the crystal lattice. Based on the results of TB calculations on large, ordered supercells, it is found that the energies of these states, $E_{\text{Bi}}^{\text{LH}}$ and $E_{\text{Bi}}^{\text{HH}}$, vary with strain as $E_{\text{Bi}}^{\text{LH}} = \Delta E_{\text{Bi}} + \delta E_{\text{Bi}}^{\text{hy}} + \delta E_{\text{Bi}}^{\text{ax}}$ and $E_{\text{Bi}}^{\text{HH}} = \Delta E_{\text{Bi}} + \delta E_{\text{Bi}}^{\text{hy}} - \delta E_{\text{Bi}}^{\text{ax}}$, with $\delta E_{\text{Bi}}^{\text{hy}}$ and $\delta E_{\text{Bi}}^{\text{ax}}$ given by $\delta E_{\text{Bi}}^{\text{hy}} = a_{\text{Bi}} \left(\epsilon_{xx} + \epsilon_{yy} + \epsilon_{zz} \right)$ and $\delta E_{\text{Bi}}^{\text{ax}} = -\frac{b_{\text{Bi}}}{2} \left(\epsilon_{xx} + \epsilon_{yy} - 2\epsilon_{zz} \right)$. The Bi-impurity state hydrostatic and axial deformation potentials—determined using large supercell TB calculations to track the evolution of the Bi-related localized states as functions of hydrostatic and axial strain—are $a_{\text{Bi}} = -1.11$ eV and $b_{\text{Bi}} = -1.71$ eV, respectively.

Finally, in the section labeled "VBAC," the VBAC interactions between the strained virtual crystal band edges and the Bi-related localized states are included to finally determine the band offsets. Examining the 12-band Hamiltonian [54,79] we see that in bulk (with $k = 0$), the CB is decoupled from the Bi-related states, so that $E_{\text{CB}}(x, \epsilon) = \widetilde{E}_{\text{CB}}(x, \epsilon)$, while the host matrix HH-band couples directly to $E_{\text{Bi}}^{\text{HH}}$, giving

$$E_{\pm}^{\text{HH}}(x, \epsilon) = \frac{\widetilde{E}_{\text{HH}}(x, \epsilon) + E_{\text{Bi}}^{\text{HH}}}{2} \pm \sqrt{\left(\frac{\widetilde{E}_{\text{HH}}(x, \epsilon) - E_{\text{Bi}}^{\text{HH}}}{2} \right)^2 + |V_{\text{Bi}}|^2} \, . \tag{10.4}$$

Since the axial component of the pseudomorphic strain couples the host matrix LH and SO bands [83], there exists a second-order strain-induced coupling between the LH-like Bi-related localized states and the host matrix SO band. As a result, the SO band cannot be treated independently and the LH, SO, and lower-energy LH-like VBAC band edges in the strained alloy are obtained as the eigenvalues $E_{\pm}^{\text{LH}}(x, \epsilon)$ and $E_{\text{SO}}(x, \epsilon)$ of the 3 × 3 Hermitian matrix

$$\begin{pmatrix} \widetilde{E}_{\text{LH}}(x, \epsilon) & -\sqrt{2}\, \delta E_{\text{VB}}^{\text{ax}} & V_{\text{Bi}} \\ & \widetilde{E}_{\text{SO}}(x, \epsilon) & 0 \\ & & E_{\text{Bi}}^{\text{LH}} \end{pmatrix}, \tag{10.5}$$

which completes the calculation.

On the basis of detailed spectroscopic measurements performed on a series of GaBi$_x$As$_{1-x}$/(Al)GaAs QW laser structures with $x \approx 2\%$ [79], the Bi-related parameters α and β of Table 10.1 have been refined in order to produce an optimized set of parameters for modeling *real* GaBi$_x$As$_{1-x}$ QWs at low x, giving $\alpha = 2.63$ eV and $\beta = 1.45$ eV, with the remaining parameters unchanged from those listed in Table 10.1. These optimized Bi-related band structure parameters have been demonstrated to produce band offsets and transitions energies which are in good, quantitative agreement with those measured for a series of GaBi$_x$As$_{1-x}$/(Al)GaAs QWs. This not only verifies the predictive capability of the 12-band model, but also emphasizes its potential for application to the calculation and analysis of the electronic and optical properties of dilute bismide QWs and related heterostructures.

10.4 GaBi$_x$As$_{1-x}$/(Al)GaAs Quantum Well Lasers

Having focused so far on constructing a consistent approach to model the (In)GaBi$_x$As$_{1-x}$ band structure, in this section we turn our attention to the theory and modeling of dilute bismide semiconductor lasers. In order to address the potential of GaBi$_x$As$_{1-x}$ for the development of long-wavelength GaAs-based lasers, we introduce a theoretical model—introduced by Broderick et al. [15]—describe the impact of Bi on the performance of real and ideal QW laser devices. After outlining the theoretical model in Section 10.4.1, in Sections 10.4.2 and 10.4.3 we analyze the effect of Bi incorporation on the electronic and optical properties of an exemplar GaBi$_x$As$_{1-x}$/(Al)GaAs QW, which has been chosen on the basis of the active region employed in the first electrically pumped dilute bismide laser [30]. Using the theoretical model we review the Bi-induced changes to the band structure, DOS, optical transition matrix elements, and optical gain, and show how the choice of barrier material plays a key role in optimizing the performance of the device at low Bi composition.

Next, in Section 10.4.4, we review gain calculations for a GaBi$_x$As$_{1-x}$/GaAs laser structure with high Bi composition ($x = 13\%$), designed to emit at 1.55 μm and to have $\Delta_{\mathrm{SO}} > E_{\mathrm{g}}$ in order to suppress CHSH Auger recombination and IVBA (cf. Figure 10.1). We highlight that the modal and differential gain at this composition is significantly higher than that at lower x due to the combined effects of increased CB offset and compressive strain at larger x. Finally, in Section 10.4.5 we review comparison of the results of these theoretical calculations directly to experimental measurements of the SE and gain [39]. These are the first measurements of SE and optical gain for a dilute bismide semiconductor laser and the only such comparison between theory and experiment to date for this material system.

The theoretical results reviewed in this section elucidate important trends in the properties and performance of GaBi$_x$As$_{1-x}$ QW lasers grown on GaAs substrates, and confirm the potential of dilute bismide QW structures for the development of highly efficient 1.55-μm lasers. Comparison to experimental data displays the predictive capability of the theoretical model, highlighting its potential for use in the design and optimization of future dilute bismide devices.

10.4.1 Theoretical Model for Dilute Bismide Quantum Well Lasers

Here, we provide a brief outline of the theoretical model developed to study the electronic and optical properties of dilute bismide QW lasers. Full details of the model can be found in Ref. [15]. The theoretical model is based upon the 12-band **k·p** Hamiltonian presented in Section 10.3.2. The QW eigenstates are calculated in the envelope function approximation using a numerically efficient plane wave expansion method, similar to that widely employed in first-principles calculations. An overview of the plane wave expansion method within the context of **k·p** theory can be found in Ref. [85].

In the plane wave approach, the QW eigenstate $|\psi_n(k_\parallel, z)\rangle$ at position z and in-plane wave vector $k_\parallel$, having energy $E_n(k_\parallel)$, is written as

$$|\psi_n(k_\parallel, z)\rangle = \frac{1}{\sqrt{L}} \sum_{b=1}^{12} \sum_{m=-M}^{M} a_{nbm}(k_\parallel)\, e^{iG_m z} |u_b\rangle\,, \tag{10.6}$$

where L is the length of the calculational supercell ($-\frac{L}{2} \leq z \leq \frac{L}{2}$), b indexes the bands of the bulk **k·p** Hamiltonian, m indexes the plane waves, and the discrete wave vectors G_m are defined as $G_m = \frac{2m\pi}{L}$. $|u_b\rangle$ are the zone-center Bloch states of the bulk 12-band **k·p** Hamiltonian. We note that this description of the QW eigenstates, which consists of $2M + 1$ independent Fourier components, is formally exact as $M \to \infty$. However, for most practical applications, a basis set containing ~ 70 plane waves is sufficient to ensure that the calculated eigenstates have converged with respect to further increases in the basis set size [86].

By substituting Equation 10.6 into the multi-band Schrödinger equation, it is straightforward to derive analytical expressions for the matrix elements of the QW Hamiltonian in the plane wave basis set defined by the set of wave vectors G_m [85,86], leading to a reciprocal space Hamiltonian matrix of size $12(2M + 1) \times 12(2M + 1)$. In this manner, an independent Hamiltonian matrix is obtained at each in-plane wave vector $k_\parallel$, diagonalization of which yields the QW band structure $E_n(k_\parallel)$ and eigenstates $|\psi_n(k_\parallel, z)\rangle$. We note that the eigenstates defined by Equation 10.6 satisfy periodic boundary conditions over the calculational supercell, meaning that a QW will tend to undergo spurious interactions with identical QWs in neighboring supercells. This can readily be mitigated by choosing a sufficiently large supercell length L, and in practice this feature of the method means that it can be used to efficiently investigate the electronic and optical properties of superlattices.

Analysis of the band offsets (cf. Section 10.3.3) suggests that the small CB offset leads to very weak electron confinement in GaBi$_x$As$_{1-x}$/GaAs QWs at low x. As such, hole-induced electrostatic confinement of electron states is likely to play an important role in determining the electronic and optical properties of $GaBi_xAs_{1-x}/GaAs$ QW lasers at low x. For selected laser structures (cf. Section 10.4.3), we therefore also consider self-consistent calculations of the electronic and optical properties, in which the QW band structure and eigenstates are calculated self-consistently by using Poisson's equation in order to determine the net carrier-induced electrostatic potential at each injected carrier density n. For the self-consistent calculations, the coupled multi-band Schrödinger–Poisson system is solved in reciprocal space, using the technique outlined in Ref. [86].

The atomistic supercall calculations reviewed in Section 10.2 clearly demonstrate that Bi-induced hybridization plays an important role in determining the character of the VB edge eigenstates. As such, in order to facilitate accurate calculation of the optical properties of dilute bismide laser structures, it is important to include the impact of Bi incorporation on the QW eigenstates. For this reason, in the theoretical model the QW band structure, envelope functions, and Hamiltonian (calculated in the axial approximation [87]) are used *directly* to compute the optical properties of a given laser structure. Key to this analysis is the calculation of the optical (momentum) transition matrix elements. Following Ref. [88], the transition matrix element describing optical recombination of an electron-hole pair at in-plane wave vector $k_\parallel$ is given in terms of the QW eigenstates described by Equation 10.6 as [88]

$$P_{n_c,n_v}\left(k_\parallel\right) = \frac{m_0}{\hbar}\left\langle \psi_{n_v}\left(k_\parallel, z\right) \middle| \frac{\partial \widehat{H}}{\partial k_\parallel} \cdot \widehat{e} \middle| \psi_{n_c}\left(k_\parallel, z\right) \right\rangle, \tag{10.7}$$

where $\widehat{e}$ is a unit vector denoting the polarization of the emitted photon, and n_c (n_v) indexes the conduction (valence) subbands. This approach to calculating the optical matrix elements then explicitly incorporates key band structure effects in the computation of the optical properties including: Bi-induced VB hybridization (VBAC) and localization, temperature- and injection-dependent carrier spillout from the QW(s), band mixing at non-zero in-plane wave vector, and pseudomorphic strain, all of which are necessary to quantitatively describe the SE and optical gain in real devices [15,39]. Note that when discussing optical transition "strengths" below we are referring to the quantity $\frac{1}{2m_0}|P_{n_c,n_v}\left(k_\parallel\right)|^2$, which has units of energy.

Using the optical matrix elements computed in this manner, the calculation of the material gain spectrum $g(\hbar\omega)$ follows the approach of Ref. [89], which consists of transforming the corresponding polarization component of the SE rate $r_{\text{spon}}(\hbar\omega)$ at fixed carrier density n according to

$$g(\hbar\omega) = \frac{3\pi^2\hbar c^2}{2n_r^2\omega^2}\left(1 - \exp\left(\frac{\hbar\omega - \Delta F}{k_B T}\right)\right) r_{\text{spon}}(\hbar\omega), \tag{10.8}$$

where $\hbar\omega$ is the photon energy, ΔF is the quasi-Fermi level separation at temperature T in the presence of an injected carrier density n, and n_r is the refractive index of the optical mode in the active region. The SE rate r_{spon} is calculated as described in Refs. [15,89]. Here, we are concerned solely with transverse electric (TE)-polarized gain, which is of larger magnitude than the transverse magnetic (TM)-polarized gain in the compressively strained structures under investigation. As such, r_{spon} is understood to represent only the TE-polarized component of the total SE rate (which gives rise to the leading factor of $\frac{3}{2}$ in Equation 10.8). This approach to calculating the optical gain spectrum has the benefits that (1) it removes the anomalous absorption at energies below the QW band gap that can plague calculations undertaken using an energy and crystal momentum-independent interband relaxation time (homogeneous spectral linewidth), and (2) by definition, $g(\Delta F) = 0$, so that the transparency point on the high-energy side of the gain peak is maintained at the correct, thermodynamically consistent energy. As we will see, when we review the comparison of theory and experiment for the optical gain in Section 10.4.5, this latter characteristic of Equation 10.8 enables a distinct carrier density n in the theoretical calculations to be associated with each current density J at which the gain spectrum is measured, thereby facilitating *direct* comparison of theoretical calculations to experiment.

Based on theory-experiment comparisons, it is found that the spectral broadening of $GaBi_xAs_{1-x}$ QWs is well described using a hyperbolic secant lineshape of the form $S(\hbar\omega) = \frac{1}{\pi\delta}\text{sech}\left(\frac{E_0-\hbar\omega}{\delta}\right)$, where δ is the (homogeneous) spectral linewidth. We note that this lineshape function was previously found to describe well the SE spectra of 1.3-μm GaInNAs dilute nitride QW lasers [90]. For the calculations in Section 10.4.2 we set $\delta = 6.6$ meV, a typical value for III–V semiconductors such as InGaAs. This enables direct comparison between the calculated SE and gain in Bi-free and Bi-containing structures, and hence quantitative analysis on the impact of Bi incorporation on the optical properties. This value of the spectral linewidth is smaller than that observed in real $GaBi_xAs_{1-x}$ devices by a factor of approximately four [15]. Larger more realistic linewidths are included in several of the calculations presented in Sections 10.4.2 and 10.4.3 by inhomogeneously broadening the calculated spectra [15], and the spectral linewidth is directly analyzed in Section 10.4.5 when review the comparison to experimental measurements of the SE and modal gain spectra [39].

For the calculation of the threshold properties in Sections 10.4.2 and 10.4.3, internal optical (cavity) losses of $\alpha_i = 4\ \text{cm}^{-1}$ are assumed, lattice-matched $Al_{0.4}Ga_{0.6}As$ cladding layers are chosen to form the separate confinement heterostructure (SCH), and an overall cavity length of 1 mm is assumed for all of the devices considered [30]. The calculated facet reflectivity for this structure is $R = 0.35$. The confinement factor Γ of the fundamental (TE-polarized) optical mode was calculated for each laser structure using an effective index approach [91]. It was found for all laser structures considered that Γ is maximized for 150-nm-thick (Al)GaAs barriers. Alternative approaches, based on similar models of the band structure, have been proposed by other authors [14,92]. However, the theoretical model presented here benefits from detailed and unambiguous parameterization based on atomistic calculations, and has been confirmed to provide reliable predictions of key trends in the properties and performance of GaAs-based dilute bismide QW lasers [15], including descriptions of the measured SE and optical gain that are in quantitative agreement with experimental measurements on real devices [39].

10.4.2 Impact of Bi Incorporation on the Band Structure and Gain of Quantum Wells

In order to quantify the impact of Bi incorporation on the electronic and optical properties of laser structures having low Bi compositions, we compare the calculated gain characteristics of a $GaBi_xAs_{1-x}$ QW laser structure at low x with those of an equivalent laser structure containing a GaAs (Bi-free, $x = 0$) QW [15]. We focus here on Bi-containing laser structures having a single 6.9-nm-thick $GaBi_{0.021}As_{0.979}$ ($x = 2.1\%$) QW with $Al_{0.144}Ga_{0.856}As$ barriers [30,79].

The calculated CB and VB offsets are $\Delta E_{CB} = 150$ meV and $\Delta E_{HH/LH} = 65$ meV in the unstrained Bi-free QW [79]. The incorporation of 2.1% Bi introduces a compressive in-plane strain of 0.25% and, when VBAC effects are taken into account, this leads to calculated HH and LH band offsets $\Delta E_{HH} = 224$ meV and $\Delta E_{LH} = 208$ meV, and a CB offset $\Delta E_{CB} = 185$ meV [15,79]. The transition energy between the lowest-energy bound electron state ($e1$) and the highest-energy bound hole state ($hh1$) is calculated to be 1.480 eV in the Bi-free QW, which is reduced to 1.289 eV in the $x = 2.1\%$ QW. Of this 191 meV reduction in the QW band gap, only 31 meV is due to the Bi-induced reduction of the CB edge energy, while the remaining 160 meV results from the VBAC-induced upward shift of the VB edge energy.

Figure 10.7 shows the calculated VB structure and DOS for the $x = 0$ and 2.1% QWs. We focus our analysis of the band structure and DOS on the VB, since Bi incorporation has a comparatively minor impact on the CB (cf. Section 10.2). The zero of energy is taken at the energy of the $hh1$ state (i.e., the QW VB edge) in each case, to facilitate a direct comparison of the DOS. Three factors contribute to differences in the band structure and DOS between the two QWs: (1) the compressive strain in the Bi-containing structure leads to a slight increase in the energy separation between $hh1$ and the highest-energy bound LH state ($lh1$), leading in turn to a reduction in the DOS at energies from 0 to −20 meV, (2) the VBAC interaction pushes the VB dispersion upward in energy at fixed $k_{\|}$, increasing the DOS in the $x = 2.1\%$ QW relative

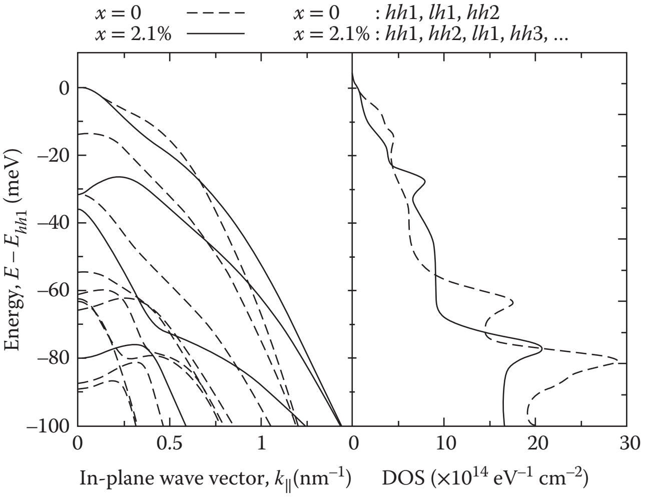


FIGURE 10.7 Calculated band structure (left panel) and density of states (DOS; right panel) in the vicinity of the VB edge for 6.9-nm-thick GaAs (Bi-free; dashed lines) and $GaBi_{0.021}As_{0.979}$ ($x = 2.1\%$; solid lines) QWs with $Al_{0.144}Ga_{0.856}As$ barriers. (Adapted from Broderick, C.A. et al., *IEEE J. Sel. Top. Quantum Electron.*, 21, 1503313, 2015.) Note that the zero of energy has been taken in each case at the energy of the highest bound hole state (*hh*1) to facilitate comparison between the two QWs.

to the Bi-free QW for energies between −20 and −60 meV, and (3) the barrier VB edge lies approximately 60 meV (200 meV) below *hh*1 in energy in the $x = 0$ ($x = 2.1\%$) QW, so that the calculated DOS in the Bi-free structure increases above that of the $x = 2.1\%$ QW for energies between −60 and −100 meV due to the large density of VB states in the AlGaAs barrier. We note that, overall, the calculated impact of Bi incorporation on the VB structure and DOS are relatively minor at low Bi compositions $x \approx 2\%$.

Having considered the impact of Bi incorporation on the QW band structure and DOS, we now turn our attention to the impact of strain and Bi-induced VBAC on the interband optical transition strengths, and then investigate the effect of Bi incorporation on the material gain. Figure 10.8a shows the TE-polarized optical transitions strengths, calculated as a function of $k_{\|}$ for the *e*1-*hh*1 and *e*1-*lh*1 transitions, for the Bi-free (dashed lines) and Bi-containing (solid lines) QWs. Incorporation of Bi reduces the optical transition strength at $k_{\|} = 0$ for transitions between the *e*1 and *hh*1/*lh*1 states. For example, the *e*1–*hh*1 transition strength at $k_{\|} = 0$ in the Bi-containing QW is approximately 67% of that in the Bi-free QW. The majority of the reduction in the optical transition strengths at $x = 2.1\%$ is due to the VBAC effect, which accounts for all but 1.1% of the calculated 33.4% reduction for the TE-polarized *e*1–*hh*1 transition in the $x = 2.1\%$ QW. This is due to the fact that the hybridization of the bound hole states in the QW with Bi-related localized states leads to a portion of the hole states consisting of an admixture of *p*-like Bi-related localized states which do not couple optically to the *s*-like states at the CB edge [54] (cf. Section 10.3.2).

Next, we investigate the variation of the peak material gain as a function of carrier density *n*, for temperatures $T = 100$, 200, and 300 K. The results of these calculations are shown in Figure 10.8b for the $x = 2.1\%$ and $x = 0$ laser structures using solid and dashed lines, respectively. Since the material gain at fixed *n* is inversely proportional to the photon energy, to facilitate a direct comparison of the gain characteristics of the two QW structures—which have different band gaps—all CB states in the $x = 0$ calculation have been shifted downward in energy by 191 meV in order to bring the *e*1–*hh*1 transition energies into coincidence for both QW structures. Therefore, the differences between the material gain for two laser structures shown in Figure 10.8b arise only due to the impact of Bi incorporation on the QW band structure and optical transition strengths. Based on a calculated optical confinement factor $\Gamma = 1.66\%$ for the

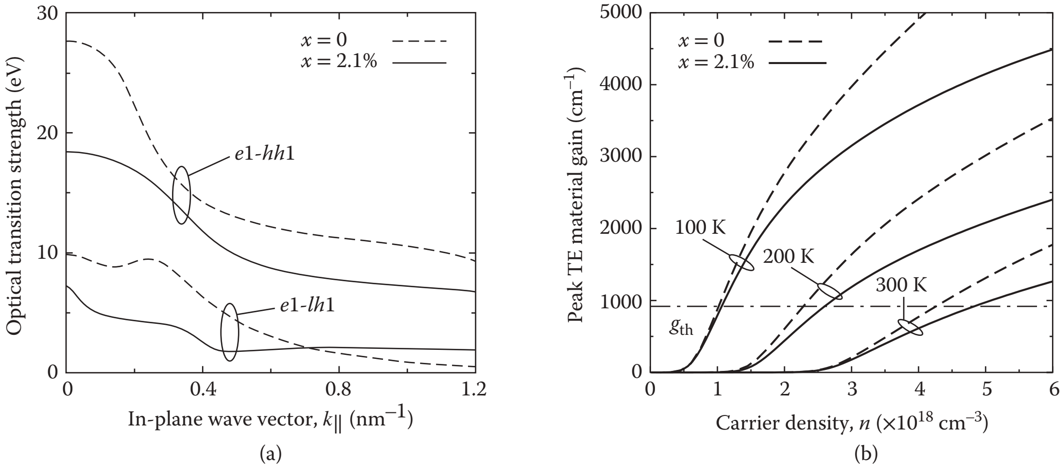


FIGURE 10.8 (a) Calculated TE-polarized optical transition strengths as a function of in-plane wave vector ($k_{||}$), for optical transitions between the lowest-energy bound electron and highest-energy bound LH- and HH-like states for 6.9-nm-thick GaAs (Bi-free; dashed lines) and GaBi$_{0.021}$As$_{0.979}$ ($x = 2.1\%$; solid lines) QWs having Al$_{0.144}$Ga$_{0.856}$As barriers. (Adapted from Broderick, C.A. et al., *IEEE J. Sel. Top. Quantum Electron.*, 21, 1503313, 2015.) (b) Calculated variation of the peak TE-polarized material gain as a function of carrier density, at temperatures of 100, 200, and 300 K for the same QWs as in (a). (Adapted from Broderick, C.A. et al., *IEEE J. Sel. Top. Quantum Electron.*, 21, 1503313, 2015.) The horizontal dash-dotted line denotes the calculated threshold material gain, $g_{th} = 919$ cm^{-1}.

$x = 2.1\%$ laser structure, we estimate a threshold material gain $g_{th} = 919$ cm^{-1} (denoted by a horizontal dash-dotted line in Figure 10.8b).

There is little change in the calculated transparency carrier density, n_{tr}, between the $x = 0$ and 2.1% laser structures for all temperatures considered. This reflects the small difference between the calculated VB edge DOS for the two QWs (cf. Figure 10.7). However, we find for $n > n_{tr}$ that (1) the gain saturates at a lower level with increasing n at $x = 2.1\%$, and (2) the threshold carrier density, n_{th}, is higher at $x = 2.1\%$ than in the Bi-free QW. Both of these features arise due primarily to the general reduction in the optical transition strengths in the Bi-containing QW structure. We also note that while the differential gain at threshold, $\frac{dg}{dn}$, is approximately equal for both the Bi-free and Bi-containing lasers structures at 100 K, $\frac{dg}{dn}$ for the Bi-containing QW decreases with respect to that of the Bi-free QW with increasing temperature. Since the modulation response frequency of a semiconductor laser is proportional to the square root of the differential gain [93,94], these calculations suggest that Bi incorporation can be expected to lead to a degradation of the dynamical performance at low Bi compositions compared to that of an equivalent Bi-free structure.

10.4.3 Optimization of GaBi$_x$As$_{1-x}$/(Al)GaAs Laser Structures

Theoretical and experimental analysis has shown that GaBi$_x$As$_{1-x}$/GaAs QWs have very small CB offsets at low x, estimated at only 55 meV for a QW with $x = 2.1\%$ [79]. Such a low CB offset is detrimental to laser operation, leading to low electron-hole spatial overlap and significant electron leakage from the QW at high temperatures and carrier densities, and is therefore a major factor limiting the material gain at low x. Electron confinement must then be improved by growing GaBi$_x$As$_{1-x}$ QWs with Al$_y$Ga$_{1-y}$As barriers to increase the CB offset, but at the cost of reducing the optical confinement in a laser structure with a fixed Al composition y in in the cladding layers. This leads to a trade-off between the material gain (which increases with increasing y due to improved electron confinement) and the optical confinement (which decreases with increasing y due to a reduced refractive index contrast between the barrier and cladding layers). The theoretical model described above has been used to determine the Al composition

y in the barrier layers required to optimize the performance of a laser structure containing the same 6.9-nm-thick $GaBi_{0.021}As_{0.979}$ QW considered in the exemplar structure of Section 10.4.2. The modal gain of the laser structure, Γg, is the quantity optimized, in order to minimize n_{th}. The gain calculations for this optimization include (1) electrostatic effects, computed self-consistently via coupling the Poisson's equation, and (2) inhomogeneous broadening, in order to produce realistic estimates of the material gain in the laser structure [15], an approach which has been shown to be in excellent agreement with experiment [39,95].

Figure 10.9 shows the calculated variation of n_{th} and g_{th} with the barrier Al composition y for this series of laser structures. Examining first the variation of g_{th} with y (closed circles, solid line) we see that the degradation of the optical confinement when Al is incorporated in the barrier layers leads to an increase in g_{th} with increasing y. This increase in g_{th} is modest at low Al compositions $y < 10\%$, with the optical confinement factor at $y = 10\%$ being 90% of its maximum calculated value $\Gamma = 1.95\%$ at $y = 0$. For $y > 10\%$ Γ degrades rapidly with increasing y, reaching values as low as 1.24% at $y = 25\%$. This results in a large increase in g_{th}, since $g_{th} \propto \Gamma^{-1}$. The net effect of this trend on the laser performance can be seen by considering the variation of n_{th} with y (open circles, dashed line). Beginning with a QW having GaAs barriers ($y = 0$), a large threshold carrier density $n_{th} = 6.3 \times 10^{18}$ cm^{-3} is calculated. When Al is incorporated in the barrier layers n_{th} decreases rapidly with increasing y, reaching a minimum value for $10\% < y < 15\%$. For $y < 10\%$, we therefore conclude that the increase in the material gain at fixed carrier density arising from the enhanced CB offset and electron confinement is sufficient to overcome the associated reduction in Γ, leading to a strong reduction in n_{th}. However, for $y > 15\%$, we find that n_{th} increases rapidly with increasing y, reflecting that any further improvement in the material gain is insufficient to overcome the rapid further degradation of Γ and associated increase in g_{th}. These theoretical calculations therefore suggest that QWs having barrier Al compositions $y = 10\%–15\%$ offer enhanced performance as compared to a QW having GaAs barriers, for $x \approx 2\%$.

While the first electrically pumped $GaBi_xAs_{1-x}$ QW laser had a nominal barrier Al composition $y = 20\%$ [30], more recently improved device performance has been demonstrated following the structural optimization outlined here [96]. Specifically, measurements confirm that choosing $Al_{0.12}Ga_{0.88}As$ as opposed to $Al_{0.20}Ga_{0.80}As$ barriers leads to a decrease of the threshold current density, J_{th}, by a factor of

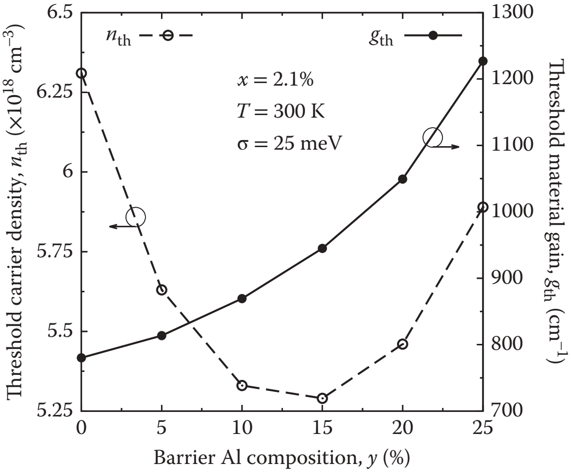


FIGURE 10.9 Calculated variation of the threshold carrier density (n_{th}; open circles, dashed lines) and threshold material gain (g_{th}; closed circles, solid lines) as a function of the barrier Al composition y for a series of 6.9-nm-thick $GaBi_{0.021}As_{0.979}/Al_yGa_{1-y}As$ QWs. (Adapted from Broderick, C.A. et al., *IEEE J. Sel. Top. Quantum Electron.*, 21, 1503313, 2015.) The calculated gain spectra, from which the values of n_{th} were determined, include an inhomogeneous broadening of 25 meV.

TABLE 10.2 Measured Room Temperature Emission Wavelength λ_{meas} and Threshold Current Density J_{th} for a Series of $GaBi_xAs_{1-x}/Al_yGa_{1-y}As$ Laser Structures Containing Nominally Identical 6.4-nm-Thick $GaBi_{0.022}As_{0.978}$ ($x = 2.2\%$) QWs

QW Material	Barrier Material	λ_{meas} (nm)	J_{th} (kA cm^{-2})
$GaBi_{0.022}As_{0.978}$	GaAs	938	7.5
$GaBi_{0.022}As_{0.978}$	$Al_{0.12}Ga_{0.88}As$	947	1.0–1.1
$GaBi_{0.022}As_{0.978}$	$Al_{0.20}Ga_{0.80}As$	947	1.5–1.6

Source: Marko et al., *J. Phys. D: Appl. Phys.*, 47, 345103, 2014.

The measured variation of J_{th} with barrier Al composition y closely follows the trend calculated for the threshold carrier density n_{th} (cf. Figure 10.9), verifying that J_{th} can be minimized by varying y to control the trade-off between the carrier and optical confinement.

approximately three. The results of these measurements, which emphasize the importance of AlGaAs barriers for the design and optimization of $GaBi_xAs_{1-x}$ QW lasers at low x, are summarized in Table 10.2. Full details of this optimization analysis can be found in Refs. [15] and [96]. Examining the experimental data in Table 10.2 we can clearly identify the trends predicted by the theoretical calculations. First, the laser structure with Al-free (GaAs) barriers offers poor material gain leading to high J_{th}. Second, incorporation of 12% Al in the barrier (close to the optimum composition suggested by the theoretical calculations) results in a strong reduction of J_{th} as a consequence of the enhanced modal gain. Finally, increasing the Al composition in the barrier from 12% to 20% brings no further improvement in the laser performance; an increase in J_{th} is observed, in line with the degradation of the optical confinement and model gain predicted by the theoretical calculations.

Since the ultimate goal is to grow $GaBi_xAs_{1-x}$ lasers with Bi compositions $x > 10\%$, this analysis has been repeated for a series of $GaBi_xAs_{1-x}$/(Al)GaAs QW laser structures as a function of x. In general, it is found that Al incorporation in the barrier layers is required to enhance the otherwise weak electron confinement for $x < 5\%$. For $x = 6\%$, the $GaBi_xAs_{1-x}$/GaAs CB offset is sufficiently large that incorporating Al in the barrier layers has a negligible impact on the material gain. We therefore conclude that AlGaAs barriers are required to optimize the performance of $GaBi_xAs_{1-x}$ QWs with $x < 6\%$. For higher Bi compositions, GaAs barriers should suffice to provide modal gain which exceeds that calculated for the $x = 2.1\%$ laser structure considered above at fixed n, provided that QWs with $x > 6\%$ can be grown with sufficiently high material, structural, and optical quality. This conclusion has proved useful, particularly for the growth of hybrid laser structures such as those presented in Ref. [35], where the cladding and part of the barrier layers are grown by metal-organic vapor phase epitaxy (MOVPE) and the active region by molecular beam epitaxy (MBE). This analysis suggests that, for sufficiently large x, GaAs can be used as the final layer grown before a laser substructure is transferred in either direction between MOVPE and MBE growth chambers, thereby minimizing the impact of AlGaAs-related degradation during transport and storage of partly grown laser structures.

10.4.4 Gain Characteristics of $GaBi_xAs_{1-x}$ Quantum Wells at $\lambda = 1.55$ μm

Having elucidated the consequences of Bi incorporation for the gain characteristics and performance of real $GaBi_xAs_{1-x}$ QW lasers at low x, we now turn our attention to an ideal 1.55-μm device at higher Bi composition. We again choose a 6.9-nm-thick $GaBi_xAs_{1-x}$ QW, in order to compare the calculated modal gain directly to that calculated for the $x = 2.1\%$ QW of Sections 10.4.2 and 10.4.3, and, in light of the analysis described in Section 10.4.3, we choose GaAs as the barrier material. For this 6.9-nm-wide $GaBi_xAs_{1-x}$/GaAs QW, we calculate that a Bi composition $x = 13\%$ produces an emission wavelength close to 1.55 μm. We note that the bulk band structure of the $GaBi_{0.13}As_{0.87}$ QW material has $\Delta_{SO} > E_g$, so that it should be possible to use $GaBi_xAs_{1-x}$ alloys to develop, on a GaAs substrate, a 1.55-μm laser in which the CHSH Auger recombination and IVBA processes are suppressed (cf. Figure 10.1) [12,15].

Incorporation of 13% Bi in the QW produces an in-plane compressive strain of 1.53%, compared to 0.25% in a QW with $x = 2.1\%$. We calculate that the tendency of the VBAC interaction to increase the effective masses of the bound hole states in the QW is overcome by the strain-induced reduction of the in-plane hole effective masses [93], leading to a significant reduction of the DOS in the vicinity of the VB edge at $x = 13\%$ [15]. This strong reduction of the DOS at energies close to the QW VB edge should then enhance the performance of this high Bi composition QW compared to the low x laser structures considered above.

Compared to the value $\Gamma = 1.66\%$ calculated for the optical confinement factor of the $x = 2.1\%$ laser structure of Section 10.4.2, we calculate $\Gamma = 1.26\%$ in the $x = 13\%$ QW. The reduction in Γ results from the longer emission wavelength of the $x = 13\%$ structure and leads to an increase of the material gain at threshold g_{th}. We note that the degradation of the optical confinement a long wavelengths can be mitigated by increasing the barrier thickness. We have not pursued this approach here, as our intention is to analyze only the impact of changing the Bi composition in otherwise identical laser structures. Based on the calculated difference in Γ between the two laser structures we calculate $g_{\text{th}} = 1208\ \text{cm}^{-1}$ at $x = 13\%$, which is 32% larger than the value of 919 cm^{-1} calculated for the $x = 2.1\%$ laser structure. However, the calculated CB offset of 245 meV in the $GaBi_{0.13}As_{0.87}/GaAs$ QW is 60 meV larger than that calculated for the $GaBi_{0.021}As_{0.979}/Al_{0.144}Ga_{0.856}As$ QW above, leading to enhanced electron confinement in the 1.55-μm QW without the need to incorporate Al in the barrier layers. The calculated optical transition strength at $k_{\|} = 0$ for the $e1$–$hh1$ optical transition in the $x = 13\%$ QW is approximately 90% of that calculated at $x = 2.1\%$, suggesting that the Bi-induced decrease of the optical transition strengths described above is most prominent at low x, but has little additional effect as x is increased. When combined with the favorable impact of strain on the VB structure and DOS at larger values of x, these effects together ensure that the material gain at fixed n is significantly improved in the $x = 13\%$ QW compared to the $x = 2.1\%$ laser structure considered above.

The $x = 13\%$ laser structure has a calculated threshold carrier density $n_{\text{th}} = 3.42 \times 10^{18}\ \text{cm}^{-3}$ at 300 K, which is approximately 60% of that calculated for the $x = 2.1\%$ laser structure. Additionally, for the $x = 13\%$ laser structure $\frac{dg}{dn} = 2.76 \times 10^{-16}\ \text{cm}^2$ for the differential (material) gain at threshold, which is approximately a factor of two larger than the value of $1.32 \times 10^{-16}\ \text{cm}^2$ calculated for the $x = 2.1\%$ laser structure. This analysis suggests that the marked improvement in the threshold characteristics of the device at $x = 13\%$ results primarily from the increased compressive strain in the QW. First, the strain-induced reduction in the DOS at the VB edge leads to an increase in the quasi-Fermi level separation at fixed n, meaning that population inversion can be achieved at lower carrier densities. Second, the reduced in-plane hole effective masses ensure that holes occupy VB states over a smaller range of $k_{\|}$ (i.e. a narrow **k**-space distribution) at $x = 13\%$ than at $x = 2.1\%$, better matching the distribution of electrons in the CB and hence ensuring that more carriers are available to contribute to the lasing mode [97]. These factors ensure that a larger increase in material gain can be obtained for a given change in n at $x = 13\%$, leading to the calculated significant improvement in $\frac{dg}{dn}$.

Previous calculations for InP-based QW structures designed to emit at 1.55 μm suggest that the maximum $\frac{dg}{dn}$ obtainable using optimized quaternary InGaAsP QWs are $\sim 14 \times 10^{-16}\ \text{cm}^2$ for a multi-QW device containing four QWs [98]. While the calculations here suggest that GaAs-based $GaBi_xAs_{1-x}$ QWs offer slightly lower $\frac{dg}{dn}$ than conventional InP-based devices, we recall that this analysis has been performed on a prototypical 1.55-μm dilute bismide structure in which the waveguide has not been optimized. We therefore conclude that 1.55-μm $GaBi_xAs_{1-x}$ QW lasers have the potential to offer differential gain, and hence dynamical performance, that compares favorably with existing InP-based devices, particularly for structures incorporating multiple $GaBi_xAs_{1-x}$ QWs. We further note that the calculated value of $\frac{dg}{dn}$ for this 1.55-μm dilute bismide device exceeds that calculated for ideal 1.3-μm GaIn(N)As/GaAs QW structures [99].

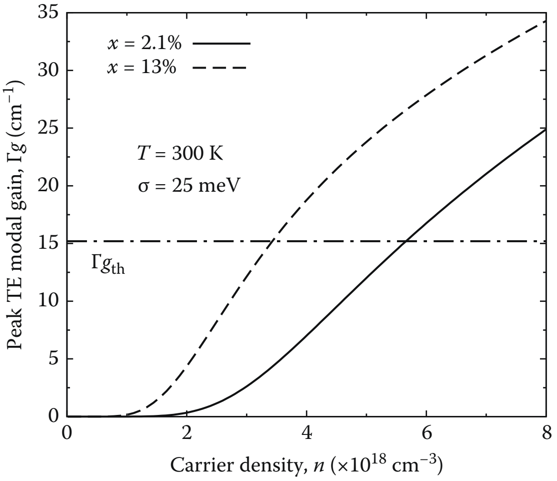


FIGURE 10.10 Calculated variation of the peak TE-polarized modal gain as a function of carrier density at 300 K for the $GaBi_{0.021}As_{0.979}/Al_{0.144}Ga_{0.856}As$ (solid line) and $GaBi_{0.13}As_{0.87}/GaAs$ (dashed line) QWs described in the text. (Adapted from C.A. Broderick et al., *IEEE J. Sel. Top. Quantum Electron.*, 21, 1503313, 2015.) The $x = 2.1\%$ structure has a room temperature emission wavelength ~950 nm, while the $x = 13\%$ structure is designed to have $\Delta_{SO} > E_g$ and to emit at 1.55 μm. The horizontal dash-dotted line denotes the calculated threshold modal gain, $\Gamma g_{th} = 15.2$ cm^{-1}.

Figure 10.10 shows the calculated variation of the modal gain as a function of n for the $x = 2.1\%$ (solid line) and $x = 13\%$ (dashed line) laser structures at 300 K. We see that, even in the presence of significant inhomogeneous broadening of the gain spectrum, the enhancement of the material gain described above is sufficient to overcome the reduction in Γ between $x = 2.1\%$ and 13%, leading to significantly improved modal gain at fixed n. Additionally, we note that the differential modal gain at threshold is significantly improved at $x = 13\%$, due to the fact that the calculated increase in the differential material gain at threshold is also sufficient to overcome the reduction in Γ between $x = 2.1\%$ and 13%. We recall that optimization of the waveguide can be undertaken for long-wavelength emission, which can be expected to further improve the performance of QW structures having high Bi compositions over that of low x structures operating at shorter wavelengths.

We therefore conclude overall that the performance of ideal $GaBi_xAs_{1-x}$ QW lasers at high Bi compositions ($x > 10\%$) should be significantly improved compared to that of shorter wavelength devices at lower x. The calculated improvement in the threshold characteristics with increasing x is primarily attributable to the associated increase in compressive strain in the QW when more Bi is incorporated. We also emphasize that the presence of a QW band structure in which $\Delta_{SO} > E_g$ promises to deliver even greater benefit for the laser operation, due to the elimination of the dominant non-radiative CHSH Auger recombination and IVBA losses.

10.4.5 Theory versus Experiment for First-Generation $GaBi_xAs_{1-x}$/(Al)GaAs Lasers

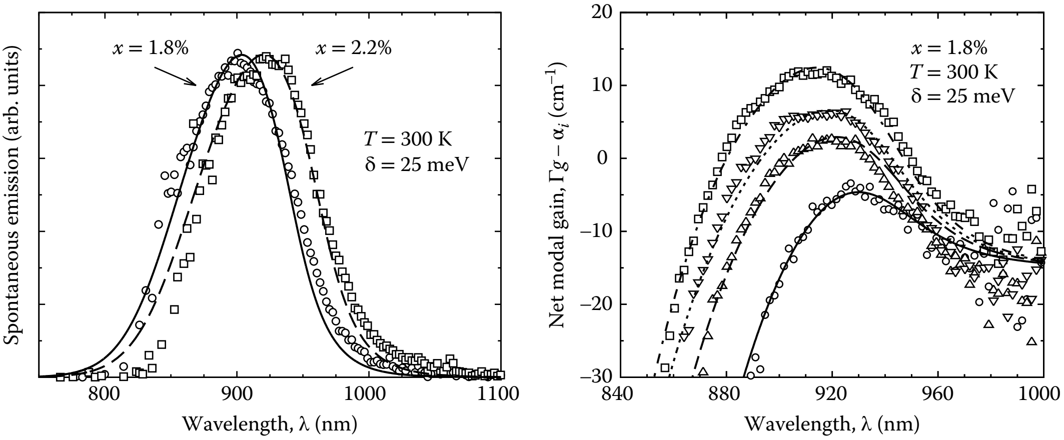
Having explored general trends in the electronic and optical properties of $GaBi_xAs_{1-x}$/(Al)GaAs QW lasers, we now compare the results of the theoretical calculations directly to experimental measurements of the SE and gain spectra. The experimental measurements were carried out using the segmented contact method, which allows measurement of the optical absorption, SE, and gain spectra of a multi-section device [100]. The multi-section devices used in this study were fabricated from a wafer on which a single QW $GaBi_xAs_{1-x}$/(Al)GaAs laser structure was grown. The laser structure is as described in Ref. [30]

and consists of $Al_{0.4}Ga_{0.6}As$ cladding layers, 150-nm-thick AlGaAs barriers and a $GaBi_xAs_{1-x}$ QW with a nominal thickness of 6.4 nm.

Two multisection devices were fabricated in order to carry out the segmented contact measurements: (1) a device fabricated from material located close to the edge of the wafer ("device 1"), and (2) a device fabricated from material located closer to the center of the wafer ("device 2"). Full details of the growth and fabrication of these devices, as well as the theory-experiment comparison to be outlined here, can be found in Ref. [39]. While the barrier has a nominal Al composition of 20%, based on the analysis of Ref. [79] we find that the actual composition is 14.4%—we use the latter in the theoretical calculations, as in Section 10.4.2.

Comparing the measured optical absorption spectra for the two multisection devices, it is found that (1) the optical (cavity) losses are in the range $\alpha_i = 10–15\ \text{cm}^{-1}$ at room temperature, and (2) the absorption edge measured for device 2 is red-shifted by approximately 20 nm compared to the that measured for device 1 [39]. The latter suggests the presence of Bi composition fluctuations across the wafer, with the Bi composition being slightly reduced toward the wafer edge. This is consistent with the measurements of Ref. [79], in which $GaBi_xAs_{1-x}$/(Al)GaAs laser structures having nominally identical $x = 2.2\%$ QWs-displayed variations of approximately 30 meV in the QW band gap (corresponding in theoretical calculations Bi compositional variations of up to ± 0.4%).

In order to investigate this behavior, we first analyze the measured SE spectra for the two devices. Figure 10.11a shows the SE spectra measured at threshold using the segmented contact method for device 1 (closed circles) and device 2 (closed squares). Both spectra have similar overall shape with the main difference between them being the wavelength of the SE peak, which is approximately 20 nm shorter in device 1. This shifted emission peak is consistent with the observed shift in the absorption edge between the two



(a) (b)

FIGURE 10.11 (a) Measured (using the segmented contact method; open symbols) and calculated (lines) SE spectra at threshold for a multisection device fabricated from material located close to the wafer edge (open circles; solid line) and fabricated from material located close to the center of the wafer (closed squares; dashed line). The ~20 nm shift in the SE peak wavelength between the two devices suggests the presence of Bi compositional variations of up to 0.4% across the wafer from which the devices were fabricated. (Adapted from Marko, I.P. et al., *Sci. Rep.*, 6, 28863, 2016.) (b) Measured (using the segmented contact method; closed symbols) and calculated (lines) net modal gain spectra for the device fabricated from material located at the wafer edge, at injected current densities $J = 0.7$, 1.4, 2.0, and 2.4 kA cm^{-2} (open circles, upright triangles, inverted triangles, and squares, respectively). The carrier densities used in the theoretical calculations (determined as outlined in the text) are $n = 5.12$, 7.11, 8.24, and $9.38 \times 10^{18}\ \text{cm}^{-3}$; the corresponding gain spectra are shown using solid, dashed, dotted, and dash-dotted lines, respectively. (Adapted from Marko, I.P. et al., *Sci. Rep.*, 6, 28863, 2016.)

devices. In order to quantify this variation in Bi composition across the wafer, and to facilitate theoretical analysis of the measured gain spectra for device 1, the theoretical model has been used to analyze the SE spectra shown in Figure 10.11a. The careful parameterization of the 12-band $\mathbf{k}\cdot\mathbf{p}$ model for the $\mathrm{GaBi}_x\mathrm{As}_{1-x}$ band structure described in Section 10.3 enables us to treat the Bi composition x in the QW to be treated as the only free variable in the analysis of the SE emission peak. Through the theoretical calculations we find that the measured SE peak wavelength of 910 nm for device 1 corresponds to a Bi composition $x = 1.8\%$ in the QW (assuming that the QW thickness to be is fixed at its nominal value of 6.4 nm).

Next, by comparing the full calculated SE spectrum for device 1 (shown in Figure 10.11a using a solid line) to the experimental data, we find that the spectral broadening is best described by a hyperbolic secant lineshape with a homogeneous linewidth $\delta = 25$ meV. The theoretical calculation was also carried out at the threshold injection level, which was determined as the carrier density required to produce the TE-polarized threshold material gain $g_{\mathrm{th}} = 1325\ \mathrm{cm}^{-1}$ (computed using an effective index calculation to determine $\Gamma = 1.60\%$ for a QW containing 1.8% Bi, and using the measured threshold current density $J_{\mathrm{th}} \approx 1.6\ \mathrm{kA\ cm}^{-2}$ in similar devices [96]). Due to uncertainty in the absolute units and relative intensity of the SE measured for the two devices, the calculated SE spectrum has been normalized to the measured value at the SE peak in order to compare the theoretical and experimental data. We note that the measured and calculated SE spectra are in good overall agreement, with the spectral broadening observed in the experiment well described by a combination of the energy broadening at room temperature of the electron and hole (quasi-Fermi) distribution functions, as well as the hyperbolic secant line broadening.

In order to quantify the variation of the Bi composition across the wafer the SE spectrum for device 2 has also been calculated. The calculation of the SE spectrum in this case proceeds as above, this time with the Bi composition again being the only parameter allowed to vary. The result of this calculation is shown using a dashed line in Figure 10.11a. We find that (1) the measured SE peak at 932 nm is well described in the theoretical calculations, assuming that the QW contains the nominal Bi composition of 2.2%, and (2) the calculated SE spectrum for $x = 2.2\%$ is in good overall agreement with experiment for this device. Overall, these theoretical results suggest that the Bi composition in the QW is close to the nominal value of 2.2% near the center of the wafer, but is slightly reduced, to 1.8%, toward the wafer edges. We recall that this variation in Bi composition is consistent with that determined previously via spectroscopic measurements [79].

The closed symbols in Figure 10.11b show the (TE-polarized) net modal gain spectra, $\Gamma g - \alpha_{\mathrm{i}}$, measured using the segmented contact method for device 1. These gain spectra, which were measured at room temperature for current densities $J = 0.7$, 1.4, 2.0, and 2.4 kA cm^{-2}, are shown using closed circles, upright triangles, inverted triangles and squares, respectively. The measured gain spectra are relatively broad, with a full width at half maximum of approximately 100 meV at a current density of 2 kA cm^{-2}, which is close to twice that observed for an InGaAs/GaAs SQW laser operating at a similar wavelength [101] and is most likely related to the strong Bi-induced inhomogeneous broadening that is characteristic of the optical spectra of $\mathrm{GaBi}_x\mathrm{As}_{1-x}$ alloys [27,28]. Based on the $\alpha_{\mathrm{i}} = 15\ \mathrm{cm}^{-1}$ optical losses measured for device 1, the estimated peak modal gain at $J = 2\ \mathrm{kA\ cm}^{-2}$ is $\Gamma g_{\mathrm{peak}} = 24\ \mathrm{cm}^{-1}$. Using the calculated optical confinement factor for this laser structure, $\Gamma = 1.60\%$, we estimate a peak material gain $g_{\mathrm{peak}} \approx 1500\ \mathrm{cm}^{-1}$ at $J = 2\ \mathrm{kA\ cm}^{-2}$, which agrees well with the value of 1560 cm^{-1}, calculated using the theoretical model.

In order to compare the calculated gain spectra to experiment we must determine the carrier density n corresponding to each current density J at which each of the gain spectra depicted by open circles in Figure 10.11b was measured. In order to do so, we recall from Equation 10.8 that the transparency point at which there is zero material/modal gain on the high-energy side of the gain peak is given when the photon energy is equal to the quasi-Fermi level separation at a given level of injection (i.e., $g = 0$ for $\hbar\omega = \Delta F$). We therefore proceed by shifting the measured net modal gain spectra, $\Gamma g - \alpha_{\mathrm{i}}$, upward by the $\alpha_{\mathrm{i}} = 15\ \mathrm{cm}^{-1}$ optical losses in order to obtain the absolute modal gain Γg in the device. Next, we use the transparency points on each of these absolute modal gain spectra to extract the quasi-Fermi level separation ΔF corresponding to each current density J in the experiment. Finally, using the theoretical model we calculate ΔF as a function of the injected carrier density n for the laser structure and, in doing so, determine the value of n which produces the extracted ΔF for each current density J. Following this procedure, we find that the current densities $J = 0.7$, 1.4, 2.0, and 2.4 kA cm^{-2} at which the gain spectra were measured, correspond to quasi-Fermi level separations of 1.375, 1.409, 1.423, and 1.435 eV, which in turn correspond to carrier densities $n = 5.12$, 7.11, 8.24, and 9.38 $\times 10^{18}$ cm^{-3}, respectively, in the theoretical calculations. The TE-polarized component of the SE rate at each of these carrier was then calculated as outlined in Section 10.4.1, from which the theoretical net modal gain spectrum was calculated at each carrier density as $\Gamma g - \alpha_i$, with $\Gamma = 1.60\%$ (calculated) and $\alpha_i = 15$ cm^{-1} (measured), and where g is the TE-polarized material gain spectrum computed using Equation 10.8.

The results of these calculations are shown in Figure 10.11b using, in order of increasing current/carrier density, solid, dashed, dotted, and dash-dotted lines. We note that the theoretical gain spectra shown using solid, dashed and dotted lines include $e1$-$hh1$ optical transitions only while the dash-dotted line, corresponding to the highest current density of 2.4 kA cm^{-2}, also includes $e1$-$lh1$ optical transitions. Including $e1$-$hh1$ transitions only at the highest carrier density was found to underestimate the peak modal gain by ∼10%, highlighting that TE-polarized optical recombination involving LH-like states contributes appreciably to the optical gain at higher levels of injection. This conclusion is consistent with further experimental analysis of the SE spectra [39]. Overall, we see that the theoretical spectra are in good, quantitative agreement with experiment: the calculated magnitude of the net modal gain is in excellent agreement with that measured using the segmented contact method across the full investigated range of current densities (confirming in particular the accuracy of the optical transition matrix elements computed within the framework of the 12-band **k·p** model), and the shape (in photon energy/wavelength) of the experimental gain spectrum is well reproduced at each current/carrier density by the theoretical model. This analysis has been repeated for a series of multisection and Fabry–Pérot laser devices having $x \approx 2\%$, demonstrating that the theoretical model is generally capable of quantitatively predicting the variation of g_{peak} with J for this new class of laser structures [39]. Overall, we note that the theoretical model presented in Section 10.4.1, which is directly underpinned by detailed analysis of the impact of Bi on the band structure (facilitated by atomistic supercell calculations), allows calculation of the device properties and performance that are generally in good, quantitative agreement with the available experimental data. The results presented here therefore confirm the predictive capability of the model, suggesting that it can be used as a reliable tool in the analysis, design and optimization of future devices.

10.5 Emerging Directions in Dilute Bismide Research

Having primarily focused so far in this chapter on describing developments in the growth, characterization, theory and applications of GaBi$_x$As$_{1-x}$ alloys grown on GaAs substrates, in this section, we turn our attention to alternative, emerging directions in research on dilute bismide alloys. Sections 10.5.1 and 10.5.2 outline the basic band structure properties of the quaternary Bi-containing alloys In$_y$Ga$_{1-y}$Bi$_x$As$_{1-x}$ and GaBi$_x$N$_y$As$_{1-x-y}$, grown on InP and GaAs substrates, respectively. By extending the **k·p** model for GaBi$_x$As$_{1-x}$ alloys outlined in Sections 10.3.2 and 10.3.3, we demonstrate how the incorporation of Bi leads in each case to highly flexible alloy band structures and long emission wavelengths beyond those currently accessible using GaAs or InP substrates. We also briefly describe the development of GaBi$_x$As$_{1-x}$/GaN$_y$As$_{1-y}$ type-II QWs grown on GaAs and demonstrate that this novel class of strain-balanced heterostructures have significant promise for the development of photonic devices operating at and beyond 1.55 µm. GaBi$_x$As$_{1-x}$/GaN$_y$As$_{1-y}$ type-II QWs also have the potential to act as a route to extending the emission wavelength beyond that which has been demonstrated to date in GaBi$_x$As$_{1-x}$/GaAs type-I QWs, which is currently limited by the challenges associated with the epitaxial growth of GaBi$_x$As$_{1-x}$ QWs having sufficiently high optical quality and Bi composition to demonstrate electrically pumped lasing at 1.55 µm. In all three cases, we demonstrate the ability to engineer the band gap, spin-orbit-splitting energy, strain, and band offsets over wide ranges, highlighting the potential to use these new material concepts to deliver enhanced capabilities in near- and mid-infrared photonic devices, as well as (at shorter wavelengths) in photovoltaics.

In addition to these three emerging directions in dilute bismide research, we note that ongoing increase of interest in Bi-containing materials has led to initial demonstrations of a number of additional classes of alloys and heterostructures for a range of potential applications. While it is beyond the scope of this chapter to discuss all of these topics in detail, we briefly list prominent examples in Section 10.5.4 and refer the reader to the references provided therein for further information.

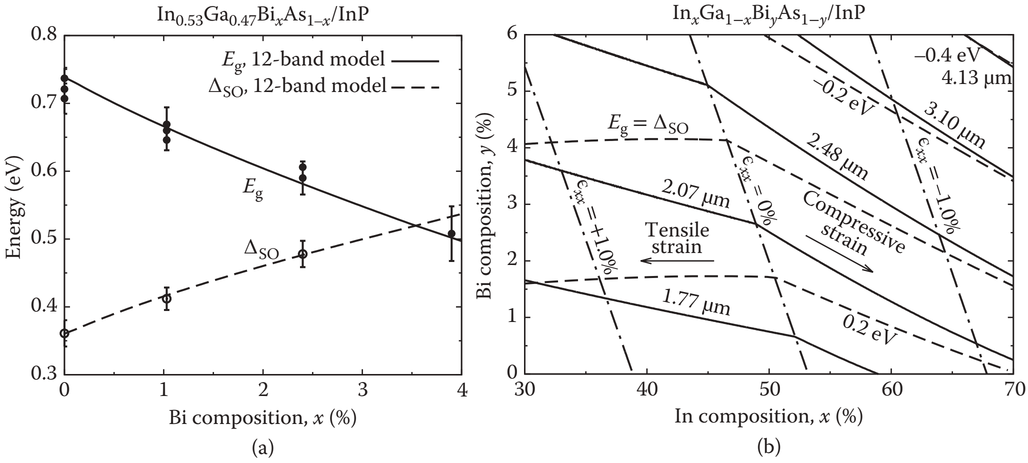
10.5.1 $\mathbf{In}_y\mathbf{Ga}_{1-y}\mathbf{Bi}_x\mathbf{As}_{1-x}$/InP Alloys

The vibrational-rotational spectra of many important gases are characterized by strong optical absorption at mid-infrared wavelengths. As such, there is broad scope for practical applications of semiconductor lasers operating in the 2- to 6-μm wavelength range, including: (1) chemical monitoring in industrial processes, (2) detection of environmental pollutants, (3) remote analysis of hazardous substances (including toxic gases and explosives), and (4) detection of biological markers in medical diagnostics. Despite this, mid-infrared semiconductor laser technology is generally less developed than the GaAs- and InP-based technologies that are employed in optical communications at wavelengths <2 μm. Despite the fact that much effort has been dedicated to developing mid-infrared emitters and detectors over the past two decades [102,103], there remains a need to develop new material and device concepts in order to overcome the limitations associated with current technologies. Existing mid-infrared semiconductor lasers typically (1) incorporate Sb-containing materials grown on nonideal and relatively expensive GaSb or InAs substrates, (2) suffer from difficulties associated with the growth of Sb-containing alloys, and/or (3) incorporate complicated heterostructures, such as in quantum (intraband) or interband cascade devices, where the laser structure must be formed of many layer repeats which require a high degree of design optimization and growth control.

In order to overcome these issues, as well as to improve device performance, there has been increasing effort directed toward developing Sb-free devices grown on InP substrates. From a practical perspective, it is desirable to develop InP-based mid-infrared devices, for several reasons. First, materials growth and processing is significantly more advanced for the InP platform than for GaSb or InAs, since the former has experienced heavy technological and commercial development for applications optical in communications. As such, the InP platform offers well-developed, reproducible, and cost-effective device fabrication. Second, from a material physics perspective, InP has several advantages over GaSb, including higher thermal conductivity and lower electrical resistance. Third, growth of InP-based materials stands to benefit from the ready availability of advanced optical components such as low-loss waveguides.

$\mathrm{In}_y\mathrm{Ga}_{1-y}\mathrm{Bi}_x\mathrm{As}_{1-x}$/InP alloys offer a promising alternative to existing approaches in this challenging wavelength range, and provide the possibility to develop interband diode lasers incorporating type-I QWs which (1) significantly extend the wavelength range accessible using InP substrates, (2) fully exploit the well-established aspects of the InP material platform [45], and (3) use the Bi-induced modifications to the band structure to suppress the Auger and IVBA loss mechanisms. Full details of our theoretical analysis of the $\mathrm{In}_y\mathrm{Ga}_{1-y}\mathrm{Bi}_x\mathrm{As}_{1-x}$ band structure and its suitability for applications in mid-infrared light-emitting devices, can be found in Refs. [13] and [42]. Figure 10.12a shows the variation of the band gap E_g and spin-orbit-splitting energy Δ_SO with Bi composition x in $\mathrm{In}_{0.53}\mathrm{Ga}_{0.47}\mathrm{Bi}_x\mathrm{As}_{1-x}$ alloys grown pseudomorphically on InP. Solid (dashed) lines show the variation of E_g (Δ_SO) calculated using the 12-band $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian described in Section 10.3, while closed (open) circles show the values of E_g (Δ_SO) measured using a combination of optical absorption, PL, and photomodulated reflectance spectroscopy [40,42]. We note that (1) the calculated variation of E_g and Δ_SO with x is in quantitative agreement with experiment, and (2) the theoretical and experimental data indicate that a band structure in which $\Delta_\mathrm{SO} > E_\mathrm{g}$ can be obtained at low Bi compositions $x > 3.5\%$ in compressively strained $\mathrm{In}_{0.53}\mathrm{Ga}_{0.47}\mathrm{Bi}_x\mathrm{As}_{1-x}$/InP alloys, which is significantly lower than the ~10% required on GaAs (cf. Figure 10.3aa).

In order to evaluate the suitability of $\mathrm{In}_y\mathrm{Ga}_{1-y}\mathrm{Bi}_x\mathrm{As}_{1-x}$/InP alloys for applications in mid-infrared devices, we have used the 12-band $\mathbf{k}\cdot\mathbf{p}$ model to investigate general trends in the band structure as a function of the Bi and In compositions x and y [44,45]. Figure 10.12b shows a composition space map of the



(a) (b)

FIGURE 10.12 (a) Variation of the band gap (E_g) and spin-orbit-splitting energy (Δ_{SO}) as a function of Bi composition x, measured for In$_{0.53}$Ga$_{0.47}$Bi$_x$As$_{1-x}$ alloys grown pseudomorphically on InP (closed and open circles, respectively), compared to calculations undertaken using the 12-band **k·p** Hamiltonian of Section 10.3 (solid and dashed lines, respectively). (Adapted from Marko, I.P. et al., *Appl. Phys. Lett.*, 101, 221108, 2012; C.A. Broderick et al., *Phys. Stat. Sol. B*, 250, 773, 2013.) (b) Calculated variation of E_g, and of the difference between the band gap and spin–orbit-splitting energy, $E_g - \Delta_{SO}$, as a function of Bi and In composition (x and y) for In$_y$Ga$_{1-y}$Bi$_x$As$_{1-x}$/InP alloys (Adapted from Broderick, C.A. et al., In *Proceedings of the 16th International Conference on Numerical Simulation of Optoelectronic Devices*, p. 47 (2016); C.A. Broderick et al., *Theory and design of In$_y$Ga$_{1-y}$As$_{1-x}$Bi$_x$ mid-infrared semiconductor lasers: type-I quantum wells for emission beyond 3* μm *on InP substrates*, submitted 2017.) Solid and dashed lines denote, respectively, paths in the composition space along which E_g and $E_g - \Delta_{SO}$ are constant. Dash-dotted lines denote paths in the composition space along which strain is constant. We see that In$_y$Ga$_{1-y}$Bi$_x$As$_{1-x}$ alloys can be grown either lattice-matched ($\epsilon_{xx} = 0$), or under compressive ($\epsilon_{xx} < 0$) or tensile ($\epsilon_{xx} > 0$) strain, and that compressively strained alloys enable emission at wavelengths >3 μm at relatively low Bi compositions. Alloys with compositions lying above the $E_g = \Delta_{SO}$ contour are alloys in which $\Delta_{SO} > E_g$, and in which suppression of the nonradiative Auger recombination and IVBA processes involving the SO band can hence be expected.

band gap (E_g), and of the difference between the band gap and spin-orbit-splitting energy ($E_g - \Delta_{SO}$), in pseudomorphically strained In$_y$Ga$_{1-y}$Bi$_x$As$_{1-x}$/InP, calculated using the 12-band model. The dash-dotted lines denote paths in the composition space along which the in-plane strain ϵ_{xx} is constant when the In$_y$Ga$_{1-y}$Bi$_x$As$_{1-x}$ alloy is grown pseudomorphically on an InP substrate and show that the alloy can be grown either lattice-matched ($\epsilon_{xx} = 0$) or under compressive ($\epsilon_{xx} < 0$) or tensile ($\epsilon_{xx} > 0$) strain. Solid lines denote alloy compositions for which E_g is constant. Examining the calculated variation of ϵ_{xx} and E_g with alloy composition, we observe that the band gap of In$_y$Ga$_{1-y}$Bi$_x$As$_{1-x}$ alloys can be varied over an extremely wide range, from ~1.55 to >4 μm, and that this broad spectral coverage can be obtained at modest strains $|\epsilon_{xx}| < 1.5\%$.

The dashed lines in Figure 10.12b denote alloy compositions for which $E_g - \Delta_{SO}$ is constant. Following the $\epsilon_{xx} = 0$ line, we calculate that (1) the amount of Bi required to bring about the $\Delta_{SO} > E_g$ band structure condition (all alloys lying above the $E_g = \Delta_{SO}$ contour) is close to 4% (a prediction in quantitative agreement with the results of a range of experimental measurements [40,43,54]), and (2) the amount of Bi required to achieve $\Delta_{SO} > E_g$ remains constant at approximately 4% in tensile strained alloys less In while, by contrast, it is significantly reduced in compressively strained alloys having higher In compositions. Compared to GaAs, In$_y$Ga$_{1-y}$As/InP alloys have reduced band gaps and comparable spin-orbit-splitting energies, so that significantly less Bi is required to achieve the $\Delta_{SO} > E_g$ band structure condition in In$_y$Ga$_{1-y}$Bi$_x$As$_{1-x}$/InP alloys than in the GaBi$_x$As$_{1-x}$/GaAs materials considered previously [13,40,42,43].

Therefore, in addition to providing the possibility of suppressing the CHSH Auger recombination and IVBA loss mechanisms in InP-based mid-infrared semiconductor lasers, the $In_yGa_{1-y}Bi_xAs_{1-x}$/InP material system also has the potential to provide an important proof of principle regarding the use of dilute bismide alloys to suppress the dominant loss mechanisms in GaAs-based semiconductor lasers operating at telecommunication wavelengths.

The difference in band structure between the compressive and tensile strained alloys shown in Figure 10.12b is readily understood: Bi, being larger than the As atoms it replaces, introduces compressive strain so that a significant fraction of In needs to be removed in order to obtain a tensile strained quaternary alloy containing Bi. This removal of In leads to a tensile strained InGaAs host matrix with an increased band gap, as well as a VB structure in which the VBAC coupling is weakened compared to the unstrained or compressively strained cases [13,42]. When these effects are combined, we find that significantly larger Bi compositions are required in tensile strained $In_yGa_{1-y}Bi_xAs_{1-x}$ in order to bring about a given band gap reduction for a fixed magnitude of the strain, as compared to the compressively strained case in which both In and Bi simultaneously contribute to the band gap reduction and compressive strain. Bearing this in mind, and examining Figure 10.12b from the perspective of laser design, we note that the composition region of interest are those lying above the $E_g = \Delta_{SO}$ contour, since it is expected that the Auger and IVBA processes involving the SO band will be suppressed in these alloys, while they are close to resonant for $2 < \lambda < 2.5$ μm, at which compositions $E_g \approx \Delta_{SO}$. Within the reduced composition range of alloys having $\Delta_{SO} > E_g$ we note that the alloy band gap covers a large range of wavelengths >2.5 μm and, significantly, that the 2.5- to >4-μm spectral range can be accessed in alloys having modest compressive strains and Bi compositions.

Our theoretical analysis of $In_yGa_{1-y}Bi_xAs_{1-x}$/InP laser structures [45] suggests that there is broad scope to use this material system to realize compressively strained type-I QWs having emission wavelengths in the 2.5- to 5 μm range. Furthermore, our analysis indicates that the physical characteristics of $In_yGa_{1-y}Bi_xAs_{1-x}$/InP QWs are intrinsically superior to existing GaSb-based heterostructures in the same wavelength range. First, as outlined above, theoretical calculations and experimental measurements have confirmed that $\Delta_{SO} > E_g$ can be achieved in $In_{0.53}Ga_{0.47}Bi_xAs_{1-x}$/InP alloys for Bi compositions as low as $x \approx 4\%$ [13,40,42,43]. This means that it should be possible to suppress the CHSH Auger recombination and IVBA processes and bring about highly efficient, temperature-stable operation in mid-infrared semiconductor lasers. Second, our calculations indicate that large type-I band offsets can readily be engineered in $In_yGa_{1-y}Bi_xAs_{1-x}$ QWs having unstrained ternary $In_{0.53}Ga_{0.47}As$ barriers. This has two significant benefits: (1) the presence of large CB and VB offsets will mitigate carrier leakage, thereby overcoming a key factor limiting the performance of GaSb-based devices at high temperature, and (2) the ability to grow lattice-matched, ternary InGaAs barriers, without the need to incorporate Al or P to enhance the carrier confinement, promises to simplify the growth of these laser structures. We note that (3) also means that it is not necessary to introduce a trade-off between the carrier and optical confinement in these structures (cf. Section 10.4.3). Furthermore, these ternary barrier materials are the same unstrained InGaAs buffer layers upon which $In_yGa_{1-y}Bi_xAs_{1-x}$ epilayers have typically been grown. Overall, this means that $In_yGa_{1-y}Bi_xAs_{1-x}$ QWs can be used to develop mid-infrared lasers in which the remainder of the laser structure is essentially identical to those commonly employed in the well-established 1.3- and 1.55-μm InP-based devices. As such, $In_yGa_{1-y}Bi_xAs_{1-x}$/InP alloys offer the possibility not only to provide type-I InP-based QW diode lasers operating at wavelengths >3 μm, but also promise to deliver several key advantages over competing mid-infrared device concepts.

While there have been some investigations of $In_yGa_{1-y}Bi_xAs_{1-x}$ alloys grown on InP substrates, this material system has to date been the subject of much less research attention than the GaAs-based $GaBi_xAs_{1-x}$ alloys discussed above. Initial growth of $In_yGa_{1-y}Bi_xAs_{1-x}$/InP materials was reported by Feng et al. [104,105], and has since been established by several other groups [106–109]. Enhanced understanding of the incorporation of Bi in InGaAs has led to growth of $In_yGa_{1-y}Bi_xAs_{1-x}$ alloys with Bi compositions as high as 7% [108], and recent experimental and theoretical analysis of a sample containing 5.8% Bi has verified the presence of a $\Delta_{SO} > E_g$ band structure up to and above room temperature [43]. Growth of

$In_yGa_{1-y}Bi_xAs_{1-x}$ alloys has, however, to date been limited to bulk-like epitaxial layers and so, in order to develop this material system for applications in mid-infrared semiconductor lasers, future studies will need to focus on (1) the growth of thin, strained QWs having high optical quality, and (2) providing evidence of reduced CHSH Auger recombination and IVBA in materials and heterostructures having $\Delta_{SO} > E_g$. While research on $In_yGa_{1-y}Bi_xAs_{1-x}$/InP alloys still faces challenges related to the establishment of growth and fabrication of high-quality materials and devices, recent successes in the development of $GaBi_xAs_{1-x}$ alloys, including the demonstration of electrically pumped lasers [30,35,96], suggest that these issues can be overcome with dedicated research effort. We therefore conclude that $In_yGa_{1-y}Bi_xAs_{1-x}$ alloys are a promising candidate for the development of highly efficient, temperature-stable semiconductor lasers, grown on InP substrates and operating in the 3- to 5-μm wavelength range.

10.5.2 $GaBi_xN_yAs_{1-x-y}$/GaAs Alloys

An alternative quaternary Bi-containing alloy which has begun to attract research interest is $GaBi_xN_yAs_{1-x-y}$. Initial theoretical analysis has demonstrated that the band structure of $GaBi_xN_yAs_{1-x-y}$ alloys grown on GaAs substrates is extremely flexible, providing significant potential to develop materials suited to a range of applications in the near- and mid-infrared [12,41,54]. First, since Bi is larger than the As atoms it replaces in GaAs to form $GaBi_xAs_{1-x}$, it introduces compressive strain when grown on a GaAs substrate. Likewise, N atoms, which are smaller than As, introduce tensile strain. As a result, independent control of the Bi and N compositions in $GaBi_xN_yAs_{1-x-y}$ allows precise strain engineering with respect to a GaAs substrate. Second, Bi or N incorporation in GaAs is known to cause a rapid reduction of the band gap with increasing composition. Theoretical calculations [54] suggest that coalloying Bi and N in GaAs to form $GaBi_xN_yAs_{1-x-y}$ enhances these effects and leads to an extremely large reduction in the band gap with increasing Bi and N composition, which is larger than that observed in either $GaBi_xAs_{1-x}$ or GaN_xAs_{1-x} alloys. As a result, the band gaps of $GaBi_xN_yAs_{1-x-y}$ alloys are capable of covering a large spectral range in the near- and mid-infrared, and come with the added benefit of being able to independently control the strain and band gap. These properties make $GaBi_xN_yAs_{1-x-y}$ alloys of particular interest for applications in multi-junction solar cells [21], since they offer the potential to provide optical absorption across an extremely wide range of wavelengths while remaining lattice-matched to either a GaAs or Ge substrate. In addition to the ability to readily control the strain in these alloys, the impact of Bi on the VB structure are retained, meaning that it should also be possible to achieve a band structure in which $\Delta_{SO} > E_g$ in order to suppress Auger and IVBA processes involving the SO band.

Detailed atomistic calculations suggest that the effects of Bi and N on the band structure are effectively decoupled [54], with the modified CB (VB) structure describable primarily in terms of the well-known effects of N-(Bi-)related localized resonant states. On this basis, Broderick et al. derived a 14-band $\mathbf{k} \cdot \mathbf{p}$ Hamiltonian to describe the band structure of $GaBi_xN_yAs_{1-x-y}$ alloys [54]. In this model the effects of N and Bi incorporation on the CB and VB structure are described independently in terms of band-anticrossing interactions between N-(Bi-)related localized states and the extended states of the GaAs host matrix CB (VB) edge (cf. Section 10.3.1), with additional virtual crystal and strain-induced modifications to the CB, VB, and SO band edge energies (cf. Section 10.3.2). The fact that N (Bi) primarily affects the CB (VB) then means that the $GaBi_xN_yAs_{1-x-y}$/GaAs band offsets are effectively independently controllable, so that there is significant potential to engineer the properties of QWs and related heterostructures. Incorporating N brings about a large reduction in the CB edge energy, resulting in a large CB offset and a small VB offset with respect to GaAs. The converse is true of Bi incorporation, which leads to a small CB offset and a large VB offset. Combining these effects means that the CB and VB offsets can be engineered relatively independently of one another in $GaBi_xN_yAs_{1-x-y}$, opening up significant potential to develop $GaBi_xN_yAs_{1-x-y}$ QWs providing tunable long-wavelength emission and absorption across a broad spectral range. Full details of our theoretical analysis of the $GaBi_xN_yAs_{1-x-y}$ band structure can be found in Refs. [12], [41] and [54].

Figure 10.13 shows a composition space map of the band gap (E_g), and of the difference between the band gap and spin–orbit-splitting energy ($E_g - \Delta_{SO}$), in pseudomorphically strained GaBi$_x$N$_y$As$_{1-x-y}$/GaAs, calculated using the 14-band **k·p** Hamiltonian described in Ref. [54]. The dashed lines denote paths in the composition space along which the in-plane strain is constant when the GaBi$_x$N$_y$As$_{1-x-y}$ alloy is grown pseudomorphically on a GaAs substrate, and show that the alloy can be grown either lattice-matched ($\epsilon_{xx} = 0$), or under compressive ($\epsilon_{xx} < 0$) or tensile ($\epsilon_{xx} > 0$) strain. Solid lines denote alloy compositions for which E_g is constant. Examining the calculated variation of ϵ_{xx} and E_g with alloy composition we observe that the band gap of GaBi$_x$N$_y$As$_{1-x-y}$ alloys can be varied over an extremely wide range, from that of GaAs (~850 nm) through the near-infrared, and out to wavelengths >3 μm in the mid-infrared. We also note that this broad spectral coverage can be obtained in alloys that are lattice-matched to GaAs, or in alloys having modest tensile or compressive strains with $|\epsilon_{xx}| < 1\%$. The dashed lines in Figure 10.13 denote alloy compositions for which $E_g - \Delta_{SO}$ is constant. The fact that N incorporation brings about a large band gap reduction while leaving the VB structure almost unaltered (apart from minor strain-induced changes) means that co-alloying N and Bi has the potential to bring about an $\Delta_{SO} > E_g$ band structure at lower Bi compositions than in the N-free GaBi$_x$As$_{1-x}$ alloy. Following the $\epsilon_{xx} = 0$ line we calculate that the amount of Bi required to bring about the $\Delta_{SO} > E_g$ band structure condition (all alloys lying to the right of the $E_g = \Delta_{SO}$ contour) is close to 7% in the lattice-matched case, which is approximately 30% less than that required in GaBi$_x$As$_{1-x}$. In addition to the aforementioned potential of this material system for applications in multijunction solar cells, we also note that the wide composition range for which $\Delta_{SO} > E_g$ indicates that there is also potential to grow GaBi$_x$N$_y$As$_{1-x-y}$ QWs for >2-μm emission, bringing about the possibility of developing GaAs-based mid-infrared diode lasers with suppressed CHSH Auger recombination and IVBA. This also provides an interesting opportunity to develop monolithic GaAs-based VCSELs and related devices operating in the mid-infrared.

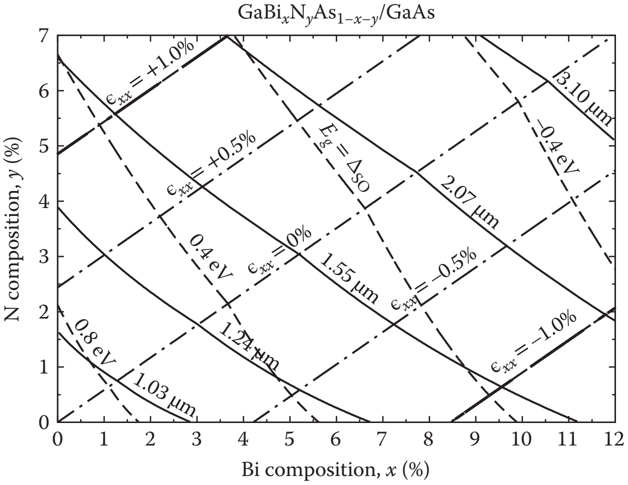


FIGURE 10.13 Calculated variation of the band gap (E_g), and of the difference between the band gap and spin-orbit-splitting energy ($E_g - \Delta_{SO}$), as a function of Bi and N compositions (x and y) for GaBi$_x$N$_y$As$_{1-x-y}$ alloys grown pseudomorphically on GaAs compositions. (Adapted from Broderick, C.A. et al., *Semicond. Sci. Technol.*, 27, 094011, 2012.) Solid and dashed lines denote, respectively, paths in the composition space along which E_g and $E_g - \Delta_{SO}$ are constant. Dash-dotted lines denote paths in the composition space along which strain is constant. We see that GaBi$_x$N$_y$As$_{1-x-y}$ alloys can be grown either lattice-matched ($\epsilon_{xx} = 0$), or under compressive ($\epsilon_{xx} < 0$) or tensile ($\epsilon_{xx} > 0$) strain, and that the band gap can be varied over an extremely wide spectral range in the near- and mid-infrared. Alloys with compositions lying to the right of the $E_g = \Delta_{SO}$ contour are alloys in which $\Delta_{SO} > E_g$, and in which suppression of the non-radiative Auger recombination and IVBA processes involving the SO band can hence be expected.

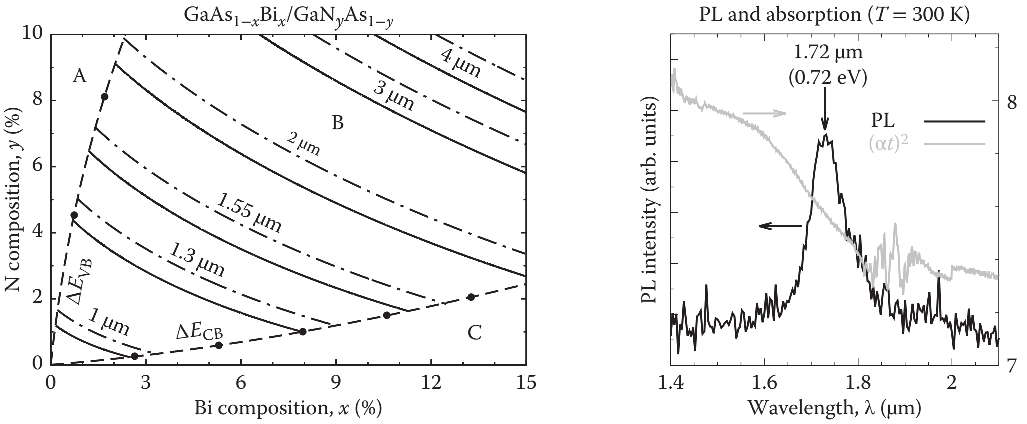
While there is increasing interest in the development of $GaBi_xN_yAs_{1-x-y}$ alloys for applications in the areas described above, to date there have been limited growth studies of this novel material system. Initial growth of $GaBi_xN_yAs_{1-x-y}$ bulk-like epitaxial layers was undertaken via MBE in the mid-2000s [110,111], while more recently epitaxial layers and QWs have been grown using MOVPE [112]. MOVPE growth has realized co-alloying of Bi and N compositions up to $x = 3.5\%$ and $y = 2.7\%$, but PL has only been observed for samples with compositions up to $x = 1.8\%$ and $y = 1.8\%$. However, the observed PL at these compositions demonstrated an alloy band gap close to 1 eV, confirming that co-alloying of Bi and N can bring about an extreme reduction of the band gap at relatively low compositions [112]. Meanwhile, MBE growth of $GaBi_xN_yAs_{1-x-y}$ alloys has demonstrated electroluminescence at wavelengths out to 1.3 μm [113], and Bi and N incorporations up to $x = 4.5\%$ and $y = 8.0\%$ have been obtained [110]. However, no measurements of the band gap have been performed on these higher composition materials, with the Bi and N compositions having been deduced on the basis of x-ray diffraction measurements. Despite the significant challenges presented by the growth of this highly mismatched quaternary alloy, initial investigations have been promising, and the realization of a 1-eV band gap at such low compositions merits further investigations in order to refine the growth of $GaBi_xN_yAs_{1-x-y}$ alloys to exploit their significant potential for practical applications.

10.5.3 $GaBi_xAs/GaN_yAs_{1-y}$ Type-II Quantum Wells on GaAs

In addition to the direct co-alloying Bi and N in GaAs to form $GaBi_xN_yAs_{1-x-y}$, Bi and N incorporation in GaAs opens up additional possibilities for band structure engineering and the development of novel heterostructures. For example, GaN_xAs_{1-x} ($GaBi_xAs_{1-x}$) can be used to form tensile (compressively) strained type-I QWs it possible to grow strain-balanced type-II QWs and superlattices in which electrons (holes) are confined in the GaN_xAs_{1-x} ($GaBi_xAs_{1-x}$) layer(s) [46,114]. Such structures offer a number of advantages for applications in semiconductor lasers, photodetectors, and solar cells, including the possibility of facilitating optical emission and absorption across a wide range of wavelengths, as well as exploiting a high degree of control over the built-in strain, carrier transport, radiative lifetimes and non-radiative recombination rates [46]. A range of type-II QWs grown on GaAs using different combinations of III-V alloys are currently under investigation for long-wavelength applications. However, strain-balancing of type-II QWs on GaAs has not been possible using the (In)Ga(N)$As_{1-x}Sb_x/In_yGa_{1-y}As$ or $GaAs_{1-x}Bi_x/In_yGa_{1-y}As$ structures that have been investigated recently [115–118], since the constituent alloys are all compressively strained when grown on GaAs. By comparison, $GaAs_{1-x}Bi_x/GaN_yAs_{1-y}$ type-II structures are highly engineerable, can be grown with little or no net strain relative to a GaAs substrate, and significant potential for applications in GaAs-based near- and mid-infrared photonic devices [46] and (at shorter wavelengths) in photovoltaics [114].

The band alignment, band offsets, and emission wavelengths achievable using $GaAs_{1-x}Bi_x/GaN_yAs_{1-y}$ structures on GaAs are summarized in Figure 10.14a. Dashed lines denote Bi and N compositions x and y for which the CB offset ΔE_{CB} and VB offset ΔE_{VB} are equal in $GaAs_{1-x}Bi_x$ and GaN_yAs_{1-y}. Above (below) the lower dashed line ΔE_{CB} is larger (smaller) in GaN_yAs_{1-y} than in $GaAs_{1-x}Bi_x$, while to the right (left) of the upper dashed line ΔE_{VB} is larger (smaller) in $GaAs_{1-x}Bi_x$ than in GaN_yAs_{1-y}. This divides the composition space into three regions. Regions A and C correspond to type-I band alignment, with electrons and holes both confined within either the GaN_yAs_{1-y} (A) or $GaAs_{1-x}Bi_x$ (C) layers. Region B corresponds to type-II band alignment, with holes (electrons) confined in the $GaAs_{1-x}Bi_x$ (GaN_yAs_{1-y}) layers. Closed circles denote increases of 50 meV in the respective band offsets, beginning from zero at $x = y = 0$ (GaAs). Solid (dash-dotted) lines in region B denote alloy compositions for which the band gap in a bulk (QW) type-II structure—between the GaN_yAs_{1-y} CB and $GaAs_{1-x}Bi_x$ VB—is constant. These calculations demonstrate that type-II $GaAs_{1-x}Bi_x/GaN_yAs_{1-y}$ QWs grown on GaAs have the potential to cover an extremely broad spectral range, through the near-infrared to mid-infrared wavelengths in excess of 3 μm [46].

Broderick et al. [46] recently presented the first demonstration of this new class of III-V heterostructures. The black and gray lines in Figure 10.14b show, respectively, the room temperature PL and optical absorption measured for a prototypical type-II GaBi$_x$As$_{1-x}$/GaN$_y$As$_{1-y}$ structure containing five QWs, having Bi and N compositions $x = 3.3\%$ and $y = 5.2\%$, and with GaBi$_x$As$_{1-x}$ and GaN$_y$As$_{1-y}$ layer thicknesses of 10.5 and 9.2 nm [46]. This structure was grown using MOVPE and characterized using high-resolution x-ray diffraction measurements, which indicated high structural quality of the strained type-II QWs. Examining Figures 10.14a and 10.14b we note that, despite the intrinsically low optical efficiency of the type-II structure (resulting from the low electron–hole overlap under optical excitation), the measured PL peak at 0.72 eV agrees well with the *e*1-*hh*1 transition energy of 0.74 eV calculated using a carefully parameterized 14-band **k·p** Hamiltonian [54] in conjunction with the plane wave expansion method described in Section 10.4.1 [46]. We note that this band gap corresponds to an emission wavelength ~1.7μm, which demonstrates that GaBi$_x$As$_{1-x}$/GaN$_y$As$_{1-y}$ type-II QWs grown on GaAs can readily provide emission at and beyond 1.55 μm with significantly reduced Bi compositions compared to those required in type-I GaBi$_x$As$_{1-x}$/(Al)GaAs structures (cf. Section 10.4). Indeed, it is promising to observe clear room temperature PL and optical absorption in initial growth of a prototype structure, particularly given the significant scope to optimize the layer ordering and thicknesses to enhance the optical efficiency [46]. Theoretical investigations are ongoing to quantify the electronic and optical properties of this new class of GaAs-based heterostructures, and refinement of epitaxial growth is expected to lead to the development of structures with improved optical efficiency for potential applications in QW solar cells, as well as in semiconductor lasers and photodetectors.



(a) (b)

FIGURE 10.14 (a) Composition space map showing regions of type-I (A and C) and type-II (B) band alignment, as well as the emission wavelengths accessible using GaAs$_{1-x}$Bi$_x$/GaN$_y$As$_{1-y}$ type-II QWs grown on GaAs. Solid (dash-dotted) lines denote paths in the composition space along which the type-II band gap in bulk (QW) GaAs$_{1-x}$Bi$_x$/GaN$_y$As$_{1-y}$ is constant. Dashed lines denote paths in the composition space along which the CB and VB offsets ΔE_{CB} and ΔE_{VB} are equal in GaAs$_{1-x}$Bi$_x$ and GaN$_y$As$_{1-y}$. Closed circles denote 50 meV increases in ΔE_{CB} and ΔE_{VB} starting from zero at $x = y = 0$. (Adapted from Broderick, C.A. et al., GaAs$_{1-x}$Bi$_x$/GaN$_y$As$_{1-y}$ type-II quantum wells: Novel strain-balanced heterostructures for GaAs-based near- and mid-infrared photonics, *Sci. Rep.*, 7, 46371, 2017) (b) Room temperature photoluminescence (PL) spectrum (black line) and squared product $(\alpha t)^2$ of the optical absorption α and propagation length *t* (from transmission measurements; gray line) measured for a prototypical GaAs$_{0.967}$Bi$_{0.033}$/GaN$_{0.062}$As$_{0.938}$ type-II QW structure, grown by MOVPE. (Adapted from C.A. Broderick et al., Type-II strain-balanced superlattices on GaAs: novel heterostructure for photonics and photovoltaics, in *Proceedings of the 17th International Conference on Numerical Simulation of Optoelectronic Devices*, 2017)

10.5.4 Further Directions: Quaternary Alloys, Type-II Structures, and Narrow-Gap Materials

Here, we briefly outline additional emerging directions in resarch on dilute bismide alloys by describing some recent work on materials and heterostructures distinct from, but related to, those discussed above. We refer the reader to the reference provided for further information.

Recently, the quaternary dilute bismide alloy $\text{GaBi}_x\text{As}_{1-x-y}\text{P}_y$ has been proposed as an alternative material system to $\text{GaBi}_x\text{N}_y\text{As}_{1-x-y}$ that can provide a 1 eV band gap while remaining lattice matched to a GaAs substrate (cf. Section 10.5.2). In contrast to co-alloying Bi with N to form $\text{GaBi}_x\text{N}_y\text{As}_{1-x-y}$, incorporation of phosphorus (P), which also compensates the compressive strain brought about by Bi incorporation, tends to increase the band gap at fixed Bi composition. This means that larger Bi compositions will be required to achieve lattice-matched alloys having a given band gap compared to those required in $\text{GaBi}_x\text{N}_y\text{As}_{1-x-y}$ alloys. However, initial investigations of $\text{GaBi}_x\text{P}_y\text{As}_{1-x-y}$ have demonstrated that incorporation of a small amount of P significantly enhances Bi incorporation during MOVPE growth [119]. Although it is likely to be related to the P-induced reduction of compressive strain in the quaternary alloy, the precise mechanism by which co-alloying P and Bi enhances Bi incorporation has yet to be definitively determined. Related structures have been grown which include compressively strained $\text{GaBi}_x\text{As}_{1-x}$ QWs having tensile strained $\text{GaAs}_{1-y}\text{P}_y$ barriers, with the first results for this new class of GaAs-based strain-compensated laser structures having been reported recently in Ref. [120]. Investigations of quatnerary dilute bismide alloys containing P (or N) are at a relatively early stage, but initial results have been promising and investigations of these novel semiconductor alloys are ongoing.

Further examples of recently emerging directions in dilute bismide research are the growth of type-II and metamorphic heterostructures. In Section 10.5.3 we described that $\text{GaBi}_x\text{As}_{1-x}/\text{GaN}_y\text{As}_{1-y}$ type-II QWs can be used to achieve long emission wavelengths on GaAs substrates. Recently, GaAs-based $\text{In}_y\text{Ga}_{1-y}\text{As}/\text{GaBi}_x\text{As}_{1-x}$ type-II QWs have been suggested as an alternative approach to achieving this goal. These heterostructures are similar in principle to the $\text{GaBi}_x\text{As}_{1-x}/\text{GaN}_y\text{As}_{1-y}$ type-II QWs discussed above, but differ in two important aspects. First, the electron-confining layer(s) of the structure are formed of the conventional alloy $\text{In}_y\text{Ga}_{1-y}\text{As}$. The ease and reproducibility with which high-quality $\text{In}_y\text{Ga}_{1-y}\text{As}$ strained layers can be grown on GaAs is likely to be of benefit for the realization of electrically pumped lasers and related devices. Second, while the electron-confining $\text{GaN}_y\text{As}_{1-y}$ layer(s) in a type-II $\text{GaBi}_x\text{As}_{1-x}/\text{GaN}_y\text{As}_{1-y}$ structure are tensile strained, the $\text{In}_y\text{Ga}_{1-y}\text{As}$ layer(s) in a type-II $\text{In}_y\text{Ga}_{1-y}\text{As}/\text{GaBi}_x\text{As}_{1-x}$ structure are compressively strained. This means that it is not possible to grow strain-balanced $\text{In}_y\text{Ga}_{1-y}\text{As}/\text{GaBi}_x\text{As}_{1-x}$ type-II structures on GaAs, which limits the ability to grow large numbers of QWs due to the accumulation of compressive strain in the device active region. This may limit the ability to grow device structure with sufficiently high optical efficiency. MBE growth of $\text{In}_y\text{Ga}_{1-y}\text{As}/\text{GaBi}_x\text{As}_{1-x}$ type-II structures has recently been established by Pan et al. [115], who reported low temperature PL at a wavelength close to 1.1 μm, thereby providing an initial confirmation of the potential of these structure to overcome the limitations associated with the growth of $\text{GaBi}_x\text{As}_{1-x}/\text{GaAs}$ type-I QWs having high Bi compositions ($\sim 10\%$).

The first metamorphic heterostructures based on dilute bismide alloys were presented recently by Gu et al. [121], who demonstrated MBE growth of InP-based type-I structures incorporating $\text{Al}_y\text{In}_{1-y}\text{As}$ metamorphic buffer layers and $\text{InBi}_x\text{As}_{1-x}$ QWs. The aim of these structures is the same as that outlined for $\text{In}_y\text{Ga}_{1-y}\text{Bi}_x\text{As}_{1-x}/\text{InP}$ in Section 10.5.1: to extend the emission wavelength in InP-based QWs beyond 3 μm. Current approaches to obtaining emission at wavelengths >3 μm on InP include InAs metamorphic QWs grown on $\text{Al}_y\text{In}_{1-y}\text{As}$ metamorphic buffer layers. It is hoped that the strong reduction of the $\text{InBi}_x\text{As}_{1-x}$ band gap with increasing Bi composition [54] should allow the wavelength to be extended significantly beyond that obtainable using existing devices [19]. The initial work of Gu et al. demonstrated low temperature PL at wavelengths beyond 3.1 μm, confirming the validity of this approach. However, further effort, is required to quantify and exploit the potential of this class of dilute bismide heterostructure for applications in mid-infrared photonic devices.

Finally, we note that there has also been a steady growth of interest in narrow-gap dilute bismide alloys such as (In)GaBi$_x$Sb$_{1-x}$. These alloys, typically grown on GaSb or InAs substrates, offer the potential to use Bi to reduce the already narrow (In)GaSb band gap, and hence cover an extremely broad range of wavelengths in the mid-infrared and beyond. As such, while investigations of Sb-containing dilute bismide alloys are at a relatively early stage, (In)GaBi$_x$Sb$_{1-x}$ alloys are attracting interest for the development of type-I and -II QW lasers and photodetectors throughout the full mid-infrared spectral range [17,122]. Furthermore, due to their intrinsically large spin-orbit coupling, in addition to the potential of using Bi incorporation to introduce large, tunable Rashba spin-splittings, interest in narrow-gap dilute bismide alloys for applications in spintronics is likely to attract significant attention in the coming years [25].

10.6 Summary and Conclusions

In this chapter we have provided an introduction to the theory and simulation of dilute bismide materials and devices, which we have framed within the context of recent and ongoing research on GaBi$_x$As$_{1-x}$ and related Bi-containing alloys. Beginning with a general overview of the electronic properties of dilute bismide alloys, we highlighted the potential to exploit the Bi-induced modifications of the band structure of conventional III–V semiconductors to develop photonic, photovoltaic, and spintronic devices with enhanced performance and capabilities. In particular, we discussed that the incorporation of Bi in (In)GaAs leads to a significant reduction in the material band gap (E_g), which is accompanied by a strong enhancement of the spin-orbit-splitting energy (Δ_{SO}), both of which are characterized by strong composition–dependent bowing. We demonstrated that it is possible, through Bi incorporation, to engineer a band structure in which $\Delta_{SO} > E_g$ at telecommunication wavelengths, which is expected to lead to suppression of the dominant nonradiative Auger recombination and IVBA processes (involving the SO band) that limit the high-temperature performance of existing InP-based 1.55-μm semiconductor lasers. As such, we concluded that GaBi$_x$As$_{1-x}$ alloys have the potential to deliver the long sought-after goal of realizing uncooled operation of highly efficient and temperature-stable telecom lasers, which promises to deliver significant energy savings in next-generation optical communications networks. Growth of long-wavelength devices on GaAs also presents the opportunity to exploit vertical-cavity architectures, meaning that dilute bismide alloys have significant potential to extend the benefits of the GaAs platform to telecommunications and mid-infrared wavelengths and, hence, facilitate the development of advanced photonics technologies, with enhanced performance and new capabilities.

Taking into account the fact that Bi incorporated into (In)GaAs acts as an isovalent impurity, in Section 10.2 we highlighted the need for detailed atomistic models to provide a quantitative understanding of the electronic properties. On this basis, we provided a review of the atomistic methods which have been developed for, and applied to, the study of GaBi$_x$As$_{1-x}$ and related alloys. Having identified the theoretical methods and models that are capable of describing the impact of Bi on the electronic structure, we then compared the results of atomistic TB calculations for realistic, disordered GaBi$_x$As$_{1-x}$ alloys to experiment. We showed that a TB approach is capable of describing, in a quantitative manner, the strong Bi-induced reduction (increase) in the band gap (spin-orbit-splitting energy), across the full composition range for which experimental data are available. We further demonstrated that a TB approach quantitatively predicts the strong Bi-induced increase in the magnitude of the electron effective g factor in pseudomorphically strained GaBi$_x$As$_{1-x}$/GaAs alloys, thereby providing detailed insight into roles played by strain and short-range alloy disorder in determining the details of the strongly perturbed VB structure and confirming the predictive capability of the theory.

In Section 10.3 we turned our attention to deriving a continuum $\mathbf{k}\cdot\mathbf{p}$ model which is suited to modeling the electronic and optical properties of dilute bismide heterostructures. Using an atomistic TB model, we reached two significant conclusions. First, we reviewed that the main features of the band structure of (In)GaBi$_x$As$_{1-x}$ alloys are well described in terms of a phenomenological Bi composition–dependent VBAC interaction between the extended VB edge states of the host matrix semiconductor and localized

Bi-related impurity states which are resonant with the host matrix VB. Second, we demonstrated how the TB method can be used to directly determine the energies of the Bi-related impurity states, as well as the coupling strength between the Bi-related impurity states and host matrix VB edge states, thereby removing the parametric ambiguity associated with band-anticrossing models parameterized solely with reference to experimental measurements. On this basis, we reviewed the derivation of a 12-band **k·p** Hamiltonian for (In)$GaBi_xAs_{1-x}$ directly from a series of detailed atomistic supercell calculations, provided a consistent set of parameters for performing calculations on real alloys, and used the **k·p** model to explicitly derive expressions for the band edge energies and offsets in pseudomorphically strained (In)$GaBi_xAs_{1-x}$ alloys and QWs, respectively.

Next, in Section 10.4, we discussed the theory and modeling of dilute bismide QW lasers grown on GaAs substrates. We applied the 12-band **k·p** model to the study of $GaBi_xAs_{1-x}$/(Al)GaAs QW heterostructures and quantified the impact of Bi incorporation on the (1) VB structure and DOS, (2) optical gain, and (3) threshold characteristics of realistic device structures. By considering the incorporation of Al in the barrier layers to improve electron confinement in the QW, we demonstrated the presence of a trade-off between the carrier and optical confinement which allows the design of $GaBi_xAs_{1-x}$/(Al)GaAs QWs at low Bi compositions $x < 6\%$ to be engineered in order to minimize the threshold current density, a prediction which has been verified experimentally. Repeating this analysis as a function of the QW Bi composition x, we demonstrated that the need to incorporate Al in the barrier layers is removed for $x > 6\%$, which should lead to improved device performance as the Bi composition in GaAs-based laser structures is increased. We extended this analysis to consider the gain characteristics of a $GaBi_xAs_{1-x}$ laser structure designed to emit at 1.55 μm and showed that QWs having higher Bi compositions stand to benefit not only from suppressed Auger recombination and IVBA, but also from the effects of Bi-induced compressive strain, with the latter leading to a favorable VB structure that should, in an ideal device, deliver low threshold current density and high differential gain. Having confirmed the potential of $GaBi_xAs_{1-x}$ QWs for the development of highly efficient and temperature-stable 1.55-μm GaAs-based semiconductor lasers, we then turned our attention to a detailed comparison between theory and experiment for the first generation of electrically pumped $GaBi_xAs_{1-x}$ laser devices. We demonstrated that the theoretical approach outlined in Sections 10.3 and 10.4 is capable of quantitatively describing the SE and optical gain of this new class of GaAs-based semiconductor lasers, confirming its predictive capability and highlighting its potential for use in the design and optimization of future $GaBi_xAs_{1-x}$ lasers and related devices.

Finally, in Section 10.5, we provided an overview of emerging directions in dilute bismide research, including quaternary Bi-containing alloys grown on InP and GaAs substrates, type-II QWs, and narrow-gap materials. We described the potential of $In_yGa_{1-y}Bi_xAs_{1-x}$ and $GaBi_xN_yAs_{1-x-y}$ alloys for respective applications in InP-based mid-infrared photonics, and in multi-junction solar cells. In both cases, we showed that Bi-containing quaternary alloys enable a large degree of control over the band structure, allowing for flexible engineering of the band gap, spin-orbit-splitting energy, band offsets, and strain over wide ranges, thereby making it possible to design new photonic and photovoltaic materials and device structures with enhanced capabilities in the near- and mid-infrared. We described that $GaBi_xAs_{1-x}$/GaN_yAs_{1-y} type-II structures grown on GaAs offer a route to emission on GaAs at wavelengths out to and beyond 3 μm, in strain-balanced QWs and superlattices. We further highlighted emerging work on $In_yGa_{1-y}As$/$GaBi_xAs_{1-x}$ type-II QWs and metamorphic $InBi_xAs_{1-x}$ QWs, which, respectively, offer alternative routes to extending the spectral ranges accessible using GaAs and InP substrates. Finally, we described growing interest in narrow-gap (In)$GaBi_xSb_{1-x}$ alloys, which have significant potential for applications across the entirety of the mid-infrared as well as in spintronic devices.

Overall, we conclude that dilute bismide alloys are a rich and rapidly growing area of semiconductor research in which there remain a range of materials physics and device engineering challenges that must be overcome in order to realize the significant potential of this new class of III–V semiconductor materials for practical applications. The theoretical models and calculations reviewed in this chapter constitute the most complete analysis to date of the electronic, optical, and spin properties of dilute bismide alloys, and are expected to serve as a platform to support, interpret, and stimulate ongoing and future research on Bi-containing materials and devices.

Acknowledgments

The writing of this chapter was supported by the Engineering and Physical Sciences Research Council, UK (project nos. EP/K029665/1, EP/H005587/01 and EP/H050787/1), and by Science Foundation Ireland (project no. 15/IA/3082). The research described in this chapter was further supported by the European Commission Seventh Framework Programme project BIANCHO (project no. FP7-257974), by the Irish Research Council (RS/2010/2766), and by Science Foundation Ireland (project no. 10/IN.1/I2994). The authors thank Dr. M. Usman (University of Melbourne, Australia), Prof. K. Volz (Philipps-Universität Marburg, Germany), Prof. A. Krotkus (Center for Physical Sciences and Technology, Lithuania), Prof. J. P. R. David (University of Sheffield, UK), Prof. T. Tiedje (University of Victoria, Canada), Dr. S. Mazzucato (Institut National des Sciences Appliquées, Toulouse, France), Prof. J. M. O. Zide (University of Delaware, USA), and their colleagues, for supplying materials, devices, and experimental data to facilitate aspects of this research, as well as for many useful discussions on the physics of dilute bismide materials and devices. C.A.B. thanks Prof. J. M. Rorison (University of Bristol, UK) for useful feedback on an earlier version of this chapter. The authors also thank their colleagues at their respective institutions for the contributions they have made to the research described in this chapter.

References

1. S. Tixier, M. Adamcyk, T. Tiedje, S. Francoeur, A. Mascarenhas, P. Wei, and F. Schiettekatte. Molecular beam epitaxy growth of $GaAs_{1-x}Bi_x$. *Appl. Phys. Lett.*, 82:2245, 2003.
2. S. Francoeur, M. J. Seong, A. Mascarenhas, S. Tixier, M. Adamcyk, and T. Tiedje. Band gap of $GaAs_{1-x}Bi_x$, $0 < x < 3.6\%$. *Appl. Phys. Lett.*, 82:3874, 2003.
3. W. Huang, K. Oe, G. Feng, and M. Yoshimoto. Molecular-beam epitaxy and characteristics of $GaN_yAs_{1-x}Bi_x$. *J. Appl. Phys.*, 98:053505, 2005.
4. K. Alberi, J. Wu, W. Walukiewicz, K. M. Yu, O. D. Dubon, S. P. Watkinsa, C. X. Wang, X. Liu, Y. J. Cho, and J. Furdyna. Valence-band anticrossing in mismatched III-V semiconductor alloys. *Phys. Rev. B*, 75:045203, 2007.
5. M. Usman, C. A. Broderick, A. Lindsay, and E. P. O'Reilly. Tight-binding analysis of the electronic structure of dilute bismide alloys of GaP and GaAs. *Phys. Rev. B*, 84:245202, 2011.
6. Z. Batool, K. Hild, T. J. C. Hosea, X. Lu, T. Tiedje, and S. J. Sweeney. The electronic band structure of GaBiAs/GaAs layers: Influence of strain and band anti-crossing. *J. Appl. Phys.*, 111:113108, 2012.
7. B. Fluegel, S. Francoeur, A. Mascarenhas, S. Tixier, E. C. Young, and T. Tiedje. Giant spin-orbit bowing in $GaAs_{1-x}Bi_x$. *Phys. Rev. Lett.*, 97:067205, 2006.
8. W. Shan, W. Walukiewicz, J. W. Ager, E. E. Haller, J. F. Geisz, D. J. Friedman, J. M. Olson, and S. R. Kurtz. Band anticrossing in GaInNAs alloys. *Phys. Rev. Lett.*, 82:1221, 1999.
9. S. J. Sweeney. Light emitting semiconductor device. Patent no. WO2010149978 A1 (filed: June 25, 2010, issued: December 29, 2010).
10. S. J. Sweeney. Bismide-alloys for higher efficiency infrared semiconductor lasers. In *Proceedings of the 22nd IEEE International Semiconductor Laser Conference (ISLC)*, p. 111, 2010.
11. S. J. Sweeney, Z. Batool, K. Hild, S. R. Jin, and T. J. C. Hosea. The potential role of bismide alloys in future photonic devices. In *Proceedings of the 13th International Conference on Transparent Optical Networks*, p. 1, 2011.
12. C. A. Broderick, M. Usman, S. J. Sweeney, and E. P. O'Reilly. Band engineering in dilute nitride and bismide semiconductor lasers. *Semicond. Sci. Technol.*, 27:094011, 2012.
13. S. Sweeney, S. R. Jin. Bismide–nitride alloys: Promising for ecient light emitting devices in the nearand mid-infrared. *J. Appl. Phys.* 113, 043110, 2013.
14. S. Imhof, C. Bückers, A. Thränhardt, J. Hader, J. V. Moloney, and S. W. Koch. Microscopic theory of the optical properties of Ga(AsBi)/GaAs quantum wells. *Semicond. Sci. Technol.*, 23:125009, 2008.

15. C. A. Broderick, P. E. Harnedy, and E. P. O'Reilly. Theory of the electronic and optical properties of dilute bismide quantum well lasers. *IEEE J. Sel. Top. Quantum Electron.*, 21:1503313, 2015.
16. A. Geižutis, V. Pačebutas, R. Butkutė, P. Svidovskya, V. Strazdienėa, and A. Krotkus. Growth and characterization of UTC photo-diodes containing $\mathrm{GaAs}_{1-x}\mathrm{Bi}_x$ absorber layer., *Solid State Electron.*, 99:101, 2014.
17. J. J. Lee, J. D. Kim, and M. Razeghi. Growth and characterization of InSbBi for long wavelength infrared photodetectors. *Appl. Phys. Lett.*, 70:3266, 1997.
18. I. C. Sandall, F. Bastiman, B. White, R. Richards, D. Mendes, J. P. R. David, and C. H. Tan. Demonstration of InAsBi photoresponse beyond 3.5 μm. *Appl. Phys. Lett.*, 104:171109, 2014.
19. Y. Gu, Y. G. Zhang, X. Y. Chen, Y. J. Ma, S. P. Xi, B. Du, and H. Li. Nearly lattice-matched short-wave infrared InGaAsBi photodetectors on InP. *Appl. Phys. Lett.*, 108:032102, 2016.
20. C. J. Hunter, F. Bastiman, A. R. Mohmad, R. Richards, J. S. Ng, S. J. Sweeney, and J. P. R. David. Absorption characteristics of $\mathrm{GaAs}_{1-x}\mathrm{Bi}_x$/GaAs diodes in the near-infrared. *IEEE Photon. Tech. Lett.*, 24:2191, 2012.
21. S. J. Sweeney, K. Hild, and S. R. Jin. The potential of GaAsBiN for multi-junction solar cells. In *Proceedings of the 39th IEEE Photovoltaic Specialists Conference*, p. 2474, 2013.
22. T. Thomas, A. Mellor, N. P. Hylton, M. Führer, D. Alonso-Álvarez, A. Braun, N. J. Ekins-Daukes, J. P. R. David, and S. J. Sweeney. Requirements for a GaAsBi 1 eV sub-cell in a GaAs-based multi-junction solar cell. *Semicond. Sci. Technol.*, 30:094010, 2015.
23. S. J. Sweeney, and K. Hild. Light receiving device. Patent no. 20160149060 (filed: June 19, 2014, issued: May 26, 2016).
24. S. Mazzucato, T. T. Zhang, H. Carrère, D. Lagarde, P. Boonpeng, A. Arnoult, G. Lacoste, A. Balocchi, T. Amand, C. Fontaine, and X. Marie. Electron spin dynamics and g-factor in GaAsBi. *Appl. Phys. Lett.*, 102:252107, 2013.
25. R. A. Simmons, S. R. Jin, S. J. Sweeney, and S. K. Clowes. Enhancement of the Rashba interaction in GaAs/AlGaAs quantum wells due to the incorporation of bismuth. *Appl. Phys. Lett.*, 107:142401, 2015.
26. P. Dongmo, Y. Zhong, P. Attia, C. Bomberger, R. Cheaito, J. F. Ihlefeld, P. E. Hopkins, and J. M. O. Zide. Enhanced room temperature electronic and thermoelectric properties of the dilute bismuthide InGaBiAs. *J. Appl. Phys.*, 112:093710, 2012.
27. M. Usman, C. A. Broderick, Z. Batool, K. Hild, T. J. C. Hosea, S. J. Sweeney, and E. P. O'Reilly. Impact of alloy disorder on the band structure of compressively strained $\mathrm{GaBi}_x\mathrm{As}_{1-x}$. *Phys. Rev. B*, 87:115104, 2013.
28. M. Usman, and E. P. O'Reilly. Atomistic tight-binding study of the electronic structure and interband optical transitions in $\mathrm{GaBi}_x\mathrm{As}_{1-x}$/GaAs quantum wells. *Appl. Phys. Lett.*, 104:071103, 2014.
29. C. A. Broderick, S. Mazzucato, H. Carrère, T. Amand, H. Makhloufi, A. Arnoult, C. Fontaine, O. Donmez, A. Erol, M. Usman, E. P. O'Reilly, and X. Marie. Anisotropic electron g factor as a probe of the electronic structure of $\mathrm{GaBi}_x\mathrm{As}_{1-x}$/GaAs epilayers. *Phys. Rev. B*, 90:195301, 2014.
30. P. Ludewig, N. Knaub, N. Hossain, S. Reinhard, L. Natterman, I. P. Marko, S. R. Jin, K. Hild, S. Chatterjee, W. Stolz, S. J. Sweeney, and K. Volz. Electrical injection Ga(AsBi)/(AlGa)As single quantum well laser. *Appl. Phys. Lett.*, 102:242115, 2013.
31. S. J. Sweeney, A. F. Phillips, A. R. Adams, E. P. O'Reilly, and P. J. A. Thijs. The effect of temperature dependent processes on the performance of 1.5-μm compressively strained InGaAs(P) MQW semiconductor diode lasers. *IEEE Photon. Tech. Lett.*, 10:1076, 1998.
32. S. J. Sweeney, A. R. Adams, M. Silver, E. P. O'Reilly, J. R. Watling, A. B. Walker, and P. J. A. Thijs. Dependence of threshold current on QW position and on pressure in 1.5-μm InGaAs(P) lasers. *Phys. Stat. Sol. B*, 211:525, 1999.
33. A. R. Adams. Band-structure engineering for low-threshold high-efficiency semiconductor lasers. *Electron. Lett.*, 22:249, 1986.

34. Y. Tominaga, K. Oe, and M. Yoshimoto. Low temperature dependence of oscillation wavelength in $\mathrm{GaAs}_x\mathrm{Bi}_{1-x}$ laser by photo-pumping. *Appl. Phys. Express*, 3:062201, 2010.
35. R. Butkutė, A. Geižutis, V. Pačebutas, B. Čechavičius, V. Bukauskas, R. Kundrotas, P. Ludewig, K. Volz, and A. Krotkus. Multi-quantum well Ga(AsBi)/GaAs laser diodes with more than 6% of bismuth. *Electron. Lett.*, 50:1155, 2014.
36. I. P. Marko, S. R. Jin, K. Hild, Z. Batool, Z. L. Bushell, P. Ludewig, W. Stolz, K. Volz, V. Pačebutas, A. Geižutis, A. Krotkus, and S. J. Sweeney. Properties of hybrid MOVPE/MBE grown GaAsBi/GaAs-based near-infrared emitting quantum well lasers. *Semicond. Sci. Technol.*, 30:094008, 2015.
37. P. Ludewig, L. Natterman, W. Stolz, and K. Volz. MOVPE growth mechanisms of dilute bismide III/V alloys. *Semicond. Sci. Technol.*, 30:094017, 2015.
38. R. D. Richards, F. Bastiman, C. J. Hunter, D. F. Mendes, A. R. Mohmad, J. S. Roberts, and J. P. R. David. Molecular beam epitaxy growth of GaAsBi using As_2 and As_4. *J. Cryst. Growth*, 390:120, 2014.
39. I. P. Marko, C. A. Broderick, S. R. Jin, P. Ludewig, W. Stolz, K. Volz, J. M. Rorison, E. P. O'Reilly, and S. J. Sweeney. Optical gain in GaAsBi/GaAs quantum well diode lasers. *Sci. Rep.*, 6:28863, 2016.
40. I. P. Marko, Z. Batool, K. Hild, S. R. Jin, N. Hossain, T. J. C. Hosea, J. P. Petropoulos, Y. Zhong, P. B. Dongmo, J. M. O. Zide, and S. J. Sweeney. Temperature and Bi-concentration dependence of the bandgap and spin–orbit splitting in InGaBiAs/InP semiconductors for mid-infrared applications. *Appl. Phys. Lett.*, 101:221108, 2012.
41. S. R. Jin, and S. J. Sweeney. InGaAsBi alloys on InP for effcient near-and mid-infrared light emitting devices. *J. Appl. Phys.* 114:213103, 2013.
42. C. A. Broderick, M. Usman, and E. P. O'Reilly. 12-band **k·p** model for dilute bismide alloys of (In)GaAs derived from supercell calculations. *Phys. Stat. Sol. B*, 250:773, 2013.
43. G. M. T. Chai, C. A. Broderick, S. R. Jin, J. P. Petropoulous, Y. Zong, P. B. Dongmo, J. M. O. Zide, E. P. O'Reilly, S. J. Sweeney, and T. J. C. Hosea. Experimental and modelling study of InGaBiAs/InP alloys with up to 5.8% Bi, and with $\Delta_{\mathrm{SO}} > E_{\mathrm{g}}$. *Semicond. Sci. Technol.*, 30:094015, 2015.
44. C. A. Broderick, W. Xiong, and J. M. Rorison. Theory of InGaBiAs dilute bismide alloys for highly efficient InP-based mid-infrared semiconductor lasers. In *Proceedings of the 16th International Conference on Numerical Simulation of Optoelectronic Devices*, p. 47, 2016.
45. C. A. Broderick, W. Xiong, S. J. Sweeney, E. P. O'Reilly, J. M. Rorison, Theory and design of $\mathrm{In}_y\mathrm{Ga}_{1-y}\mathrm{As}_{1-x}\mathrm{Bi}_x$ mid-infrared semiconductor lasers: type-I quantum wells for emission beyond 3 μm on InP substrates. Submitted, 2017.
46. C. A. Broderick, S. R. Jin, I. P. Marko, K. Hild, P. Ludewig, Z. L. Bushell, W. Stolz, J. M. Rorison, E. P. O'Reilly, K. Volz, and S. J. Sweeney. $\mathrm{GaAs}_{1-x}\mathrm{Bi}_x/\mathrm{GaN}_y\mathrm{As}_{1-y}$ type-II quantum wells: Novel strain-balanced heterostructures for GaAs-based near- and mid-infrared photonics. *Sci. Rep.* 7, 46371, 2017.
47. A. Janotti, S. H. Wei, and S. B. Zhang. Theoretical study of the effects of isovalent coalloying of Bi and N in GaAs. *Phys. Rev. B*, 65:115203, 2002.
48. P. Carrier, and S. H. Wei. Calculated spin-orbit splitting of all diamondlike and zinc-blende semiconductors: Effects of $p_{1/2}$ local orbitals and chemical trends. *Phys. Rev. B*, 70:035212, 2004.
49. Y. Zhang, A. Mascarenhas, and L. W. Wang. Similar and dissimilar aspects of III–V semiconductors containing Bi versus N. *Phys. Rev. B*, 71:155201, 2005.
50. E. P. O'Reilly, A. Lindsay, P. J. Klar, A. Polimeni, and M. Capizzi. Trends in the electronic structure of dilute nitride alloys. *Semicond. Sci. Technol.*, 24:033001, 2009.
51. P. Vogl, H. P. Hjalmarson, and J. D. Dow. A semi-empirical tight-binding theory of the electronic structure of semiconductors. *J. Phys. Chem. Solids*, 44:365, 1983.
52. E. P. O'Reilly, A. Lindsay, S. Tomić, and M. Kamal-Saadi. Tight-binding and **k·p** models for the electronic structure of Ga(In)NAs and related alloys. *Semicond. Sci. Technol.*, 17:870, 2002.
53. A. Lindsay, and E. P. O'Reilly. Unification of the band-anticrossing and cluster-state models of dilute nitride semiconductor alloys. *Phys. Rev. Lett.*, 93:196402, 2004.
54. C. A. Broderick, M. Usman, and E. P. O'Reilly. Derivation of 12- and 14-band **k·p** Hamiltonians for dilute bismide and bismide–nitride alloys. *Semicond. Sci. Technol.*, 28:125025, 2013.

55. F. A. Trumbore, M. Gershenzon, and D. G. Thomas. Luminescence due to the isoelectronic substitution of bismuth for phosphorus in gallium phosphide. *Appl. Phys. Lett.*, 9:4, 1966.
56. H. X. Deng, J. Li, S. S. Li, H. Peng, J. B. Xia, L. W. Wang, and S. H. Wei. Band crossing in isovalent semiconductor alloys with large size mismatch: First-principles calculations of the electronic structure of Bi and N incorporated GaAs. *Phys. Rev. B*, 82:193204, 2010.
57. R. S. Joshya, A. J. Ptak, R. France, A. Mascarenhas, and R. N. Kini. Resonant state due to Bi in the dilute bismide alloy $GaAs_{1-x}Bi_x$. *Phys. Rev. B*, 90:165203, 2014.
58. K. Alberi, D. A. Beaton, and A. Mascarenhas. Direction observation of the $E\_$ resonant state in $GaAs_{1-x}Bi_x$. *Phys. Rev. B*, 92:241201, 2015.
59. V. Virkkala, V. Havu, F. Tuomisto, and M. J. Puska. Modeling Bi-induced changes in the electronic structure of $GaAs_{1-x}Bi_x$ alloys. *Phys. Rev. B*, 88:235201, 2013.
60. O. Rubel, A. Bokhanchuk, S. J. Ahmed, and E. Assmann. Unfolding the band structure of disordered solids: From bound states to high-mobility Kane fermions. *Phys. Rev. B*, 90:115202, 2014.
61. J. Hader, J. V. Moloney, O. Rubel, S. C. Badescu, S. Johnson, and S. W. Koch. Microscopic modelling of opto-electronic properties of dilute bismide materials for the mid-IR. *Proc. SPIE*, 9767:976709, 2016.
62. M. Mbarki and A. Rebey. First-principles calculation of the physical properties of $GaAs_{1-x}Bi_x$ alloys. *Semicond. Sci. Technol.*, 26:105020, 2011.
63. R. Kudraweic, J. Kopaczek, M. P. Polak, P. Scharoch, M. Gladysiewicz, R. D. Richards, F. Bastiman, and J. P. R. David. Experimental and theoretical studies of band gap alignment in $GaAs_{1-x}Bi_x$/GaAs quantum wells. *J. Appl. Phys.*, 116:233508, 2014.
64. M. P. Polak and P. Scharoch, and R. Kudraweic. First-principles calculations of bismuth-induced changes in the band structure of dilute Ga–V–Bi and In–V–Bi alloys: Chemical trends versus experimental data. *Semicond. Sci. Technol.*, 30:094001, 2015.
65. M. Ferhat and A. Zaoui. Structural and electronic properties of III–V bismuth compounds. *Phys. Rev. B*, 73:115107, 2006.
66. A. Belabbes, A. Zaoui, and M. Ferhat. Lattice dynamics study of bismuth III–V compounds. *J. Phys.: Condens. Matter*, 20:415221, 2008.
67. H. Jacobsen, B. Puchala, T. F. Keuch, and D. Morgan. Ab initio study of the strain dependent thermodynamics of Bi doping in GaAs. *Phys. Rev. B*, 86:085207, 2012.
68. G. Luo, S. Yang, J. Li, M. Arjmand, I. Szlufarska, A. S. Brown, T. F. Keuch, and D. Morgan. First-principles studies on molecular beam epitaxy growth of $GaAs_{1-x}Bi_x$. *Phys. Rev. B*, 92:035415, 2015.
69. M. P. J. Punkinnen, P. Laukkanen, M. Kuzmin, H. Levämäki, J. Løang, M. Tuominen, M. Yasir, J. Dahl, S. Lu, E. K. Delczeg-Czirjak, L. Vitos, and K. Kokko. Does Bi form clusters in $GaAs_{1-x}Bi_x$ alloys. *Semicond. Sci. Technol.*, 29:115007, 2014.
70. M. P. J. Punkinnen, A. Lahti, P. Laukkanen, M. Kuzmin, M. Tuominen, M. Yasir, J. Dahl, J. Makela, H. L. Zhang, L. Vitos, and K. Kokko. Thermodynamics of pseudobinary $GaAs_{1-x}Bi_x$ $(0 \leq x \leq 1)$ alloys studied by different exchange-correlation functionals, special quasi-random structures and Monte Carlo simulations. *Computat. Condens. Matter*, 5:7, 2015
71. P. N. Keating. Effect of invariance requirements on the elastic strain energy of crystals with application to the diamond structure. *Phys. Rev.*, 145:637, 1966.
72. O. L. Lazarenkova, P. von Allmen, F. Oyafuso, S. Lee, and G. Klimeck. Effect of anharmonicity of the strain energy on band offsets in semiconductor nanostructures. *Appl. Phys. Lett.*, 85:4193, 2004.
73. J. C. Slater and G. F. Koster. Simplified LCAO method for the periodic potential problem. *Phys. Rev.*, 94:1498, 1954.
74. J. Yoshida, T. Kita, O. Wada, and K. Oe. Temperature dependence of $GaAs_{1-x}Bi_x$ band gap studied by photo-reflectance spectroscopy. *Jpn. J. Appl. Phys.*, 42:371, 2003.
75. G. Hendorfer and J. Schneider. g-factor and effective mass anisotropies in pseudomorphic strained layers. *Semicond. Sci. Technol.*, 6:595, 1991.
76. P. L. Jeune, D. Robart, X. Marie, T. Amand, M. Brousseau, J. Barrau, V. Kalevich, and D. Rodichev. Anisotropy of the electron Landé g-factor in quantum wells. *Semicond. Sci. Technol.*, 12:380, 1997.

77. M. Oestreich and S. Hallstein, W. W. Rühle. Spin quantum beats in semiconductors. *IEEE J. Sel. Top. Quantum Electron.*, 2:747, 1996.
78. A. P. Heberle, W. W. Rühle, and K. Ploog. Quantum beats of electron Larmor precession in GaAs wells. *Phys. Rev. Lett.*, 72:3887, 1994.
79. C. A. Broderick, P. E. Harnedy, P. Ludewig, Z. L. Bushell, K. Volz, R. J. Manning, and E. P. O'Reilly. Determination of type-I band offsets in $\text{GaBi}_x\text{As}_{1-x}$ quantum wells using polarisation-resolved photovoltage spectroscopy and 12-band $\mathbf{k}\cdot\mathbf{p}$ calculations. *Semicond. Sci. Technol.*, 30:094009, 2015.
80. T. B. Bahder. 8-Band $\mathbf{k}\cdot\mathbf{p}$ model of strained zinc-blende crystals. *Phys. Rev. B*, 41:11992, 1990.
81. G. L. Bir, G. E. Pikus. *Symmetry and Strain-Induced Effects in Semiconductors* New York: Wiley, 1974.
82. C. G. Van de Walle. Band lineups and deformation potentials in the model-solid theory. *Phys. Rev. B*, 39:1871, 1989.
83. M. P. C. M. Krijn. Heterojunction band offsets and effective masses in III–V quaternary alloys. *Semicond. Sci. Technol.*, 6:27, 1991.
84. I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan. Band parameters for III–V compound semiconductors and their alloys. *J. Appl. Phys.*, 89:5815, 2001.
85. M. Ehrhardt and T. Koprucki, editors. *Multiband Effective Mass Approximations: Advanced Mathematical Models and Numerical Techniques*. Berlin: Springer, 2014.
86. S. B. Healy, and E. P. O'Reilly. Influence of electrostatic confinement on optical gain in GaInNAs quantum-well lasers. *IEEE J. Quant. Electron.*, 42:608, 2006.
87. A. T. Meney, B. Gonul, and E. P. O'Reilly. Evaluation of various approximations used in the envelope-function method. *Phys. Rev. B*, 50:10893, 1994.
88. F. Szmulowicz. Derivation of a general expression for the momentum matrix elements within the envelope-function approximation. *Phys. Rev. B*, 51:1613, 1995.
89. C. S. Chang, S. L. Chuang, J. R. Minch, W. C. W. Fang, Y. K. Chen, and T. Tanbun-Ek. Amplified spontaneous emission spectroscopy in strained quantum-well lasers. *IEEE J. Sel. Top. Quantum Electron.*, 1:1100, 1995.
90. S. Tomić, E. P. O'Reilly, R. Fehse, S. J. Sweeney, A. R. Adams, A. D. Andreev, S. A. Choulis, T. J. C. Hosea, and H. Riechert. Theoretical and experimental analysis of 1.3-μm InGaAsN/GaAs lasers. *IEEE J. Sel. Top. Quantum Electron.*, 9:1228, 2003.
91. K. Kawano and T. Kitoh. *Introduction to Optical Waveguide Analysis: Solving Maxwell's Equation and the Schrödinger Equation*. New York, NY: Wiley, 2001.
92. M. Gladysiewicz, R. Kudrawiec, and M. S. Wartak. 8-Band and 14-band $\mathbf{k}\cdot\mathbf{p}$ modeling of electronic band structure and material gain in Ga(In)AsBi quantum wells grown on GaAs and InP substrates. *J. Appl. Phys.*, 118:055702, 2015.
93. E. P. O'Reilly and A. R. Adams. Band-structure engineering in strained semiconductor lasers. *IEEE J. Quant. Electron.*, 30:366, 1994.
94. S. L. Chuang. *Physics of Photonic Devices*. New York, NY: Wiley, 2009.
95. C. A. Broderick, I. P. Marko, J. M. Rorison, S. J. Sweeney, and E. P. O'Reilly. GaAs-based dilute bismide semiconductor lasers: Theory vs. experiment. In *Proceedings of the 16th International Conference on Numerical Simulation of Optoelectronic Devices*, p. 209, 2016.
96. I. P. Marko, P. Ludewig, Z. L. Bushell, S. R. Jin, K. Hild, Z. Batool, S. Reinhard, L. Natterman, W. Stolz, K. Volz, and S. J. Sweeney. Physical properties and optimization of GaBiAs/(Al)GaAs based near-infrared laser diodes grown by MOVPE with up to 4.4% Bi. *J. Phys. D: Appl. Phys.*, 47:345103, 2014.
97. A. Ghiti, M. Silver, and E. P. O'Reilly. Low threshold current and high differential gain in ideal tensile- and compressive-strained quantum-well lasers. *J. Appl. Phys.*, 71:4626, 1992.
98. M. Silver and E. P. O'Reilly. Optimization of long wavelength InGaAsP strained quantum-well lasers. *IEEE J. Quant. Electron.*, 31:1193, 1995.
99. S. Tomić and E. P. O'Reilly. Optimization of material parameters in 1.3-μm InGaAsN-GaAs lasers. *IEEE Photon. Tech. Lett.*, 15:6, 2003.

100. P. Blood, G. M. Lewis, P. M. Smowton, H. Summers, J. Thomson and J. Lutti. Characterization of semiconductor laser gain media by the segmented contact method. *IEEE J. Sel. Top. Quantum Electron.*, 9:1275, 2003.
101. D. J. Bossert and D. Gallant. Gain, refractive index, and α-parameter in InGaAs–GaAs SQW broad-area lasers. *IEEE Photon. Tech. Lett.*, 8:322, 1996.
102. A. Bauer, K. Rößner, T. Lehnhardt, M. Kamp, S. Höfling, L. Worschech, and A. Forchel. Mid-infrared semiconductor heterostructure lasers for gas sensing applications. *Semicond. Sci. Technol.*, 26:014032, 2011.
103. E. Tournié and A. N. Baranov. Mid-infrared semiconductor lasers: A review. *Semiconduct. Semimet.*, 86:183, 2012.
104. G. Feng, M. Yoshimoto, K. Oe, A. Chayahara, and Y. Horino. New III–V semiconductor InGaAsBi alloy grown by molecular beam epitaxy. *Jpn. J. Appl. Phys.*, 44:L1161, 2005.
105. G. Feng, K. Oe, and M. Yoshimoto. Bismuth containing III–V quaternary alloy InGaAsBi grown by MBE. *Phys. Stat. Sol. A*, 203:2760, 2006.
106. J. P. Petropoulos, Y. Zhong, and J. M. O. Zide. Optical and electrical characterization of InGaAsBi for use as a mid-infrared optoelectronic material. *Appl. Phys. Lett.*, 99:031110, 2011.
107. Y. Zhong, P. B. Dongmo, J. P. Petropoulos, and J. M. O. Zide. Effects of molecular beam epitaxy growth conditions on composition and optical properties of $In_xGa_{1-x}Bi_yAs_{1-y}$. *Appl. Phys. Lett.*, 100:112110, 2012.
108. J. Devenson, V. Pačebutas, R. Butkutė, A. Baranov, and A. Krotkus. Structure and optical properties of InGaAsBi with up to 7% bismuth. *Appl. Phys. Express*, 5:015503, 2012.
109. S. Zhou, M. Qi, L. Ai, A. Xu, and S. Wang. Effects of buffer layer preparation and Bi concentration on InGaAsBi layers grown by gas source molecular beam epitaxy. *Semicond. Sci. Technol.*, 30:125001, 2015.
110. M. Yoshimoto, W. Huang, Y. Takehara, J. Saraie, A. Chayahara, Y. Horino, and K. Oe. New semiconductor GaNAsBi alloy grown by molecular beam epitaxy. *Jpn. J. Appl. Phys.*, 43:L845, 2004.
111. S. Tixier, S. E. Webster, E. C. Young, T. Tiedje, S. Francoeur, A. Mascarenhas, P. Wei, and F. Schiettekatte. Band gaps for the dilute quaternary alloys $GaN_xAs_{1-x-y}Bi_y$ and $Ga_{1-y}In_yN_xAs_{1-x}$. *Appl. Phys. Lett.*, 86:112113, 2005.
112. Z. L. Bushell, P. Ludewig, N. Knaub, Z. Batool, K. Hild, W. Stolz, S. J. Sweeney, and K. Volz. Growth and characterisation of Ga(AsNBi) alloy by metal-organic vapour phase epitaxy. *J. Cryst. Growth*, 396:79, 2014.
113. M. Yoshimoto, W. Huang, G. Feng, Y. Tanaka, and K. Oe. Molecular-beam epitaxy of GaNAsBi layer for temperature-insensitive wavelength emission. *J. Cryst. Growth*, 301–302:975, 2007.
114. J. Hwang, J. D. Phillips. Band structure of strain-balanced GaAsBi/GaAsN superlattices on GaAs. *Phys. Rev. B*, 83:195327, 2011.
115. W. Pan, L. Zhang, L. Zhu, Y. Li, X. Chen, X. Wu, F. Zhang, J. Shao, and S. Wang. Optical properties and band bending in InGaAs/GaAsBi/InGaAs type-II quantum well grown by gas source molecular beam epitaxy. *J. Appl. Phys.*, 120:105702, 2016.
116. J. F. Klem, O. Blum, S. R. Kurtz, I. J. Fritz, and K. D. Choquette. GaAsSb/InGaAs type-II quantum wells for long-wavelength lasers on GaAs substrates. *J. Vac. Sci. Technol. B*, 18:1605, 2000.
117. J. Y. Yeh, L. J. Mawst, A. A. Khandekar, T. F. Kuech, I. Vurgaftman, J. R. Meyer, and N. Tansu. Long wavelength emission of InGaAsN/GaAsSb type II “W” quantum wells. *Appl. Phys. Lett.*, 88:051115, 2006.
118. C. Berger, C. Möller, P. Hens, C. Fuchs, W. Stolz, S. W. Koch, A. R. Perez, J. Hader, and J. V. Moloney. Novel type-II material system for laser applications in the near-infrared regime. *AIP Adv.*, 5:047105, 2015.
119. K. Forghani, Y. Guan, M. Losurdo, G. Luo, D. Morgan, S. E. Babcock, A. S. Brown, L. J. Mawst, and T. F. Kuech. $GaAs_{1-y-z}P_yBi_z$, an alternative reduced band gap alloy system lattice-matched to GaAs. *Appl. Phys. Lett.*, 105:111101, 2014.

120. H. Kim, K. Forghani, Y. Guan, G. Luo, A. Anand, D. Morgan, T. F. Keuch, L. J. Mawst, Z. R. Lingley, B. J. Foran, and Y. Sin. Strain-compensated $GaAs_{1-y-z}P_y/GaAs_{1-x}Bi_x/GaAs_{1-y}P_y$ quantum wells for laser applications. *Semicond. Sci. Technol.*, 30:094011, 2015.
121. Y. Gu, Y. G. Zhang, X. Y. Chen, Y. J. Ma, S. P. Xi, B. Du, W. Y. Ji, and Y. H. Shi. Metamorphic $InAs_{1-x}Bi_x/In_{0.83}Al_{0.17}As$ quantum well structures on InP for mid-infrared emission. *Appl. Phys. Lett.*, 109:122102, 2016.
122. M. K. Rajpalke, W. M. Linhart, M. Birkett, K. M. Yu, D. O. Scanlon, J. Buckeridge, T. S. Jones, M. J. Ashwin, and T. D. Veal. Growth and properties of GaSbBi alloys. *Appl. Phys. Lett.*, 103:142106, 2013.

III

Nanostructures

11 Quantum Wells *Seoung-Hwan Park and Doyeol Ahn* 365
Introduction • Band Structures of Bulk Semiconductors • Band Structures of Strained-Layer QW • Optical Matrix Elements • Band Structures of Bulk Semiconductors with an Arbitrary Crystal Orientation • Optical Gain Model with Many-Body Effects • Numerical Example • Summary

12 Nanowires *Oliver Marquardt, Vladimir M. Kaganer, and Pierre Corfdir* 397
Introduction: Semiconductor Nanowires • Basic Charge Confining Mechanisms in NWs • Doping and the Influence of Surfaces • Crystal-Phase Bandgap Engineering • Optical Anisotropy of NWs—The Antenna Effect

13 Quantum Dots *Stanko Tomić and Nenad Vukmirović*.. 419
Introduction • The Core/Shell QD Structure • The $\mathbf{k}\cdot\mathbf{p}$ Method for SP Electronic Structure • PW Implementation of the $\mathbf{k}\cdot\mathbf{p}$ Hamiltonian and Coulomb Integrals for QD Structures • The Effect of Dielectric Confinement • The Configuration Interaction Hamiltonian for NQD • Transport Properties of QD Nanocrystal Arrays • Practical Examples: Cd-Based Chalcogenide QDs • Conclusion

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

11

Quantum Wells

11.1 Introduction 365
11.2 Band Structures of Bulk Semiconductors 366
Zinc-Blende Hamiltonian of the (001) Orientation • Wurtzite Hamiltonian of the (0001) Orientation
11.3 Band Structures of Strained-Layer QW 371
Zinc-Blende Semiconductor • Wurtzite Semiconductor
11.4 Optical Matrix Elements 377
Zinc-Blende Structure • Wurtzite Structure
11.5 Band Structures of Bulk Semiconductors with an Arbitrary Crystal Orientation 379
Zinc-Blende Structure • Wurtzite Structure • Interband Optical-Matrix Elements for QW with an Arbitrary Crystal Orientation
11.6 Optical Gain Model with Many-Body Effects 387
11.7 Numerical Example 389
Zinc-Blende Structure • Wurtzite Structure
11.8 Summary 395

Seoung-Hwan Park
and
Doyeol Ahn

11.1 Introduction

A quantum well (QW) is a heterostructure in which one thin-well layer is surrounded by two barrier layers. This layer is so thin that both electrons and holes are quantized. The electronic and the optical properties of quantized states offer new opportunities in developing practical devices, such as QW infrared photodetectors, quantum cascade lasers, all-optical switches, modulators, and many others [1–4]. Hence, it is very important to obtain eigenvalues and wave functions for the design of the active region in these optoelectronic devices based on QW structures. In this chapter, we review theoretical formalism to calculate eigenvalues and wave functions of (001)-oriented zinc-blende and (0001)-oriented wurtzite QW structures [5–10]. We block diagonalize zinc-blende and wurtzite Luttinger–Kohn 6×6 Hamiltonians for the valence bands to two 3×3 Hamiltonians, which have analytical solutions for eigenvalues and eigenvectors. We derive several important forms such as interband optical matrix elements and optical gains [11–15]. Also, as a numerical example, we calculate eigenvalues and wave functions for zinc-blende and wurtzite Hamiltonians using a finite-difference method (FDM) [4]. On the basis of this information, we discuss crystal orientation effects on electronic and optical properties of strained zinc-blende and wurtzite QW structures, including the Hamiltonian for nonpolar wurtzite QW structures.

11.2 Band Structures of Bulk Semiconductors

11.2.1 Zinc-Blende Hamiltonian of the (001) Orientation

11.2.1.1 6 × 6 Hamiltonian for the Valence Band

The Luttinger–Kohn Hamiltonian for the valence band of the (001)-oriented zinc-blende semiconductor is given by [2,4,5]

$$H^{\mathrm{LK}}(\vec{k},\bar{\bar{\varepsilon}}) = -\begin{pmatrix} P+Q & -S & R & 0 & -S/\sqrt{2} & \sqrt{2}R \\ -S^{\dagger} & P-Q & 0 & R & -\sqrt{2}Q & \sqrt{3/2}S \\ R^{\dagger} & 0 & P-Q & 0 & \sqrt{3/2}S^{\dagger} & \sqrt{2}Q \\ 0 & R^{+} & S^{\dagger} & P+Q & -\sqrt{2}R^{\dagger} & -S^{\dagger}/\sqrt{2} \\ -S^{\dagger}/\sqrt{2} & 0 & \sqrt{3/2}S & -\sqrt{2}R & P+\Delta & 0 \\ \sqrt{2}R^{\dagger} & \sqrt{3/2}S^{\dagger} & \sqrt{2}Q^{\dagger} & -S/\sqrt{2} & 0 & P+\Delta \end{pmatrix}, \tag{11.1}$$

where

$$\begin{aligned} &P = P_k + P_\varepsilon, \quad Q = Q_k + Q_{\varepsilon,} \\ &R = R_k + R_\varepsilon, \quad S = S_k + S_\varepsilon, \\ &P_k = \left(\frac{\hbar^2}{2m_0}\right)\gamma_1(k_x^2 + k_x^2 + k_z^2), \quad Q_k = \left(\frac{\hbar^2}{2m_0}\right)\gamma_2(k_z^2 + k_x^2 - 2k_z^2) \\ &R_k = \left(\frac{\hbar^2}{2m_0}\right)\sqrt{3}[-\gamma_2(k_x^2 - k_x^2) + 2i\gamma_3 k_x k_y], \quad S_k = \left(\frac{\hbar^2}{2m_0}\right)2\sqrt{3}\gamma_3(k_x - ik_y)k_z, \\ &P_\varepsilon = -a_v(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz}), \quad Q_\varepsilon = -\frac{b}{2}(\varepsilon_{xx} + \varepsilon_{yy} - 2\varepsilon_{zz}), \\ &R_\varepsilon = \frac{\sqrt{3}}{2}b(\varepsilon_{xx} - \varepsilon_{yy}) - id\varepsilon_{xy}, \quad S_\varepsilon = -d(\varepsilon_{zx} - i\varepsilon_{yz}), \end{aligned} \tag{11.2}$$

and $\bar{\bar{\varepsilon}} = (\varepsilon_{ij})$ for $i, j = x, y, z$ is a symmetric strain tensor; γ_1, γ_2, and γ_3 are the Luttinger parameters; a_{v}, b, and d are the Bir–Pikus deformation potentials; Δ is the spin–orbit split-off energy; m_0 is the free electron mass; $\hbar$ is Planck's constant divided by 2π; and k_i is the wave vector. The superscript † means taking both transpose (~) and complex conjugate (*). In the Hamiltonian H, we restricted ourselves to the biaxial strain case for simplicity, namely,

$$\varepsilon_{xx} = \varepsilon_{yy} \neq \varepsilon_{zz}, \quad \varepsilon_{ij} = 0 \quad \text{for} \quad i \neq j. \tag{11.3}$$

For the case of a strained-layer semiconductor pseudomorphically grown on a (001)-oriented substrate,

$$\varepsilon_{xx} = \varepsilon_{yy} = \frac{a_{\mathrm{s}}^z - a_{\mathrm{l}}^z}{a_{\mathrm{l}}^z}, \quad \varepsilon_{zz} = -\frac{2C_{12}}{C_{11}}\varepsilon_{xx}, \tag{11.4}$$

where a_{s}^z and a_{l}^z are the lattice constants of the substrate (s) and the layer (l) material, and C_{11} and C_{12} are the stiffness constants for the zinc-blende structure. The bases for the Hamiltonian are

$$|u_1\rangle = \left|\frac{3}{2}, \frac{3}{2}\right\rangle = -\frac{1}{\sqrt{2}}\left|(X + iY)\uparrow\right\rangle,$$

$$
\begin{aligned}
\left|u_2\right\rangle &= \left|\frac{3}{2},\frac{1}{2}\right\rangle = \frac{1}{\sqrt{6}}\left|-(X-iY)\downarrow +2Z\uparrow\right\rangle,\\
\left|u_3\right\rangle &= \left|\frac{3}{2},-\frac{1}{2}\right\rangle = \frac{1}{\sqrt{6}}\left|(X-iY)\uparrow +2Z\downarrow\right\rangle,\\
\left|u_4\right\rangle &= \left|\frac{3}{2},-\frac{3}{2}\right\rangle = \frac{1}{\sqrt{2}}\left|(X-iY)\downarrow\right\rangle, \qquad (11.5)\\
\left|u_5\right\rangle &= \left|\frac{1}{2},\frac{1}{2}\right\rangle = \frac{1}{\sqrt{3}}\left|(X+iY)\downarrow +Z\uparrow\right\rangle,\\
\left|u_6\right\rangle &= \left|\frac{1}{2},-\frac{1}{2}\right\rangle = \frac{1}{\sqrt{3}}\left|(X-iY)\uparrow -Z\downarrow\right\rangle.
\end{aligned}
$$

11.2.1.2 Block-Diagonalized 3 × 3 Hamiltonian

Under the axial approximation [2], we write the R_k term

$$
\begin{aligned}
R_k &= -\left(\frac{\hbar^2}{2m_0}\right)\sqrt{3}\left[\frac{\gamma_2+\gamma_2}{2}(k_x-ik_x)^2+\frac{\gamma_2-\gamma_3}{2}(k_x+ik_x)^2\right]\\
&\cong -\left(\frac{\hbar^2}{2m_0}\right)\sqrt{3}\left[\bar{\gamma}(k_x-ik_x)^2\right],
\end{aligned} \tag{11.6}
$$

where $\bar{\gamma}=\frac{\gamma_2+\gamma_3}{2}$. In this approximation, we assume $\gamma_2 \cong \gamma_3$ in the R_k term only, whereas we still use γ_2 and γ_3 in the other terms. When we define the angle φ by

$$k_x + ik_x = k_{||}e^{i\varphi}, \tag{11.7}$$

we can write

$$R_k = R_\rho e^{-2i\varphi},\quad S_k = S_\rho e^{-i\varphi},\quad R_\rho = -\left(\frac{\hbar^2}{2m_0}\right)\sqrt{3}\bar{\gamma}k_{||}^2,\quad S_\rho = \left(\frac{\hbar^2}{2m_0}\right)2\sqrt{3}\gamma_3 k_{||}k_z, \tag{11.8}$$

where $k_{||}=\sqrt{k_x^2+k_x^2}$. Then, the 6×6 Hamiltonian can be block-diagonalized into two 3×3 Hamiltonians by using the transformation matrix U

$$
\begin{aligned}
H &= UH^{\mathrm{LK}}U^{+}\\
&= -\begin{pmatrix}
P+Q & -R_\rho-iS_\rho & -\sqrt{2}R_\rho+i\sqrt{\frac{1}{2}}S_\rho & 0 & 0 & 0\\
-R_\rho+iS_\rho & P-Q & \sqrt{2}Q+i\sqrt{\frac{3}{2}}S_\rho & 0 & 0 & 0\\
-\sqrt{2}R_\rho-i\sqrt{\frac{1}{2}}S_\rho & \sqrt{2}Q-i\sqrt{\frac{3}{2}}S_\rho & P+\Delta & 0 & 0 & 0\\
0 & 0 & 0 & P+Q & -R_\rho+iS_\rho & -\sqrt{2}R_\rho-i\sqrt{\frac{1}{2}}S_\rho\\
0 & 0 & 0 & -R_\rho-iS_\rho & P-Q & \sqrt{2}Q-i\sqrt{\frac{3}{2}}S_\rho\\
0 & 0 & 0 & -\sqrt{2}R_\rho+i\sqrt{\frac{1}{2}}S_\rho & \sqrt{2}Q+i\sqrt{\frac{3}{2}}S_\rho & P+\Delta
\end{pmatrix}
\end{aligned} \tag{11.9}
$$

and

$$U = \begin{pmatrix} \alpha & 0 & 0 & i\alpha^* & 0 & 0 \\ 0 & -i\beta & -\beta^* & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & i\beta & -\beta^* \\ \alpha & 0 & 0 & -i\alpha^* & 0 & 0 \\ 0 & i\beta & -\beta^* & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -i\beta & -\beta^* \end{pmatrix}, \tag{11.10}$$

where

$$\alpha = \frac{1}{\sqrt{2}} e^{i\frac{3}{2}\varphi}, \quad \beta = \frac{1}{\sqrt{2}} e^{i\frac{1}{2}\varphi}. \tag{11.11}$$

A new basis set is given by using the basis transformation, $|i>= \sum T_{ij}|u_j>$, where $T = U^*$. That is,

$$\begin{aligned}
|1\rangle &= -\frac{1}{2}\Big|(X+iY)\uparrow\Big\rangle e^{-i\frac{3}{2}\varphi} - i\frac{1}{2}\Big|(X-iY)\downarrow\Big\rangle e^{-i\frac{3}{2}\varphi} \\
|2\rangle &= i\frac{1}{2\sqrt{3}}\Big|-(X-iY)\downarrow +2Z\uparrow\Big\rangle e^{-i\frac{1}{2}\varphi} - \frac{1}{2\sqrt{3}}\Big|(X-iY)\uparrow +2Z\downarrow\Big\rangle e^{i\frac{1}{2}\varphi} \\
|3\rangle &= -i\frac{1}{\sqrt{6}}\Big|(X+iY)\downarrow +Z\uparrow\Big\rangle e^{-i\frac{1}{2}\varphi} - \frac{1}{\sqrt{6}}\Big|(X-iY)\uparrow -Z\downarrow\Big\rangle e^{i\frac{1}{2}\varphi} \\
|4\rangle &= -\frac{1}{2}\Big|(X+iY)\uparrow\Big\rangle e^{-i\frac{3}{2}\varphi} + i\frac{1}{2}\Big|(X-iY)\downarrow\Big\rangle e^{i\frac{3}{2}\varphi} \\
|5\rangle &= -i\frac{1}{2\sqrt{3}}\Big|-(X-iY)\downarrow +2Z\uparrow\Big\rangle e^{-i\frac{1}{2}\varphi} - \frac{1}{2\sqrt{3}}\Big|(X-iY)\uparrow +2Z\downarrow\Big\rangle e^{i\frac{1}{2}\varphi} \\
|6\rangle &= i\frac{1}{\sqrt{6}}\Big|(X+iY)\downarrow +Z\uparrow\Big\rangle e^{-i\frac{1}{2}\varphi} - \frac{1}{\sqrt{6}}\Big|(X-iY)\uparrow -Z\downarrow\Big\rangle e^{i\frac{1}{2}\varphi}.
\end{aligned} \tag{11.12}$$

11.2.2 Wurtzite Hamiltonian of the (0001) Orientation

11.2.2.1 6 × 6 Hamiltonian for the Valence Band

The *c*-plane Hamiltonian for the valence band of the (0001)-oriented wurtzite semiconductor in $\{|1\rangle, |2\rangle, |3\rangle, |4\rangle, |5\rangle, |6\rangle\}$ bases is given by [4,6]

$$H(\vec{k}, \bar{\bar{\varepsilon}}) = \begin{pmatrix} F & -K^* & -H^* & 0 & 0 & 0 \\ -K & G & H & 0 & 0 & \Delta \\ -H & H* & \lambda & 0 & \Delta & 0 \\ 0 & 0 & 0 & F & -K & H \\ 0 & 0 & \Delta & -K* & G & -H^* \\ 0 & \Delta & 0 & H^* & -H & \lambda \end{pmatrix}, \tag{11.13}$$

where

$$\begin{aligned}
F &= \Delta_1 + \Delta_2 + \lambda + \vartheta, \\
G &= \Delta_1 - \Delta_2 + \lambda + \vartheta, \\
\lambda &= \frac{\hbar^2}{2m_0}\left[A_1 k_z^2 + A_2\left(k_x^2 + k_y^2\right)\right] + D_1\varepsilon_{zz} + D_2\left(\varepsilon_{xx} + \varepsilon_{yy}\right),
\end{aligned}$$

$$\begin{aligned}
\vartheta &= \frac{\hbar^2}{2m_0}\left[A_3 k_z^2 + A_4\left(k_x^2 + k_y^2\right)\right] + D_3\varepsilon_{zz} + D_4\left(\varepsilon_{xx} + \varepsilon_{yy}\right),\\
K &= \frac{\hbar^2}{2m_0}A_5\left(k_x + ik_y\right)^2 + D_5\left(\varepsilon_{xx} - \varepsilon_{yy}\right),\\
H &= \frac{\hbar^2}{2m_0}A_6\left(k_x + ik_y\right)k_z + D_6\left(\varepsilon_{xz} + i\varepsilon_{yz}\right),\\
\Delta &= \sqrt{2}\Delta_3,
\end{aligned} \tag{11.14}$$

and the A_i's are the valence-band effective-mass parameters analogous to the Luttinger parameters for the zinc-blende semiconductors, the D_i's are the deformation potentials for wurtzite semiconductors, Δ_1 is the crystal-field split energy, and Δ_2 and Δ_3 are the spin–orbit interaction energies. Under the cubic approximation [6,10], the following relations hold for the parameters A_i's and D_i's:

$$A_1 - A_2 = -A_3 = 2A_4, \quad A_3 + 4A_5 = \sqrt{2}A_6, \quad D_1 - D_2 = -D_3 = 2D_4, \quad D_3 + 4D_5 = \sqrt{2}D_6. \tag{11.15}$$

Here, we restricted ourselves to the biaxial strain case for simplicity, namely,

$$\varepsilon_{xx} = \varepsilon_{yy} \neq \varepsilon_{zz}, \quad \varepsilon_{ij} = 0 \quad \text{for} \quad i \neq j. \tag{11.16}$$

For the case of a strained-layer semiconductor pseudomorphically grown on a (0001)-oriented substrate,

$$\varepsilon_{xx} = \varepsilon_{yy} = \frac{a_{\mathrm{s}}^{\mathrm{w}} - a_{\mathrm{l}}^{\mathrm{w}}}{a_{\mathrm{l}}^{\mathrm{w}}}, \quad \varepsilon_{zz} = -\frac{2C_{13}}{C_{33}}\varepsilon_{xx}, \tag{11.17}$$

where $a_{\mathrm{s}}^{\mathrm{w}}$ and $a_{\mathrm{l}}^{\mathrm{w}}$ are the lattice constants of the substrate (s) and the layer (l) material, and C_{13} and C_{33} are the stiffness constants for the wurtzite structure. The bases for the Hamiltonian are

$$\begin{aligned}
\left|u_1\right\rangle &= -\frac{1}{\sqrt{2}}\left|(X + iY)\uparrow\right\rangle,\\
\left|u_2\right\rangle &= \frac{1}{\sqrt{2}}\left|(X - iY)\uparrow\right\rangle,\\
\left|u_3\right\rangle &= \left|Z\uparrow\right\rangle,\\
\left|u_4\right\rangle &= \frac{1}{\sqrt{2}}\left|(X - iY)\downarrow\right\rangle,\\
\left|u_5\right\rangle &= -\frac{1}{\sqrt{2}}\left|(X + iY)\downarrow\right\rangle,\\
\left|u_6\right\rangle &= \left|Z\downarrow\right\rangle.
\end{aligned} \tag{11.18}$$

11.2.2.2 Block-Diagonalized 3 × 3 Hamiltonian

When we define the angle φ by

$$k_x + ik_x = k_{||}e^{i\varphi}, \tag{11.19}$$

we can write

$$K = K_t e^{2i\varphi}, \quad H = H_t e^{i\varphi}, \quad K_t = \left(\frac{\hbar^2}{2m_0}\right) A_5 k_{||}^2, \quad H_t = \left(\frac{\hbar^2}{2m_0}\right) A_6 k_{||} k_z. \tag{11.20}$$

Then, the 6×6 Hamiltonian can be block-diagonalized into two 3×3 Hamiltonians following a similar procedure to that of the zinc-blende structure by using the transformation matrix U [4,6,7]

$$H' = UHU^{+} = \begin{pmatrix} F & K_t & -iH_t & 0 & 0 & 0 \\ K_t & G & \Delta - iH_t & 0 & 0 & 0 \\ iH_t & \Delta + iH_t & \lambda & 0 & 0 & 0 \\ 0 & 0 & 0 & F & K_t & iH_t \\ 0 & 0 & 0 & K_t & G & \Delta + iH_t \\ 0 & 0 & 0 & -iH_t & \Delta - iH_t & \lambda \end{pmatrix}, \tag{11.21}$$

and

$$U = \begin{pmatrix} \alpha & 0 & 0 & \alpha^* & 0 & 0 \\ 0 & \beta^* & 0 & 0 & \beta & 0 \\ 0 & 0 & \beta & 0 & 0 & \beta^* \\ \alpha & 0 & 0 & -\alpha^* & 0 & 0 \\ 0 & \beta^* & 0 & 0 & -\beta & 0 \\ 0 & 0 & -\beta & 0 & 0 & \beta^* \end{pmatrix} \tag{11.22}$$

where

$$\alpha = \frac{1}{\sqrt{2}} e^{i\left(\frac{3}{4}\pi + \frac{3}{2}\varphi\right)}, \quad \beta = \frac{1}{\sqrt{2}} e^{i\left(\frac{1}{4}\pi + \frac{1}{2}\varphi\right)}. \tag{11.23}$$

Also, a new basis set is given by using the basis transformation, $|i\rangle = \sum T_{ij} |u_j\rangle$, where $T = U^*$. That is,

$$\begin{aligned}
|1\rangle &= -\frac{1}{\sqrt{2}}\left|(X+iY)\uparrow\right\rangle \frac{1}{\sqrt{2}} e^{i\left(\frac{3}{4}\pi + \frac{3}{2}\varphi\right)} + \frac{1}{\sqrt{2}}\left|(X-iY)\downarrow\right\rangle \frac{1}{\sqrt{2}} e^{-i\left(\frac{3}{4}\pi + \frac{3}{2}\varphi\right)} \\
|2\rangle &= \frac{1}{\sqrt{2}}\left|(X-iY)\uparrow\right\rangle \frac{1}{\sqrt{2}} e^{-i\left(\frac{1}{4}\pi + \frac{1}{2}\varphi\right)} - \frac{1}{\sqrt{2}}\left|(X+iY)\downarrow\right\rangle \frac{1}{\sqrt{2}} e^{i\left(\frac{1}{4}\pi + \frac{1}{2}\varphi\right)} \\
|3\rangle &= \left|Z\uparrow\right\rangle \frac{1}{\sqrt{2}} e^{i\left(\frac{1}{4}\pi + \frac{1}{2}\varphi\right)} + \left|Z\downarrow\right\rangle e^{i\frac{1}{2}\varphi} \frac{1}{\sqrt{2}} e^{-i\left(\frac{1}{4}\pi + \frac{1}{2}\varphi\right)} \\
|4\rangle &= -\frac{1}{\sqrt{2}}\left|(X+iY)\uparrow\right\rangle \frac{1}{\sqrt{2}} e^{i\left(\frac{3}{4}\pi + \frac{3}{2}\varphi\right)} - \frac{1}{\sqrt{2}}\left|(X-iY)\downarrow\right\rangle \frac{1}{\sqrt{2}} e^{-i\left(\frac{3}{4}\pi + \frac{3}{2}\varphi\right)} \\
|5\rangle &= \frac{1}{\sqrt{2}}\left|(X-iY)\uparrow\right\rangle \frac{1}{\sqrt{2}} e^{-i\left(\frac{1}{4}\pi + \frac{1}{2}\varphi\right)} + \frac{1}{\sqrt{2}}\left|(X+iY)\downarrow\right\rangle \frac{1}{\sqrt{2}} e^{i\left(\frac{1}{4}\pi + \frac{1}{2}\varphi\right)} \\
|6\rangle &= -\left|Z\uparrow\right\rangle \frac{1}{\sqrt{2}} e^{i\left(\frac{1}{4}\pi + \frac{1}{2}\varphi\right)} + \left|Z\downarrow\right\rangle e^{i\frac{1}{2}\varphi} \frac{1}{\sqrt{2}} e^{-i\left(\frac{1}{4}\pi + \frac{1}{2}\varphi\right)}.
\end{aligned} \tag{11.24}$$

11.3 Band Structures of Strained-Layer QW

11.3.1 Zinc-Blende Semiconductor

Here, we consider a strained-layer QW structure, assuming that the growth direction is along the z-axis and the strain caused by lattice mismatch is entirely elastically accommodated in the QW.

11.3.1.1 Conduction Band

For the unstrained QW, the effective mass theory for the conduction band is obtained from the dispersion relation

$$E(k) = \frac{\hbar^2 k^2}{2m_e}, \tag{11.25}$$

where the effective mass of the electron in the conduction band is $m_e = m_b$ in the barrier region and $m_e = m_w$ in the QW. The potential for the unstrained QW is given by

$$V_c(z) = \begin{cases} \Delta E_c & |z| > \frac{L_w}{2} \\ 0 & |z| \le \frac{L_w}{2} \end{cases}, \tag{11.26}$$

where ΔE_c is the conduction-band offset. The energies are all measured from the conduction-band edge of unstrained QW. For a strained QW, the effective mass equation for a single band is

$$\left[-\frac{\hbar^2}{2m_e^c}\frac{\partial^2}{\partial z^2} + \frac{\hbar^2}{2m_e^c}\nabla_t^2 + V(z) + a_c(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz})\right]\psi(\vec{r}) = E\psi(\vec{r}), \tag{11.27}$$

where m_e^c is the effective mass in the conduction band for zinc-blende structure and a_c is the conduction-band deformation potential. In general, the wave function $\psi(\vec{r})$ can be written

$$\psi^{c\eta}(\vec{r}) = \frac{e^{i(k_x x + k_y y)}}{\sqrt{A}} f(z)|S, \eta > \tag{11.28}$$

and

$$\left[-\frac{\hbar^2}{2m_e^c}\frac{\partial^2}{\partial z^2} + V(z) + a_c(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz})\right] f(z) = \left(E(k_t) - \frac{\hbar^2 k_t^2}{2m_e^c}\right) f(z), \tag{11.29}$$

where η is the electron spin and $|S >$ is the basis function near the zone center in the conduction band. The eigenvalues and eigenfunctions in the conduction band are obtained from the earlier equation. When we ignore the k_t dependence of $f(z)$, Equation 11.29 is solved at $k_t = 0$ for the nth sub-band energy $E_n(0)$ and we have $E_n(k_t) = E_n(0) + \hbar^2 k_t^2/2m^*$.

11.3.1.2 Valence Band

For the unstrained QW, a QW potential is given by

$$V_h(z) = \begin{cases} -\Delta E_v & |z| > \frac{L_w}{2} \\ 0 & |z| \le \frac{L_w}{2} \end{cases}, \tag{11.30}$$

where ΔE_v is the valence-band offset. The eigenvalues and eigenfunctions for the strained QW can be obtained by solving the effective mass equation for a QW potential $V_\mathrm{h}(z)$ given earlier:

$$\left[H^\mathrm{LK}\left(k_x, k_y, k_z = -i\frac{\partial}{\partial z}\right) + V_\mathrm{h}(z)\right] \cdot \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \\ F_6 \end{bmatrix} = E \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \\ F_6 \end{bmatrix}, \tag{11.31}$$

where H^LK is the 6×6 Hamiltonian given by Equation 11.1 and the envelope functions F_1, F_2, F_3, F_4, F_5, and F_6 can be written in the vector form

$$\vec{F}_{\vec{k}}(\vec{r}\,) = \begin{bmatrix} F_1 \\ F_2 \\ F_3 \\ F_4 \\ F_5 \\ F_6 \end{bmatrix} = \begin{bmatrix} g_{3/2,3/2}(k_x, k_y, z) \\ g_{3/2,1/2}(k_x, k_y, z) \\ g_{3/2,-1/2}(k_x, k_y, z) \\ g_{3/2,03/2}(k_x, k_y, z) \\ g_{1/2,1/2}(k_x, k_y, z) \\ g_{1/2,-1/2}(k_x, k_y, z) \end{bmatrix} \frac{e^{i(k_x x + k_y y)}}{\sqrt{A}}. \tag{11.32}$$

The wave function in component form is expressed as

$$\begin{aligned} \psi_{\vec{k}}(\vec{r}) &= F_1\left|\frac{3}{2},\frac{3}{2}\right\rangle + F_2\left|\frac{3}{2},\frac{1}{2}\right\rangle + F_3\left|\frac{3}{2},-\frac{1}{2}\right\rangle + F_4\left|\frac{3}{2},-\frac{3}{2}\right\rangle + F_5\left|\frac{1}{2},\frac{1}{2}\right\rangle + F_6\left|\frac{1}{2},-\frac{1}{2}\right\rangle \\ &= \frac{e^{i(k_x x + k_y y)}}{\sqrt{A}} \sum_{\upsilon} g_{3/2,\upsilon}(k_x, k_y, z)\left|\frac{3}{2},\upsilon\right\rangle, \end{aligned} \tag{11.33}$$

where $\upsilon = \frac{3}{2}, \frac{1}{2}, -\frac{1}{2}, -\frac{3}{2}, \frac{1}{2}$, and $-\frac{1}{2}$. Denoting $\vec{k}_t = k_x\hat{x} + k_y\hat{y}$, we can write

$$\left[H^\mathrm{LK}\left(\vec{k}_t, k_z = -i\frac{\partial}{\partial z}\right) + V_\mathrm{h}(z)\right] \cdot \begin{bmatrix} g_{3/2,3/2}(\vec{k}_t, z) \\ g_{3/2,1/2}(\vec{k}_t, z) \\ g_{3/2,-1/2}(\vec{k}_t, z) \\ g_{3/2,03/2}(\vec{k}_t, z) \\ g_{1/2,1/2}(\vec{k}_t, z) \\ g_{1/2,-1/2}(\vec{k}_t, z) \end{bmatrix} = E(\vec{k}_t) \begin{bmatrix} g_{3/2,3/2}(\vec{k}_t, z) \\ g_{3/2,1/2}(\vec{k}_t, z) \\ g_{3/2,-1/2}(\vec{k}_t, z) \\ g_{3/2,03/2}(\vec{k}_t, z) \\ g_{1/2,1/2}(\vec{k}_t, z) \\ g_{1/2,-1/2}(\vec{k}_t, z) \end{bmatrix}. \tag{11.34}$$

However, it is convenient to solve eigenvalues and eigenfunctions with a block-diagonalized 3×3 Hamiltonian for a numerical calculation. Let us look at the upper 3×3 Hamiltonian in Equation 11.9. The wave functions can be written as

$$\psi^U(\vec{k}_t, \vec{r}) = \frac{e^{i\vec{k}_t \cdot \vec{r}_t}}{\sqrt{A}} \left[g^{(1)}(k_z, z)\,|1\rangle + g^{(2)}(k_z, z)\,|2\rangle + g^{(3)}(k_z, z)\,|3\rangle\right]. \tag{11.35}$$

The wave function satisfies the upper Hamiltonian equation

$$-\begin{pmatrix} P+Q & -R_\rho - iS_\rho & -\sqrt{2}R_\rho + i\sqrt{\frac{1}{2}}S_\rho \\ -R_\rho + iS_\rho & P-Q & \sqrt{2}Q + i\sqrt{\frac{3}{2}}S_\rho \\ -\sqrt{2}R_\rho - i\sqrt{\frac{1}{2}}S_\rho & \sqrt{2}Q - i\sqrt{\frac{3}{2}}S_\rho & P+\Delta \end{pmatrix}\begin{bmatrix} g^{(1)}(\vec{k}_t, z) \\ g^{(2)}(\vec{k}_t, z) \\ g^{(3)}(\vec{k}_t, z) \end{bmatrix}$$

$$+\begin{pmatrix} V_\mathrm{h}(z) & 0 & 0_\rho \\ 0 & V_\mathrm{h}(z) & 0 \\ 0 & 0 & V_\mathrm{h}(z) \end{pmatrix}\begin{bmatrix} g^{(1)}(\vec{k}_t, z) \\ g^{(2)}(\vec{k}_t, z) \\ g^{(3)}(\vec{k}_t, z) \end{bmatrix} = E(\vec{k}_t)\begin{bmatrix} g^{(1)}(\vec{k}_t, z) \\ g^{(2)}(\vec{k}_t, z) \\ g^{(3)}(\vec{k}_t, z) \end{bmatrix}. \tag{11.36}$$

A similar procedure holds for the lower 3 × 3 Hamiltonian. That is, the wave function is given by

$$\psi^L(\vec{k}_t, \vec{r}) = \frac{e^{i\vec{k}_t \cdot \vec{r}_t}}{\sqrt{A}}\left[g^{(4)}(k_z, z)\,|\rangle + g^{(5)}(k_z, z)\,|\rangle + g^{(6)}(k_z, z)\,|\rangle\right], \tag{11.37}$$

which satisfies the following 3 × 3 Hamiltonian equation

$$-\begin{pmatrix} P+Q & -R_\rho + iS_\rho & -\sqrt{2}R_\rho - i\sqrt{\frac{1}{2}}S_\rho \\ -R_\rho - iS_\rho & P-Q & \sqrt{2}Q - i\sqrt{\frac{3}{2}}S_\rho \\ -\sqrt{2}R_\rho + i\sqrt{\frac{1}{2}}S_\rho & \sqrt{2}Q + i\sqrt{\frac{3}{2}}S_\rho & P+\Delta \end{pmatrix}\begin{bmatrix} g^{(4)}(\vec{k}_t, z) \\ g^{(5)}(\vec{k}_t, z) \\ g^{(6)}(\vec{k}_t, z) \end{bmatrix}$$

$$+\begin{pmatrix} V_\mathrm{h}(z) & 0 & 0_\rho \\ 0 & V_\mathrm{h}(z) & 0 \\ 0 & 0 & V_\mathrm{h}(z) \end{pmatrix}\begin{bmatrix} g^{(1)}(\vec{k}_t, z) \\ g^{(2)}(\vec{k}_t, z) \\ g^{(3)}(\vec{k}_t, z) \end{bmatrix} = E(\vec{k}_t)\begin{bmatrix} g^{(4)}(\vec{k}_t, z) \\ g^{(5)}(\vec{k}_t, z) \\ g^{(6)}(\vec{k}_t, z) \end{bmatrix}. \tag{11.38}$$

11.3.2 Wurtzite Semiconductor

We consider nitride-based semiconductors as an example of wurtzite materials. We assume that a strained-layer semiconductor is pseudomorphically grown along the (0001) direction (*c*-axis) on a (0001)-oriented substrate [16]. The wurtzite structure differs from the zinc-blende structure in several aspects. First, in the (0001) wurtzite GaN-based QW structures, there exists a large internal field due to the strain-induced piezoelectric (PZ) and spontaneous (SP) polarizations [17–21]. Second, there are energy splittings in the valence band such as a crystal field splitting and the splitting due to spin–orbit interaction [6].

11.3.2.1 Internal Field

For binary AB compounds such as a GaN wurtzite structure, the sequence of the atomic layers of the constituents A and B is reversed along the [0001] and [000$\bar{1}$] directions. Defining the $+z$ direction with a vector pointing from a Ga atom to the nearest neighbor N atom, the Ga face means that Ga is on the top position of the {0001} bilayer, corresponding to the [0001] polarity. The induced piezoelectric polarization is given by [20]

$$P_z^{\mathrm{PZ}} = e_{31}(\varepsilon_{xx} + \varepsilon_{yy}) + e_{33}\varepsilon_{zz} = 2\left(e_{31} - \frac{C_{13}}{C_{33}}e_{33}\right)\varepsilon_{xx}, \tag{11.39}$$

where e_{ij} are piezoelectric constants. In the case of AlGaN system, for example, the piezoelectric polarization is negative for tensile strain and positive for compressive strain because $2\left(e_{31} - \frac{C_{13}}{C_{33}}e_{33}\right) < 0$ for the whole range of Al compositions. Also, there exists a spontaneous polarization along the c-axis. That is, the spontaneous polarization for GaN (or AlN) has been found to be negative, meaning that the polarization for Ga-face (or Al-face) is pointing toward the substrate. As a result, in the AlGaN system, the alignment of the piezoelectric polarization and the spontaneous polarization is parallel in the case of tensile strain and antiparallel in the case of compressive strain [18]. In the case of multiple QW (MQW) structure, the internal field F_z is determined from the periodic boundary condition and the difference between the sum of spontaneous and piezoelectric polarizations of the well and barrier layers. The continuity of the displacement vector (D_z) normal to the surface gives

$$D_z = \varepsilon_{\mathrm{w}} F_{\mathrm{w}} + P_{\mathrm{w}} = \varepsilon_{\mathrm{b}} F_{\mathrm{b}} + P_{\mathrm{b}}, \tag{11.40}$$

where the subscripts w and b mean the well region and the barrier region, respectively, $P_{\mathrm{w}} = P_{\mathrm{w}}^{\mathrm{PZ}} + P_{\mathrm{w}}^{\mathrm{SP}}$, and $P_{\mathrm{b}} = P_{\mathrm{b}}^{\mathrm{PZ}} + P_{\mathrm{b}}^{\mathrm{SP}}$. We consider the periodic boundary condition that the net voltage drop over one period is zero. That is,

$$F_{\mathrm{w}} L_{\mathrm{w}} + F_{\mathrm{b}} L_{\mathrm{b}} = 0, \tag{11.41}$$

where L_{w} and L_{b} are the well width and the barrier width, respectively. Then, we find the electric fields in the well and the barrier:

$$\begin{aligned} F_{\mathrm{w}} &= \frac{L_{\mathrm{b}}}{\varepsilon_{\mathrm{b}} L_{\mathrm{w}} + \varepsilon_{\mathrm{w}} L_{\mathrm{b}}}(P_{\mathrm{b}} - P_{\mathrm{w}}) \\ F_{\mathrm{b}} &= \frac{-L_{\mathrm{w}}}{\varepsilon_{\mathrm{b}} L_{\mathrm{w}} + \varepsilon_{\mathrm{w}} L_{\mathrm{b}}}(P_{\mathrm{b}} - P_{\mathrm{w}}). \end{aligned} \tag{11.42}$$

Hence, the self-consistent solution, which solves the Schrödinger equation and Poisson equation simultaneously, is necessary for the wurtzite QWs.

11.3.2.2 Energy Splitting

At the zone center ($k_x = k_y = k_z = 0$), we can obtain the following doubly degenerate band-edge energies

$$\begin{aligned} E_1 &= \Delta_1 + \Delta_2 \\ E_2 &= \frac{\Delta_1 - \Delta_2}{2} + \sqrt{\left(\frac{\Delta_1 - \Delta_2}{2}\right)^2 + 2\Delta_3^2} \\ E_3 &= \frac{\Delta_1 - \Delta_2}{2} - \sqrt{\left(\frac{\Delta_1 - \Delta_2}{2}\right)^2 + 2\Delta_3^2}. \end{aligned} \tag{11.43}$$

Without the spin–orbit interaction, $\Delta_2 = \Delta_3 = 0$, we have the top two degenerate bands and the lower band as the reference level E_{v}. That is,

$$\begin{aligned} E_1 &= E_2 = E_{\mathrm{v}} + \Delta_1 = \Delta_1 \\ E_3 &= E_v = 0. \end{aligned} \tag{11.44}$$

Here, we set the reference energy E_v to 0. Then, with the spin–orbit interaction, the top valence-band energy is $E_1 = \Delta_1 + \Delta_2$, and the conduction-band edge is given by adding the bandgap energy E_g to E_v:

$$E_c = E_g + \Delta_1 + \Delta_2. \tag{11.45}$$

11.3.2.3 Self-Consistent Calculations with the Screening of Eigenvalues and Eigenfunctions in the Conduction Band and the Valence Band

The total potential profiles for the electrons and the holes are

$$V_c(z) = \begin{cases} E_c + \Delta E_c + |e|\, F_b z - |e|\, \varphi(z) & |z| > \frac{L_w}{2} \\ E_c + |e|\, F_w z - |e|\, \varphi(z) & |z| \le \frac{L_w}{2} \end{cases} \tag{11.46}$$

and

$$V_h(z) = \begin{cases} -\Delta E_v + |e|\, F_b z - |e|\, \varphi(z) & |z| > \frac{L_w}{2} \\ |e|\, F_w z - |e|\, \varphi(z) & |z| \le \frac{L_w}{2} \end{cases}. \tag{11.47}$$

Here, $\varphi(z)$ is the screening potential induced by the charged carriers and satisfies the Poisson equation

$$\frac{d}{dz}\left(\varepsilon(z)\frac{d}{dz}\right)\varphi(z) = -|e|[p(z) - n(z)], \tag{11.48}$$

where $\varepsilon(z)$ is the dielectric constant and we assume that there is no doping in the well and the barrier. The electron and the hole concentrations, $n(z)$ and $p(z)$, are related to the wave functions of the nth conduction sub-band and the mth valence sub-band by

$$n(z) = \frac{kTm_e}{\pi\hbar^2}\sum_n |f_n(z)|^2 \ln\left(1 + e^{[E_{fc}-E_{cn}]/kT}\right) \tag{11.49}$$

and

$$p(z) = \sum_{\sigma=U,L}\sum_m \int dk_{||}\frac{k_{||}}{2\pi}\sum_\nu |g^{\sigma(\nu)}_{mk_{||}}(z)|^2 \frac{1}{1 + e^{[E_{fv}-E_{vm}(k_{||})]/kT}}, \tag{11.50}$$

where n and m are the quantized sub-band indices for the conduction and the valence bands, E_{fc} and E_{fv} are the quasi-Fermi levels of the electrons and the holes, respectively, and E_{cn} and $f_n(z)$ are the quantized energy level of the electrons at a band-edge and eigenfunctions, respectively. The eigenvalues and eigenfunctions in the conduction band are obtained by solving

$$\left[-\frac{\hbar^2}{2m^w_{ez}}\frac{\partial^2}{\partial z^2} + V_c(z) + a^w_{ct}(\varepsilon_{xx} + \varepsilon_{yy}) + a^w_{cz}\varepsilon_{zz}\right] f(z) = \left(E(k_{||}) - \frac{\hbar^2 k_{||}^2}{2m^w_{et}}\right) f(z), \tag{11.51}$$

where a^w_{ct} and a^w_{cz} are the conduction-band deformation potentials along the c-axis and perpendicular to the c axis, respectively. Usually, it is convenient to obtain eigenvalues in the conduction band with $E_c = 0$ in Equation 11.46. The term $(E_g + \Delta_1 + \Delta_2)$ can be added when we calculate optical properties such as spontaneous emission coefficient and optical gain.

Also, $E_{\mathrm{vm}}(k_{||})$ is the energy for the *m*th sub-band in the valence band, σ denotes the upper (U) and the lower (L) blocks of the Hamiltonian, $k_{||}$ is the in-plane wave vector, ν refers to the new bases for the Hamiltonian, and $g^{\sigma(\nu)}_{\mathrm{mk}_{||}}(z)$ is the envelope function in the valence band.

The eigenvalues and eigenfunctions for the upper Hamiltonian equation of the valence band are obtained by solving

$$-\begin{pmatrix} F & K_t & -iH_t \\ K_t & G & \Delta - iH_t \\ iH_t & \Delta + iH_t & P + \Delta \end{pmatrix} \begin{bmatrix} g^{(1)}(\vec{k}_t, z) \\ g^{(2)}(\vec{k}_t, z) \\ g^{(3)}(\vec{k}_t, z) \end{bmatrix}$$

$$+\begin{pmatrix} V_{\mathrm{h}}(z) & 0 & 0_{\rho} \\ 0 & V_{\mathrm{h}}(z) & 0 \\ 0 & 0 & V_{\mathrm{h}}(z) \end{pmatrix} \begin{bmatrix} g^{(1)}(\vec{k}_t, z) \\ g^{(2)}(\vec{k}_t, z) \\ g^{(3)}(\vec{k}_t, z) \end{bmatrix} = E(\vec{k}_t) \begin{bmatrix} g^{(1)}(\vec{k}_t, z) \\ g^{(2)}(\vec{k}_t, z) \\ g^{(3)}(\vec{k}_t, z) \end{bmatrix}. \tag{11.52}$$

A similar procedure holds for the lower 3×3 Hamiltonian, which satisfies the following 3×3 Hamiltonian equation

$$-\begin{pmatrix} F & K_t & iH_t \\ K_t & G & \Delta + iH_t \\ -iH_t & \Delta - iH_t & P + \Delta \end{pmatrix} \begin{bmatrix} g^{(4)}(\vec{k}_t, z) \\ g^{(5)}(\vec{k}_t, z) \\ g^{(6)}(\vec{k}_t, z) \end{bmatrix}$$

$$+\begin{pmatrix} V_{\mathrm{h}}(z) & 0 & 0_{\rho} \\ 0 & V_{\mathrm{h}}(z) & 0 \\ 0 & 0 & V_{\mathrm{h}}(z) \end{pmatrix} \begin{bmatrix} g^{(1)}(\vec{k}_t, z) \\ g^{(2)}(\vec{k}_t, z) \\ g^{(3)}(\vec{k}_t, z) \end{bmatrix} = E(\vec{k}_t) \begin{bmatrix} g^{(4)}(\vec{k}_t, z) \\ g^{(5)}(\vec{k}_t, z) \\ g^{(6)}(\vec{k}_t, z) \end{bmatrix}. \tag{11.53}$$

The potential $\varphi(z)$ is obtained by integration

$$\varphi(z) = -\int_{-L/2}^{z} E(z')\mathrm{d}z', \tag{11.54}$$

where

$$E(z) = \int_{-L/2}^{z} \frac{1}{\varepsilon(z)} \rho(z')\mathrm{d}z'. \tag{11.55}$$

The procedures for the self-consistent calculations consist of the following steps:

1. Start with the potential profiles V_{c} and V_{h} with $\varphi^{(0)}(z) = 0$ in Equations 11.46 and 11.47.
2. Solve the Schrödinger equation (for electrons) and the block-diagonalized Hamiltonian (for holes) with the potential profiles $\varphi^{(n-1)}(z)$ in step (1) to obtain band structures and wave functions.
3. For a given carrier density, obtain the Fermi energies from Equations 11.49 and 11.50 by using the band structures and the charge distribution by using the wave functions.
4. Solve Poisson's equation to find $\varphi^{(n)}(z)$.

   Check if $\varphi^{(n)}(z)$ converges to $\varphi^{(n-1)}(z)$. If not, set $\varphi^{(n)}(z) = w\varphi^{(n)}(z) + (1-w)\varphi^{(n-1)}(z)$, $n = n+1$; then, return to step (2). If yes, the band structures and the wave functions obtained with $\varphi^{(n-1)}(z)$ are solutions. An adjustable parameter w ($0 < w < 1$) is typically set to be 0.5 at low carrier densities. With increasing carrier densities, a smaller value of w is needed for rapid convergence.

11.4 Optical Matrix Elements

11.4.1 Zinc-Blende Structure

The optical momentum matrix elements are defined as

$$|\hat{\varepsilon}\cdot M_{lm}^{\eta\sigma}(k_{||})|^2 = |<\psi_l^{\eta}|\hat{\varepsilon}\cdot\vec{p}|\psi_m^{\nu\sigma}>|^2, \tag{11.56}$$

which represents the interband transition probability between electrons and hole. Using the expressions given in Equations 11.28 and 11.35 and taking the φ integration of the momentum matrix elements, we obtain the following momentum matrix elements for the upper Hamiltonian.

Transverse electric (TE) polarization ($\hat{\varepsilon}=\hat{x}$ or $\hat{\varepsilon}=\hat{y}$):

$$\begin{aligned}
|M_{\mathrm{TE}}|^2 &= |<\psi_l^{\eta}|p_x|\psi_m^{\nu\sigma}>|^2 = |<\psi_l^{\eta}|p_y|\psi_m^{\nu\sigma}>|^2\\
&= \delta_{\vec{k'}_t,\vec{k}_t} <f_n|g^{(1)}> <S\uparrow|p_x|1> + <f_n|g^{(2)}><S\uparrow|p_x|2> + <f_n|g^{(3)}><S\uparrow|p_x|3>|^2\\
&= \delta_{\vec{k'}_t,\vec{k}_t} |<S\uparrow|p_x|X>|^2| >f_n|g^{(1)}> \left(-\frac{1}{2}e^{-i\frac{3}{2}\varphi}\right)\\
&\quad + <f_n|g^{(2)}> \left(-\frac{1}{2\sqrt{3}}e^{i\frac{1}{2}\varphi}\right) + <f_n|g^{(3)}> \left(-\frac{1}{\sqrt{6}}e^{i\frac{1}{2}\varphi}\right)|^2\\
&= \delta_{\vec{k'}_t,\vec{k}_t} |<S\uparrow|p_x|X>|^2 \frac{1}{12}\\
&\quad\times\left\{\begin{array}{l} 3<f_n|g^{(1)}>^2 + [<f_n|g^{(2)}> + \sqrt{2}<f_n|g^{(3)}>]^2\\ +2\sqrt{3}\cos 2\phi <f_n|g^{(1)}><f_n|g^{(2)}> + 2\sqrt{6}\cos 2\phi <f_n|g^{(1)}><f_n|g^{(3)} \end{array}\right\}.
\end{aligned} \tag{11.57}$$

When we calculate the absorption coefficient, the summation over k_t is given by

$$\frac{1}{V}\sum_{k_t} = \frac{1}{L_z}\int_0^{\infty}\frac{k_t}{2\pi}\mathrm{d}k_t\int_0^{2\pi}\frac{\mathrm{d}\varphi}{2\pi}. \tag{11.58}$$

Thus, the term containing the $\cos 2\varphi$ factor does not contribute to the absorption coefficient because the integration over φ vanishes. We obtain

$$|M_{\mathrm{TE}}|^2 = \delta_{\vec{k'}_t,\vec{k}_t}|<S\uparrow|p_x|X>|^2\frac{1}{12}\left\{3<f_n|g^{(1)}>^2 + [<f_n|g^{(2)}> + \sqrt{2}<f_n|g^{(3)}>]^2\right\}. \tag{11.59}$$

Transverse magnetic (TM) polarization ($\hat{\varepsilon}=\hat{z}$):

$$\begin{aligned}
|M_{\mathrm{TM}}|^2 &= |<\psi_l^{\eta}|p_z|\psi_m^{\nu\sigma}>|^2\\
&= \delta_{\vec{k'}_t,\vec{k}_t}|<f_n|g^{(1)}><S\uparrow|p_z|1> + <f_n|g^{(2)}><S\uparrow|p_z|2> + <f_n|g^{(3)}><S\uparrow|p_z|3>|^2\\
&= \delta_{\vec{k'}_t,\vec{k}_t}|<S\uparrow|p_z|Z>|^2|<f_n|g^{(2)}> i\frac{1}{\sqrt{3}}e^{-i\frac{1}{2}\varphi} - <f_n|g^{(3)}> i\frac{1}{\sqrt{6}}e^{-i\frac{1}{2}\varphi}|^2\\
&= \delta_{\vec{k'}_t,\vec{k}_t}|<S\uparrow|p_z|Z>|^2\frac{1}{3}\left|<f_n|g^{(2)}> - <f_n|g^{(3)}>\frac{1}{\sqrt{2}}\right|.
\end{aligned} \tag{11.60}$$

Here,

$$|< S\uparrow|p_x|X>|^2 = |< S\uparrow|p_z|Z>|^2 = \frac{m_o}{2}\left(\frac{m_o}{m_e}-1\right)\frac{(E_g+\Delta)E_g}{(E_g+2\Delta/3)}. \tag{11.61}$$

The momentum matrix elements for the lower Hamiltonian can be also obtained by a similar procedure:

$$|M_{\mathrm{TE}}|^2 = \delta_{\vec{k}'_t,\vec{k}_t}|< S\downarrow|p_x|X>|^2\frac{1}{12}\left\{3 < f_n|g^{(6)}>^2 + [< f_n|g^{(4)}> + \sqrt{2} < f_n|g^{(5)}>]^2\right\} \tag{11.62}$$

and

$$|M_{\mathrm{TM}}|^2 = \delta_{\vec{k}'_t,\vec{k}_t}|< S\downarrow|p_z|Z>|^2\frac{1}{3}\left|< f_n|g^{(5)}> - < f_n|g^{(4)}> \frac{1}{\sqrt{2}}\right||^2. \tag{11.63}$$

11.4.2 Wurtzite Structure

Similarly to the zinc-blende case, we can obtain the optical momentum matrix elements as follows.

TE polarization ($\widehat{\varepsilon} = \hat{x}$ or $\widehat{\varepsilon} = \hat{y}$):

$$\begin{aligned}|M_{\mathrm{TE}}|^2 &= |<\psi_l^{\eta}|p_x|\psi_m^{\nu\sigma}>|^2 = |<\psi_l^{\eta}|p_y|\psi_m^{\nu\sigma}>|^2 \\ &= \delta_{\vec{k'}_t,\vec{k}_t}|<f_n|g^{(1)}>< S\uparrow|p_x|1> + < f_n|g^{(2)}>< S\uparrow|p_x|2> + < f_n|g^{(3)}>< S\uparrow|p_x|3>|^2 \\ &= \delta_{\vec{k'}_t,\vec{k}_t}|< S\uparrow|p_x|X>|^2\left|-\frac{1}{2} < f_n|g^{(1)}> e^{i\left(\frac{3}{4}\pi+\frac{3}{2}\varphi\right)} + < f_n|g^{(2)}> \frac{1}{2}e^{-i\left(\frac{1}{4}\pi+\frac{1}{2}\varphi\right)}\right|^2 \\ &= \delta_{\vec{k'}_t,\vec{k}_t}|< S\uparrow|p_x|X>|^2\frac{1}{4}\left|< f_n|g^{(1)}>^2 + < f_n|g^{(2)}>^2 + 2\cos\varphi < f_n|g^{(1)}>< f_n|g^{(2)}>\right|.\end{aligned} \tag{11.64}$$

Finally,

$$|M_{TE}|^2 = |\delta_{\vec{k}'_t,\vec{k}_t}|< S\uparrow|p_x|X>|^2\frac{1}{4}\left\{< f_n|g^{(1)}>^2 + < f_n|g^{(2)}>^2\right\}. \tag{11.65}$$

TM polarization ($\widehat{\varepsilon} = \hat{z}$):

$$\begin{aligned}|M_{\mathrm{TM}}|^2 &= |<\psi_l^{\eta}|p_z|\psi_m^{\nu\sigma}>|^2 \\ &= \delta_{\vec{k'}_t,\vec{k}_t}|<f_n|g^{(1)}>< S\uparrow|p_z|1> + < f_n|g^{(2)}>< S\uparrow|p_z|2> + < f_n|g^{(3)}>< S\uparrow|p_z|3>|^2 \\ &= \delta_{\vec{k'}_t,\vec{k}_t} < S\uparrow|p_z|Z>^2 |<f_n|g^{(3)}> \frac{1}{\sqrt{2}}e^{i\left(\frac{1}{4}\pi+\frac{1}{2}\varphi\right)}|^2 \\ &= \delta_{\vec{k'}_t,\vec{k}_t} < S\uparrow|p_z|Z>^2 \frac{1}{2} < f_n|g^{(3)}>^2 .\end{aligned} \tag{11.66}$$

The momentum matrix elements for the lower Hamiltonian also can be obtained by a similar procedure:

$$|M_{\mathrm{TE}}|^2 = |\delta_{\vec{k'}_t,\vec{k}_t}|< S\downarrow|p_x|X>|^2\frac{1}{4}\left\{< f_n|g^{(4)}>^2 + < f_n|g^{(5)}>^2\right\} \tag{11.67}$$

and

$$|M_{\mathrm{TM}}|^2 = \delta_{\vec{k'}_t,\vec{k}_t} < S\downarrow|p_z|Z>^2 \frac{1}{2} < f_n|g^{(6)}>^2 . \tag{11.68}$$

11.5 Band Structures of Bulk Semiconductors with an Arbitrary Crystal Orientation

11.5.1 Zinc-Blende Structure

The Hamiltonian for an arbitrary crystal orientation can be obtained using a rotation matrix

$$U = \begin{pmatrix} \cos\theta\cos\varphi & \cos\theta\sin\varphi & -\sin\theta \\ -\sin\varphi & \cos\varphi & 0 \\ \sin\theta\cos\varphi & \sin\theta\sin\varphi & \cos\theta \end{pmatrix}, \tag{11.69}$$

The rotation with the Euler angles θ and φ transforms physical quantities from the (x, y, z) coordinates to the (x', y', z') coordinates. Figure 11.1 shows configuration of the $[lmn]$-oriented coordinate system to the conventional [001]-oriented coordinate system. The relation between the coordinate systems for vectors and tensors is expressed as

$$\begin{aligned} k_\alpha &= U_{i\alpha}k'_i, \\ \varepsilon_{\alpha\beta} &= U_{i\alpha}U_{j\beta}\varepsilon'_{ij}, \end{aligned} \tag{11.70}$$

where the summation over repeated indices is assumed. Our object is to obtain $P_{k'}$, $Q_{k'}$, $R_{k'}$, and $S_{k'}$ in the (x', y', z') coordinates. First we obtain the 4×4 Hamiltonian for general crystal orientation by using invariant method and then extending these results to obtain the 6×6 Hamiltonian. Here, for simplicity, we consider the case of the QWs with $(11n)$ orientations with $\theta = \arctan(\sqrt{2}/n)$ and $\varphi = \pi/4$. For example, (110) means $\theta = \pi/2$ and $\varphi = \pi/4$. The 4×4 Hamiltonian to describe the interaction between the heavy-hole and light-hole bands along with the Bir–Pikus Hamiltonian for strain can be formally written as

$$H_t^{\mathrm{v}} = H_0^{\mathrm{v}}(\vec{k}) + H_{\mathrm{s}}^{\mathrm{v}}, \tag{11.71}$$

with $H_0^{\mathrm{v}}(\vec{k})$ representing the valence-band Hamiltonian and $H_{\mathrm{s}}^{\mathrm{v}}$ the valence-band strain or Bir–Pikus Hamiltonian. The Luttinger formulation of the most general Hamiltonian for the (001) crystal orientation

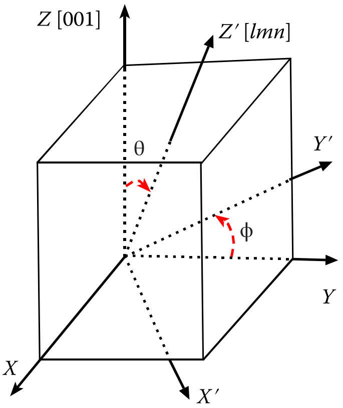


FIGURE 11.1 Configuration of the $[lmn]$-oriented coordinate system to the conventional [001]-oriented coordinate system.

is given by [22,23]

$$
\begin{aligned}
H_0^{\mathrm{v}} &= \frac{\hbar^2}{2m_{\mathrm{o}}}(\gamma_1 k^2 I_4 - 2\gamma_2[(J_x^2 - 1/3J^2)k_x^2 + (J_y^2 - 1/3J^2)k_y^2 + (J_z^2 - 1/3J^2)k_z^2] \\
&\quad - 4\gamma_3[\{J_xJ_y\}\{k_xk_y\} + \{J_yJ_z\}\{k_yk_z\} + \{J_zJ_x\}\{k_zk_x\}]), \\
H_{\mathrm{s}}^{\mathrm{v}} &= -a_v(\varepsilon_{xx} + \varepsilon_{yy} + \varepsilon_{zz})I_4 + b[(J_x^2 - 1/3J^2)\varepsilon_{xx} + (J_y^2 - 1/3J^2)\varepsilon_{yy} + (J_z^2 - 1/3J^2)\varepsilon_{zz}] \\
&\quad + 2d/\sqrt{3}[\{J_xJ_y\}\{\varepsilon_{xx}\} + \{J_yJ_z\}\{\varepsilon_{yy}\} + \{J_zJ_x\}\{\varepsilon_{zz}\}]),
\end{aligned}
\tag{11.72}
$$

where the J_i's are the angular momentum matrices for a state with spin 3/2; I_4 is a 4×4 identity matrix; and $\{k_xk_y\} = (k_xk_y + k_yk_x)/2$. The angular momentum matrices are any three matrices, $(J_xJ_yJ_z)$, which satisfy the rules of commutation for angular momentum. These matrices given by Luttinger are as follows:

$$
\begin{aligned}
J_x &= \begin{pmatrix} 0 & 0 & \sqrt{3}/2 & 0 \\ 0 & 0 & 1 & \sqrt{3}/2 \\ \sqrt{3}/2 & 1 & 0 & 0 \\ 0 & \sqrt{3}/2 & 0 & 0 \end{pmatrix}, \quad
J_y = \begin{pmatrix} 0 & 0 & -i\sqrt{3}/2 & 0 \\ 0 & 0 & i & -i\sqrt{3}/2 \\ i\sqrt{3}/2 & -i & 0 & 0 \\ 0 & i\sqrt{3}/2 & 0 & 0 \end{pmatrix}, \\
J_z &= \begin{pmatrix} 3/2 & 0 & 0 & 0 \\ 0 & -1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 0 \\ 0 & 0 & 0 & -3/2 \end{pmatrix}, \quad
I_4 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}.
\end{aligned}
\tag{11.73}
$$

We can obtain the (001)-oriented Hamiltonian by substituting Equation 11.73 into 11.72:

$$
H_0^{\mathrm{v}} = \begin{pmatrix} P+Q & R & -S & 0 \\ R^* & P-Q & 0 & S^* \\ -S^* & 0 & P-Q & R \\ 0 & S^* & R^* & P+Q \end{pmatrix}.
\tag{11.74}
$$

This is exactly the same from as the upper 4×4 Hamiltonian given in Equation 11.1. However, note that, in this case, the ordering of the basis functions for the choice of the J_i's is $\left|\frac{3}{2}, \frac{3}{2}\right\rangle, \left|\frac{3}{2}, -\frac{1}{2}\right\rangle, \left|\frac{3}{2}, \frac{1}{2}\right\rangle, \left|\frac{3}{2}, -\frac{3}{2}\right\rangle$. Next, we consider the case of QWs with $(11n)$ orientations with $\theta = \arctan(\sqrt{2}/n)$ and $\varphi = \pi/4$. Then, from Equations 11.69 and 11.70,

$$
\begin{aligned}
P_{k'} &= \left(\frac{\hbar^2}{2m_0}\right)\gamma_1(k_{x'}^2 + k_{y'}^2 + k_{z'}^2), \\
Q_{k'} &= \left(\frac{\hbar^2}{2m_0}\right)\left\{\left[\frac{(n^2-4)(n^2-1)}{(n^2+2)^2}\gamma_2 + \frac{9n^2}{(n^2+2)^2}\gamma_3\right]k_{x'}^2 + \left[\frac{(n^2+2)(n^2-1)}{(n^2+2)^2}\gamma_2 + \frac{3}{(n^2+2)}\gamma_3\right]k_{y'}^2\right. \\
&\quad \left. + \left[-2\frac{(n^2-1)^2}{(n^2+2)^2}\gamma_2 - \frac{6+12n^2}{(n^2+2)}\gamma_3\right]k_{z'}^2\right\}, \\
R_{k'} &= \left(\frac{\hbar^2}{2m_0}\right)\left\{\sqrt{3}\left[\frac{(n^2-4)}{(n^2+2)^2}\gamma_2 - \frac{5n^2+n^4}{(n^2+2)^2}\gamma_3\right]k_{x'}^2 + \sqrt{3}\left[\frac{1}{(n^2+2)}\gamma_2 + \frac{1+n^2}{(n^2+2)^2}\gamma_3\right]k_{y'}^2\right.
\end{aligned}
$$

$$
\begin{aligned}
&+ 2\sqrt{3}\left[-\frac{(n^2-1)}{(n^2+2)^2}(\gamma_2-\gamma_3)\right] k_{z'}^2 + i2\sqrt{3}\left[\frac{(n^2\gamma_2+2\gamma_3)}{(n^2+2)}\right] k_{x'}k_{y'} \\
&+ 2\sqrt{6}\left[i\frac{n}{(n^2+2)}(\gamma_2-\gamma_3)\right] k_{y'}k_{z'} + 6\sqrt{6}\left[\frac{n}{(n^2+2)^2}(\gamma_2-\gamma_3)\right] k_{x'}k_{z'} \Bigg\},
\end{aligned}
$$

$$
\begin{aligned}
-S_{k'} &= \left(\frac{\hbar^2}{2m_0}\right)\Bigg\{\left[\sqrt{6}\frac{n(4-n^2)}{(n^2+2)^2}(\gamma_2-\gamma_3)\right] k_{x'}^2 + \left[\sqrt{6}\frac{n}{(n^2+2)}(\gamma_3-\gamma_2)\right] k_{y'}^2 \\
&+ \left[2\sqrt{6}\frac{n(n^2-1)}{(n^2+2)^2}(\gamma_2-\gamma_3)\right] k_{z'}^2 + i2\sqrt{6}\left[\frac{n}{(n^2+2)}(\gamma_2-\gamma_3)\right] k_{x'}k_{y'} \\
&+ i2\sqrt{3}\left[\frac{2\gamma_2+n^2\gamma_3}{(n^2+2)}\right] k_{y'}k_{z'} + \sqrt{3}\left[-\frac{12n^2}{(n^2+2)^2}\gamma_2 - 2\frac{(4-2n^2+n^4)}{(n^2+2)^2}\gamma_3\right] k_{x'}k_{z'} \Bigg\},
\end{aligned}
$$

$$
\begin{aligned}
P_{\varepsilon'} &= -a_{\mathrm{v1}}(\varepsilon_{x'x'} + \varepsilon_{y'y'} + \varepsilon_{z'z'}), \\
Q_{\varepsilon'} &= -\frac{b(n^2-1)(-\Gamma + 3\sqrt{2}\varepsilon_{x'z'}n + \Gamma n^2) + d\sqrt{3}[\Gamma + \sqrt{2}\varepsilon_{x'z'}n + n^2(2\Gamma - \sqrt{2}\varepsilon_{x'z'}n)]}{(n^2+2)^2}, \\
R_{\varepsilon'} &= -\frac{b\sqrt{3}(-\Gamma + 3\sqrt{2}\varepsilon_{x'z'}n + \Gamma n^2) + d(\Gamma - 3\sqrt{2}\varepsilon_{x'z'}n - \Gamma n^2)]}{(n^2+2)^2}, \\
-S_{\varepsilon'} &= -\frac{b\sqrt{3}n(\sqrt{2}\Gamma - 6\varepsilon_{x'z'}n - \sqrt{2}\Gamma n^2) + d[\sqrt{2}n(n^2-1)\Gamma + \varepsilon_{x'z'}(-4+2n^2-n^4)]}{(n^2+2)^2},
\end{aligned}
\tag{11.75}
$$

where $\Gamma = \varepsilon_{||} - \varepsilon_{z'z'}$, $\varepsilon_{||} = (a_{\mathrm{s}} - a_{\mathrm{e}})/a_{\mathrm{e}}$, and a_{s} and a_{e} are lattice constants for the substrate and the epilayer materials, respectively. The biaxial strain components for a general crystal orientation are determined from the condition that the layer is grown pseudomorphically and these strain coefficients should minimize the strain energy of the layer simultaneously [24,25]. That is,

$$
\varepsilon_{x'x'} = \varepsilon_{y'y'} = \varepsilon_{||}, \quad \varepsilon_{x'y'} = \varepsilon_{y'z'} = 0, \quad \varepsilon_{z'z'} = -2\frac{K_3}{\sqrt{2}K_2}\varepsilon_{||}, \quad \varepsilon_{x'z'} = -2\frac{K_1}{\sqrt{2}K_2}\varepsilon_{||}, \tag{11.76}
$$

where

$$
\begin{aligned}
K_1 &= (C_{11} + 2C_{12})(-C_{11} + C_{12} + 2C_{44})n(n^2-1), \\
K_2 &= 2C_{11}C_{44} + 2C_{12}C_{44} + 2C_{44}^2 + (C_{11}^2 + C_{11}C_{12} - 2C_{12}^2 + 2C_{11}C_{44} - 4C_{12}C_{44} + C_{11}C_{44}n^2)n^2, \\
K_2 &= -2\left[C_{11}C_{44} + 3C_{12}C_{44} - 2C_{44}^2 + (C_{11}^2 + C_{11}C_{12} - 2C_{12}^2 - C_{11}C_{44} + C_{12}C_{44}n^2)n^2\right],
\end{aligned}
\tag{11.77}
$$

and C_{ij} are the stiffness constants in the strained epilayers. Then, the 4 × 4 Hamiltonian of $(11n)$-oriented zinc-blende crystal is given by

$$
H'_{4\times4} = \begin{pmatrix} P'+Q' & R' & -S' & 0 \\ R'^* & P'-Q' & 0 & S'^* \\ -S'^* & 0 & P'-Q' & R' \\ 0 & S'^* & R'^* & P'+Q' \end{pmatrix}, \tag{11.78}
$$

where

$$
P' = P_{k'} + P_{\varepsilon'}, \quad Q' = Q_{k'} + Q_{\varepsilon'}, \quad R' = R_{k'} + R_{\varepsilon'}, \quad S' = S_{k'} + S_{\varepsilon'}. \tag{11.79}
$$

Off-diagonal strains in zinc-blende structure semiconductors induce a polarization given by [25]

$$P_i^{\mathrm{S}} = 2e_{14}\varepsilon_{jk}, \tag{11.80}$$

where P_i^{S} is the induced polarization, e_{14} is the piezoelectric constant, and $i, j, k = x, y, z$. Here, i, j, k are in cyclic order. A strained layer with a [001] growth direction has only diagonal strains. Thus, the (001)-oriented layer will not have strain-induced piezoelectric polarization fields. However, strained layers with any other growth direction will have piezoelectric polarization fields. The piezoelectric polarization along the growth direction is important because the electric field in the QW originates from polarization charges at heterojunction interfaces. The electric field induced perpendicular to the growth direction due to the strain can be calculated as

$$E_{\perp} = -\frac{\vec{P}^{\mathrm{S}} \cdot \hat{u}}{\varepsilon}, \tag{11.81}$$

where the unit vector $\hat{u}$ along the growth direction is given by $\hat{u} = \sqrt{1/(n^2+2)}(\hat{\imath} + \hat{\jmath} + n\hat{k})$. For an $[11n]$ growth direction,

$$E_{\perp} = -\frac{2e_{14}}{\sqrt{n^2+2}}(n^2\varepsilon_{xy} + \varepsilon_{yz} + \varepsilon_{zx}), \tag{11.82}$$

where

$$\begin{aligned}
\varepsilon_{xy} &= \frac{1}{(n^2+2)}\left(-\varepsilon_{x'x'} + \sqrt{2}n\varepsilon_{x'z'} + 2n\varepsilon_{z'z'}\right), \\
\varepsilon_{yz} &= \frac{1}{2(n^2+2)}\left(-2n\varepsilon_{x'x'} - \sqrt{2}(2-n^2)\varepsilon_{x'z'} + 2n\varepsilon_{z'z'}\right) \\
\varepsilon_{zx} &= \varepsilon_{yz}.
\end{aligned} \tag{11.83}$$

It is straightforward to obtain the 6 × 6 Hamiltonian of $(11n)$-oriented zinc-blende crystal using components P', Q', R', and S' in the (x', y', z') coordinates. That is, the 6×6 Hamiltonian for the valence-band structure in the (x', y', z') coordinates can be written as

$$H'_{6\times6} = -\begin{pmatrix}
P'+Q' & -S' & R' & 0 & -S'/\sqrt{2} & \sqrt{2}R' \\
-S'^{\dagger} & P'-Q' & 0 & R' & -\sqrt{2}Q' & \sqrt{3/2}S' \\
R'^{\dagger} & 0 & P'-Q' & 0 & \sqrt{3/2}S'^{\dagger} & \sqrt{2}Q' \\
0 & R'^{+} & S'^{\dagger} & P'+Q' & -\sqrt{2}R'^{\dagger} & -S'^{\dagger}/\sqrt{2} \\
-S'^{\dagger}/\sqrt{2} & 0 & \sqrt{3/2}S' & -\sqrt{2}R' & P'+\Delta & 0 \\
\sqrt{2}R'^{\dagger} & \sqrt{3/2}S'^{\dagger} & \sqrt{2}Q' & -S'/\sqrt{2} & 0 & P'+\Delta
\end{pmatrix}
\begin{matrix}
|3/2, 3/2\rangle' \\
|3/2, 1/2\rangle' \\
|3/2, -1/2\rangle' \\
|3/2, -3/2\rangle' \\
|1/2, 1/2\rangle' \\
|1/2, -1/2\rangle'
\end{matrix}. \tag{11.84}$$

Here, the basis set in the (x', y', z') coordinates consists of the following basis functions:

$$|u_1\rangle' = \left|\frac{3}{2}, \frac{3}{2}\right\rangle' = -\frac{1}{\sqrt{2}}\left|(X' + iY')\uparrow'\right\rangle,$$

$$|u_2\rangle' = \left|\frac{3}{2}, \frac{1}{2}\right\rangle' = \frac{1}{\sqrt{6}}\left|-(X' - iY')\downarrow' + 2Z'\uparrow'\right\rangle,$$

$$|u_3\rangle' = \left|\frac{3}{2}, -\frac{1}{2}\right\rangle' = \frac{1}{\sqrt{6}}\left|(X' - iY')\uparrow' + 2Z'\downarrow'\right\rangle,$$

$$|u_4\rangle' = \left|\frac{3}{2}, -\frac{3}{2}\right\rangle' = \frac{1}{\sqrt{2}}\left|(X' - iY') \downarrow'\right\rangle, \tag{11.85}$$

$$|u_5\rangle' = \left|\frac{1}{2}, \frac{1}{2}\right\rangle' = \frac{1}{\sqrt{3}}\left|(X' + iY') \downarrow' + Z' \uparrow'\right\rangle,$$

$$|u_6\rangle' = \left|\frac{1}{2}, -\frac{1}{2}\right\rangle' = \frac{1}{\sqrt{3}}\left|(X' - iY') \uparrow' - Z' \downarrow'\right\rangle.$$

11.5.2 Wurtzite Structure

Analytical expressions for strain components for general crystal orientation can be obtained from the condition that the layer is grown pseudomorphically and the condition that strain energy density should be minimal. Figure 11.2 shows (a) configuration of the coordinate system (x', y', z') in $(hkil)$-oriented crystals and (b) a wurtzite primitive cell. The z-axis corresponds to the c-axis [0001] and the growth axis or the z'-axis is normal to the QW plane $(hkil)$. We define the unit vectors $\hat{x}', \hat{y}'$ and $\hat{z}'$ along the $\vec{x}'-, \vec{y}'$-, and $\vec{z}'$-axes and they are related to unit vectors $\hat{x}$, $\hat{y}$, and $\hat{z}\hat{z}'$ along the $\vec{x}$-, $\vec{y}$-, and $\vec{z}$-axes of the original crystal orientation (0001) by the rotation matrix Equation 11.69. We define the hexagonal primitive translational vectors as

$$\begin{aligned}\vec{\alpha}_i &= a_i \hat{x},\\ \vec{\beta}_i &= -\frac{a_i}{2}\hat{x} + \frac{\sqrt{3}a_i}{2}\hat{y},\\ \vec{\gamma}_i &= c_i \hat{z},\end{aligned} \tag{11.86}$$

where a_i and c_i are lattice constants of the hexagonal structure with the subscript i denoting the epilayer (e) and substrate (s), respectively. When the crystal is strained the primitive translation vectors become

$$\begin{aligned}\vec{\alpha}''_i &= a_i \hat{x}'',\\ \vec{\beta}''_i &= -\frac{a_i}{2}\hat{x}'' + \frac{\sqrt{3}a_i}{2}\hat{y}'',\\ \vec{\gamma}''_i &= c_i \hat{z}'',\end{aligned} \tag{11.87}$$

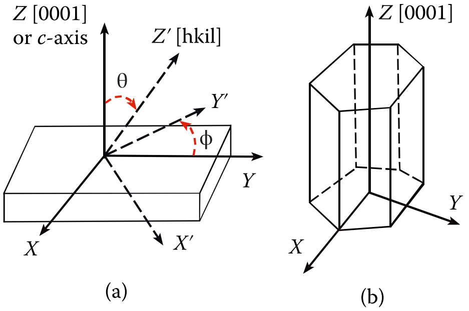


FIGURE 11.2 (a) Configuration of the coordinate system (x', y', z') in $(hkil)$-oriented crystals and (b) a wurtzite primitive cell.

with

$$\begin{aligned}\hat{x}'' &= (1+\varepsilon_{xx})\hat{x} + \varepsilon_{xy}\hat{y} + \varepsilon_{xz}\hat{z},\\ \hat{y}'' &= \varepsilon_{yx}\hat{x} + (1+\varepsilon_{yy})\hat{y} + \varepsilon_{yz}\hat{z},\\ \hat{z}'' &= \varepsilon_{zx}\hat{x} + \varepsilon_{zy}\hat{y} + (1+\varepsilon_{zz})\hat{z}.\end{aligned} \tag{11.88}$$

When a pseudomorphic interface is formed during the epitaxial growth, the translation vectors of each strained layer must have the same projection onto the growth plane, which is so-called the lattice commensurability constraint. By applying the lattice commensurability constraint on the QW and after some mathematical manipulations, we obtain the following analytical expressions for the strain tensors:

$$\begin{aligned}\varepsilon_{xx} &= \varepsilon_{xx}^{(0)} + \frac{\sin\theta\cos\phi}{\cos\theta}\varepsilon_{xz},\\ \varepsilon_{xy} &= \frac{\sin\theta\sin\phi}{\cos\theta}\varepsilon_{xz},\\ \varepsilon_{yy} &= \varepsilon_{xx}^{(0)} + \frac{\sin\theta\sin^2\phi}{\cos\theta\cos\phi}\varepsilon_{xz},\\ \varepsilon_{yz} &= \frac{\sin\phi}{\cos\phi}\varepsilon_{xz},\\ \varepsilon_{zz} &= \varepsilon_{zz}^{(0)} + \frac{\cos\theta}{\sin\theta\cos\phi}\varepsilon_{xz},\end{aligned} \tag{11.89}$$

where $\varepsilon_{xx}^{(0)} = \frac{a_s - a_e}{a_e}$ and $\varepsilon_{zz}^{(0)} = \frac{c_s - c_e}{c_e}$. The strain tensor ε_{xz} can be evaluated by minimizing the strain energy density which is given by

$$W = \frac{1}{2}\left[C_{11}\varepsilon_{xx}^2 + C_{11}\varepsilon_{yy}^2 + C_{33}\varepsilon_{12}^2 + 2C_{12}\varepsilon_{xx}\varepsilon_{yy} + 2C_{13}(\varepsilon_{xx}\varepsilon_{zz} + \varepsilon_{yy}\varepsilon_{zz}) + 4C_{44}\varepsilon_{zz}^2\right]. \tag{11.90}$$

By minimizing the strain energy density with respect to ε_{xz}, we get

$$\varepsilon_{xz} = -\frac{N(\theta,\phi)\sin\theta\cos\theta}{D(\theta,\phi)\cos\phi}\varepsilon_{xx}^{(0)}, \tag{11.91}$$

where

$$N(\theta,\phi) = \sin^2\theta\left(C_{11} + C_{12} + C_{13}\frac{\varepsilon_{zz}^{(0)}}{\varepsilon_{xx}^{(0)}}\right) + \cos^2\theta\left(2C_{13} + C_{33}\frac{\varepsilon_{zz}^{(0)}}{\varepsilon_{xx}^{(0)}}\right), \tag{11.92}$$

and

$$D(\theta,\phi) = \sin^4\theta\left(C_{11}\frac{\sin^4\phi + \cos^4\phi}{\cos^2\phi} + 2C_{12}\sin^2\phi\right) + C_{33}\frac{\cos^4\theta}{\cos^2\phi} + 2\left(\frac{C_{13}}{\cos^2\phi} + 2C_{44}\right)\sin^2\theta\cos^2\theta. \tag{11.93}$$

Using Equations 11.70 and 11.89, general expressions for the Hamiltonian for the valence-band structure in (x', y', z') coordinates can be written as

$$\tilde{H}'(\vec{k}', \bar{\bar{\varepsilon}}') = \begin{pmatrix} F' & -K'^* & -H'^* & 0 & 0 & 0 \\ -K' & G' & H' & 0 & 0 & \Delta \\ -H' & H'^* & \lambda' & 0 & \Delta & 0 \\ 0 & 0 & 0 & F' & -K' & H' \\ 0 & 0 & \Delta & -K'^* & G' & -H'^* \\ 0 & \Delta & 0 & H'^* & -H' & \lambda' \end{pmatrix}, \tag{11.94}$$

where

$$\begin{aligned}
F' &= \Delta_1 + \Delta_2 + \lambda' + \vartheta', \\
G' &= \Delta_1 - \Delta_2 + \lambda' + \vartheta', \\
\lambda' &= \frac{\hbar^2}{2m_0} A_1 \left(\sin^2\theta k_x'^2 - 2\sin\theta\cos\theta k_x' k_z' + \cos^2\theta k_z'^2\right) \\
&\quad + \frac{\hbar^2}{2m_0} A_2 \left(\cos^2\theta k_x'^2 + s\cos\theta\sin\theta k_x' k_z' + \sin^2\theta k_z'^2 + k_y'^2\right) \\
&\quad + D_1 \left(\varepsilon_{zz}^{(0)} - \frac{N(\theta,\phi)\cos^2\theta}{D(\theta,\phi)\cos^2\phi}\varepsilon_{xx}^{(0)}\right) + D_2 \left(2 - \frac{N(\theta,\phi)\sin^2\theta}{D(\theta,\phi)\cos^2\phi}\right)\varepsilon_{xx}^{(0)}, \\
\vartheta' &= \frac{\hbar^2}{2m_0} A_3 \left(\sin^2\theta k_x'^2 - 2\sin\theta\cos\theta k_x' k_z' + \cos^2\theta k_z'^2\right) \\
&\quad + \frac{\hbar^2}{2m_0} A_4 \left(\cos^2\theta k_x'^2 + s\cos\theta\sin\theta k_x' k_z' + \sin^2\theta k_z'^2 + k_y'^2\right) \\
&\quad + D_3 \left(\varepsilon_{zz}^{(0)} - \frac{N(\theta,\phi)\cos^2\theta}{D(\theta,\phi)\cos^2\phi}\varepsilon_{xx}^{(0)}\right) + D_4 \left(2 - \frac{N(\theta,\phi)\sin^2\theta}{D(\theta,\phi)\cos^2\phi}\right)\varepsilon_{xx}^{(0)}, \\
K' &= \frac{\hbar^2}{2m_0} A_5 e^{2i\phi} \left(\cos^2\theta k_x'^2 - k_y'^2 + \sin^2\theta k_z'^2 + 2\sin\theta\cos\theta k_x' k_y' + 2i\cos\theta k_x' k_y' + 2i\sin\theta k_y' k_z'\right) \\
&\quad - D_5 \frac{N(\theta,\phi)\left(\cos^2\phi - \sin^2\phi - 2i\sin\phi\cos\phi\right)\sin^2\theta}{D(\theta,\phi)\cos^2\phi}\varepsilon_{xx}^{(0)}, \\
H' &= \frac{\hbar^2}{2m_0} A_6 e^{i\phi} \left\{\sin\theta\cos\theta\left(k_z'^2 - k_x'^2\right) + \cos^2\theta k_x' k_z' - \sin^2\theta k_x' k_z' + i\left(\cos\theta k_y' k_z' - \sin\theta k_x' k_y'\right)\right\} \\
&\quad - D_6 e^{i\phi} \frac{N(\theta,\phi)\sin\theta\cos\theta}{D(\theta,\phi)\cos^2\phi}\varepsilon_{xx}^{(0)}.
\end{aligned} \tag{11.95}$$

We note that the old bases $\{|1\rangle, |2\rangle, |3\rangle, |4\rangle, |5\rangle, |6\rangle\}$ are used in the matrix representation of the Hamiltonian in Equation 11.94 and we assumed the hexagonal symmetry in the calculation. The Hamiltonian for nonpolar wurtzite QW structures can be obtained by substituting $\theta = \pi/2$ into the above equations.

11.5.3 Interband Optical-Matrix Elements for QW with an Arbitrary Crystal Orientation

11.5.3.1 (11*n*)-Oriented Zinc-Blende QW

The hole wave function in a QW can be written as

$$\Psi_m(z';k'_t) = \frac{e^{i\vec{k}_t\cdot\vec{r}_t}}{\sqrt{A}} \sum_{\nu=1}^{6} g_m^{(\nu)}(z';k'_t)\,|u_\nu\rangle\,, \tag{11.96}$$

where $g_m^{(\nu)}(z';k'_t)(\nu = 1,2,3,4,5,$ and 6) is the wave function for the *m*th sub-band in (x',y',z') coordinates, $|\rangle\; u_\nu\rangle'$ is given by Equation 11.85, and $g_m'^{(i)}$ follows the normalization rules

$$\sum_{\nu=1}^{6} \int \mathrm{d}z'\; |g_m^{(\nu)}(z';\; k'_t)|^2 = 1. \tag{11.97}$$

Since we solve one Schrödinger equation for electrons and 6 × 6 Hamiltonian for holes, we need only the optical matrix elements for two cases: spin up and spin down. The polarization-dependent interband momentum-matrix elements are written as follows.

TE polarization ($\hat{e}' = \cos\varphi'\hat{x}' + \sin\varphi'\hat{y}'$):

$$\begin{aligned}
\left|\hat{e}'\cdot\vec{M}'^{\uparrow}\right|^2 &= \left|\cos\varphi'\left\{-\frac{1}{\sqrt{2}}P_x\left\langle g_m'^{(1)}|\phi_l\right\rangle + \frac{1}{\sqrt{6}}P_x\left\langle g_m'^{(3)}|\phi_l\right\rangle + \frac{1}{\sqrt{3}}P_x\left\langle g_m'^{(6)}|\phi_l\right\rangle\right\}\right.\\
&\quad\left. + \sin\varphi'\left\{-i\frac{1}{\sqrt{2}}P_x\left\langle g_m'^{(1)}|\phi_l\right\rangle - i\frac{1}{\sqrt{6}}P_x\left\langle g_m'^{(3)}|\phi_l\right\rangle - i\frac{1}{\sqrt{3}}P_x\left\langle g_m'^{(6)}|\phi_l\right\rangle\right\}\right|^2\\
\left|\hat{e}'\cdot\vec{M}'^{\downarrow}\right|^2 &= \left|\cos\varphi'\left\{-\frac{1}{\sqrt{6}}P_x\left\langle g_m'^{(2)}|\phi_l\right\rangle + \frac{1}{\sqrt{2}}P_x\left\langle g_m'^{(4)}|\phi_l\right\rangle + \frac{1}{\sqrt{3}}P_x\left\langle g_m'^{(5)}|\phi_l\right\rangle\right\}\right.\\
&\quad\left. + \sin\varphi'\left\{-i\frac{1}{\sqrt{6}}P_x\left\langle g_m'^{(2)}|\phi_l\right\rangle - i\frac{1}{\sqrt{2}}P_x\left\langle g_m'^{(4)}|\phi_l\right\rangle + i\frac{1}{\sqrt{3}}P_x\left\langle g_m'^{(5)}|\phi_l\right\rangle\right\}\right|^2.
\end{aligned} \tag{11.98}$$

TE polarization ($\hat{e}' = \hat{z}'$):

$$\begin{aligned}
\left|\hat{e}'\vec{M}'^{\uparrow}\right|^2 &= \left|-\frac{2}{\sqrt{6}}P_z\left\langle g_m'^{(2)}|\phi_l\right\rangle + \frac{1}{\sqrt{3}}P_z\left\langle g_m'^{(5)}|\phi_l\right\rangle\right|^2\\
\left|\hat{e}'\cdot\vec{M}'^{\downarrow}\right|^2 &= \left|\frac{2}{\sqrt{6}}P_x\left\langle g_m'^{(3)}|\phi_l\right\rangle - \frac{1}{\sqrt{3}}P_x\left\langle g_m'^{(6)}|\phi_l\right\rangle\right|^2.
\end{aligned} \tag{11.99}$$

11.5.3.2 Wurtzite QW

The optical matrix elements general crystal orientation are given by

$$|\hat{e}'\cdot\vec{M}'^{\eta}|^2 = \left|\left\langle\Phi_l'^{\eta}\right|\hat{e}'\cdot\vec{p}'\;|\rangle\,\Psi'_m\rangle\right|^2. \tag{11.100}$$

The polarization-dependent interband momentum-matrix elements can be written as follows.

TE polarization $\left(\hat{e}' = \cos\varphi'\hat{x} + \sin\varphi'\hat{y}\right)$:

$$
\begin{aligned}
\hat{e}'\cdot\vec{M}'^{\uparrow} = & -\cos\varphi'\sin\theta P_z\left\langle\varphi_l'\,\middle|\,g_m'^{(3)}\right\rangle \\
& -\frac{1}{\sqrt{2}}\cos\varphi'\cos\theta\cos\phi P_x\left\{(1+i)\left\langle\varphi_l'\,\middle|\,g_m'^{(1)}\right\rangle-(1-i)\left\langle\varphi_l'\,\middle|\,g_m'^{(2)}\right\rangle\right\} \\
& +\frac{1}{\sqrt{2}}\sin\varphi' P_x\left\{\sin\phi\left(\left\langle\varphi_l'\,\middle|\,g_m'^{(1)}\right\rangle-\left\langle\varphi_l'\,\middle|\,g_m'^{(2)}\right\rangle\right)-i\cos\phi\left(\left\langle\varphi_l'\,\middle|\,g_m'^{(1)}\right\rangle+\left\langle\varphi_l'\,\middle|\,g_m'^{(2)}\right\rangle\right)\right\}, \\
\hat{e}'\cdot\vec{M}'^{\downarrow} = & -\cos\varphi'\sin\theta P_z\left\langle\varphi_l'\,\middle|\,g_m'^{(6)}\right\rangle \\
& +\frac{1}{\sqrt{2}}\cos\varphi'\cos\theta\cos\phi P_x\left\{(1-i)\left\langle\varphi_l'\,\middle|\,g_m'^{(4)}\right\rangle-(1+i)\left\langle\varphi_l'\,\middle|\,g_m'^{(5)}\right\rangle\right\} \\
& -\frac{1}{\sqrt{2}}\sin\varphi' P_x\left\{\sin\phi\left(\left\langle\varphi_l'\,\middle|\,g_m'^{(4)}\right\rangle-\left\langle\varphi_l'\,\middle|\,g_m'^{(5)}\right\rangle\right)+i\cos\phi\left(\left\langle\varphi_l'\,\middle|\,g_m'^{(4)}\right\rangle+\left\langle\varphi_l'\,\middle|\,g_m'^{(5)}\right\rangle\right)\right\}.
\end{aligned}
\tag{11.101}
$$

TM polarization $\left(\hat{e}' = \hat{z}'\right)$:

$$
\begin{aligned}
\hat{e}'\cdot\vec{M}'^{\uparrow} = & \cos\theta P_z\left\langle\varphi_l'\,\middle|\,g_m'^{(3)}\right\rangle+\frac{1}{\sqrt{2}}\sin\theta\cos\phi\left\{-\left\langle\varphi'_l\,\middle|\,g_m'^{(1)}\right\rangle+\left\langle\varphi_l'\,\middle|\,g_m'^{(2)}\right\rangle\right\} \\
& -\frac{i}{\sqrt{2}}\sin\theta\sin\phi\left\{\left\langle\varphi_l'\,\middle|\,g_m'^{(1)}\right\rangle+\left\langle\varphi_l'\,\middle|\,g_m'^{(2)}\right\rangle\right\}, \\
\hat{e}'\cdot\vec{M}'^{\downarrow} = & \cos\theta P_z\left\langle\varphi_l'\,\middle|\,g_m'^{(6)}\right\rangle+\frac{1}{\sqrt{2}}\sin\theta\cos\phi\left\{\left\langle\varphi_l'\,\middle|\,g_m'^{(4)}\right\rangle-\left\langle\varphi_l'\,\middle|\,g_m'^{(5)}\right\rangle\right\} \\
& -\frac{i}{\sqrt{2}}\sin\theta\sin\phi\left\{\left\langle\varphi_l'\,\middle|\,g_m'^{(4)}\right\rangle+\left\langle\varphi_l'\,\middle|\,g_m'^{(5)}\right\rangle\right\},
\end{aligned}
\tag{11.102}
$$

11.6 Optical Gain Model with Many-Body Effects

The optical gain spectra are calculated using the non-Markovian gain model with many-body effects [11,13]. The many-body effects include the plasma screening, band-gap renormalization (BGR), and the excitonic or the Coulomb enhancement (CE) of the interband transition probability [14]. The simplest non-Markovian quantum kinetics is Gaussian line-shape function, which is connected with memory effects in the system–reservoir interaction [13]. The optical gain with many-body effects including the effects of anisotropy on the valence-band dispersion is given by

$$
g(\omega)=\sqrt{\frac{\mu_{\mathrm{o}}}{\varepsilon}}2\left(\frac{e^2}{m_{\mathrm{o}}^2\omega}\right)\int_0^{2\pi}\mathrm{d}\varphi_0\int_0^{\infty}\mathrm{d}k_{||}\frac{2k_{||}}{(2\pi)^2L_{\mathrm{w}}}\left|M_{nm}(k_{||},\varphi_0)\right|^2\left[f_l^{\mathrm{c}}(k_{||},\varphi_0)-f_m^{\mathrm{v}}(k_{||},\varphi_0)\right]L(\omega,k_{||},\varphi_0),
\tag{11.103}
$$

where ω is the angular frequency, μ_{o} is the vacuum permeability, ε is the dielectric constant, e is the charge of an electron, m_{o} is the free electron mass, $k_{||}$ is the magnitude of the in-plane wave vector in the QW plane, L_w is the well width, $\left|M_{nm}(k_{||},\varphi_0)\right|^2$ is the momentum matrix element in the strained QW, f_l^c and f_m^v are the Fermi functions for occupation probability by the electrons in the conduction sub-band states and the valence sub-band states, respectively, and the indices l and m denote the electron states in conduction sub-band and heavy hole (light hole) sub-band states, respectively. Also, $E_{lm}(\hbar\omega,k_{||},\varphi_0)=E_l^\mathrm{c}(k_{||},\varphi_0)-E_m^\mathrm{v}(k_{||},\varphi_0)+E_\mathrm{g}+\Delta E_\mathrm{SX}+\Delta E_\mathrm{CH}-\hbar\omega$ is the renormalized transition energy between electrons and holes, where E_g is the band gap of the material, and ΔE_SX and ΔE_CH are the screened exchange and the Coulomb-hole contributions [14] to the BGR, respectively. The factor $Q(\hbar\omega,k_{||},\varphi_0)$ accounts for the excitonic or CE of the interband transition probability [14,15]. The Gaussian line-shape function $L(E_{lm}(\hbar\omega,k_{||},\varphi_0)$ renormalized with many-body effects is given by

$$\mathrm{Re}[L(E_{lm}(\hbar\omega,k_{||},\varphi_0))]=\sqrt{\frac{\pi\tau_{in}(\hbar\omega,k_{||},\varphi_0)\tau_\mathrm{c}}{2\hbar^2}}\exp\left(-\frac{\tau_{in}(\hbar\omega,k_{||},\varphi_0)\tau_\mathrm{c}}{2\hbar^2}E_{lm}^2(\hbar\omega,k_{||},\varphi_0)\right)$$

$$\mathrm{Im}[L(E_{lm}(\hbar\omega,k_{||},\varphi_0))]=\frac{\tau_\mathrm{c}}{\hbar}\int_0^\infty \mathrm{d}t\exp\left(-\frac{\tau_\mathrm{c}}{2\tau_\mathrm{in}(\hbar\omega,k_{||},\varphi_0)}t^2\right)\sin\left(\frac{\tau_\mathrm{c}E_{lm}(\hbar\omega,k_{||},\varphi_0)}{\hbar}t\right). \tag{11.104}$$

The correlation time τ_c is related to the non-Markovian enhancement of optical gain [13] and is assumed to be constant. The τ_in and the τ_c used in the calculation are 25 and 10 fs, respectively. The BGR is given as a summation of the Coulomb-hole self-energy and the screened-exchange shift [14]. The φ_0 dependence of the BGR is very small and neglected for simplicity. The Coulomb-hole contribution to the BGR is written as

$$\Delta E_\mathrm{CH}=-2E_\mathrm{R}a_\mathrm{o}\lambda_\mathrm{s}\ln\left(1+\sqrt{\frac{32\pi NL_\mathrm{w}}{C\lambda_\mathrm{s}^3a_\mathrm{o}}}\right), \tag{11.105}$$

where N is the carrier density, λ_s is the inverse screening length, and C is a constant usually taken between 1 and 4. The Rydberg constant E_R and the exciton Bohr radius a_o are given by

$$\Delta E_\mathrm{R}(eV)=13.6\frac{m_\mathrm{o}/m_\mathrm{r}}{(\varepsilon/\varepsilon_0)^2} \tag{11.106}$$

and

$$a_\mathrm{o}(\text{Å})=0.53\frac{\varepsilon/\varepsilon_0}{m_\mathrm{o}/m_\mathrm{r}}, \tag{11.107}$$

where m_r is the reduced electron–hole mass defined by $1/m_\mathrm{r}=1/m_\mathrm{e}+1/m_\mathrm{h}$. The exchange contribution to the BGR is given by

$$\Delta E_\mathrm{SX}=-\frac{2E_\mathrm{R}a_\mathrm{o}}{\lambda_\mathrm{s}}\int_0^\infty \mathrm{d}k_{||}k_{||}\frac{1+\frac{C\lambda_\mathrm{s}a_\mathrm{o}k_{||}^2}{32\pi NL_\mathrm{w}}}{1+\frac{k_{||}}{\lambda_\mathrm{s}}+\frac{Ca_\mathrm{o}k_{||}^3}{32\pi NL_\mathrm{w}}}\left[f_n^\mathrm{c}(k_{||})+1-f_m^\mathrm{v}(k_{||})\right]. \tag{11.108}$$

The factor $1/(1 - Q(k_{||}, \hbar\omega))$ represents the CE in the Padé approximation. Here, the factor $Q(k_{||}, \hbar\omega)$ is given by [11,14]

$$Q(k_{||}, \hbar\omega) = i\frac{E_{\mathrm{R}} a_{\mathrm{o}}}{\pi\lambda_{\mathrm{s}} \left|M_{nm}(k_{||})\right|^2} \int_0^{\infty} \mathrm{d}k'_{||} k'_{||} \left|M_{nm}(k_{||})\right| [f_n^{\mathrm{c}}(k'_{||}) - f_m^{\mathrm{v}}(k'_{||})] \Xi(E_{lm}(\hbar\omega, k_{||})) \Theta(k_{||}, k'_{||}), \quad (11.109)$$

where

$$\Theta(k_{||}, k'_{||}) = \int_0^{\infty} \mathrm{d}\theta \frac{1 + \dfrac{C\lambda_{\mathrm{s}} a_{\mathrm{o}} q_{||}^2}{32\pi N L_{\mathrm{w}}}}{1 + \dfrac{q_{||}}{\lambda_{\mathrm{s}}} + \dfrac{C a_{\mathrm{o}} q_{||}^3}{32\pi N L_{\mathrm{w}}}} \quad (11.110)$$

and

$$q = \left|k_{||} - k'_{||}\right|. \quad (11.111)$$

The spontaneous emission coefficient $g_{\mathrm{SP}}(\omega)$ can be obtained by replacing $[f_l^{\mathrm{c}}(k_{||}, \varphi_0) - f_m^{\mathrm{v}}(k_{||}, \varphi_0)]$ in Equation 11.103 by $f_l^{\mathrm{c}}(k_{||}, \varphi_0)(1 - f_m^{\mathrm{v}}(k_{||}, \varphi_0))$.

$$g_{\mathrm{SP}}(\omega) = \sqrt{\frac{\mu_{\mathrm{o}}}{\varepsilon}} 2 \left(\frac{e^2}{m_{\mathrm{o}}^2 \omega}\right) \int_0^{2\pi} \mathrm{d}\varphi_0 \int_0^{\infty} \mathrm{d}k_{||} \frac{2k_{||}}{(2\pi)^2 L_{\mathrm{w}}} \left|M_{nm}(k_{||}, \varphi_0)\right|^2 f_l^{\mathrm{c}}(k_{||}, \varphi_0)(1 - f_m^{\mathrm{v}}(k_{||}, \varphi_0)) L(\omega, k_{||}, \varphi_0). \quad (11.112)$$

The spontaneous emission rate $r_{\mathrm{spon}}(E)$ can be obtained from our calculated spontaneous emission spectrum $g_{\mathrm{SP}}(E)$ by using

$$r_{\mathrm{spon}}(E) = \left(\frac{4n^2}{\hbar\lambda^2}\right) g_{\mathrm{SP}}(E),$$

where n is the refractive index of the QW and $E = 2\pi c\hbar/\lambda$ with c being the speed of light.

11.7 Numerical Example

As a numerical example, we calculate valence-band structure, optical matrix element, and optical gain as a function of crystal angle for zinc-blende and wurtzite GaN/AlGaN QW structures. For zinc-blende case, we consider the compressively strained GaN well by assuming that QW structures are grown on AlGaN substrate. On the other hand, in the case of the wurtzite semiconductor, we assume that QW structures are grown on GaN substrate. The material parameters for zinc-blende GaN and AlN used in the calculation were taken from Refs. [9,26–32] and references therein. Also, the material parameters for wurtzite GaN and AlN used in the calculation were taken from Refs. [33–35] and references therein. All parameters used in the calculation are summarized in Table 11.1.

11.7.1 Zinc-Blende Structure

Figure 11.3 shows the valence-band structures along $k_{x'}$ and $k_{y'}$ of (a) (001)-, (b) (111)-, and (c) (110)-oriented zinc-blende GaN/Al$_x$Ga$_{1-x}$N QWs ($L_{\mathrm{w}} = 2.5$ nm). The sub-bands are labeled HH$_i$ and LS$_i$, where i denotes the sub-band level. Here SL is the acronym for "splitoff-hole-light-hole," and the first letter denotes the dominant component for the wave function. The crystal angles for QWs with $(11n)$ orientations

TABLE 11.1 Parameters Used in the Calculation

	Wurtzite			Zinc-Blende	
Parameters	GaN	AlN		GaN	AlN
Lattice constant (Å)					
a	3.1892	3.112	a	4.460	4.342
c	5.185	4.982			
Energy parameter					
E_g(eV)	3.44	6.16	E_g(eV)	3.1	4.9
$\Delta_1 = \Delta_{cr}$(meV)	22.0	−58.5	Δ (meV)	11.0	11.0
$\Delta_2 = \Delta/3$ (meV)	15.0	20.4			
$\Delta_3 = \Delta_2$(meV)	5.0	6.8			
Conduction-band effective masses					
m_{ez}^{w}/m_o	0.2	0.3	m_e^c/m_o	0.13	0.21
$m_{et}^{w} = m_{ez}^{w}$					
Valence-band effective-mass parameters					
A_1	−6.56	−3.95	γ_1	3.06	2.42
A_2	−0.91	−0.27	γ_2	0.91	0.58
A_3	−3.13	−1.95	γ_3	1.03	0.71
Deformation potentials					
a_c	−4.6	−4.5	a_c	−2.77	−6.8
D_1	−1.7	−2.89	a_v	3.63	2.3
D_2	6.3	4.89	b	−2.67	−1.5
D_3	−4.0	−3.34	d	−4.62	−4.5
Dielectric constant					
ε	10.0	8.5	ε	10.69	8.5
Elastic stiffness constant					
C_{11}	39.0	39.8	C_{11}	29.6	30.4
C_{12}	14.5	14.0	C_{12}	15.4	15.2
C_{13}	10.6	12.7			
C_{33}	39.8	38.2			
C_{44}	10.5	9.6	C_{44}	20.0	19.9
C_{66}	12.3	12.9			
Piezoelectric constant					
d_{31}($\times10^{-12}$ m/V)	−1.7	−2.0	e_{14}(C/m^2)	−1.11	−0.526
Spontaneous polarization (C/m^2)					
P_{SP}	−0.029	−0.081			

($\varphi = \pi/4$) are obtained by using the relation $\theta = \tan^{-1}(\sqrt{2}/n)$. For example, crystal angles for (111) and (110) are $\theta \approx 55°$ and $\theta \approx 90°$, respectively. The valence-band structures of the self-consistent model are calculated at a carrier density of $N_{2D} = 10 \times 10^2$ cm^{-2}. The strain-induced piezoelectric field in the GaN well for the (111) orientation is about 1.33 MV/cm. The sub-band structures for different orientations are apparently different. For the (001) and the (111) orientations, the sub-band structures are nearly spherically symmetric. On the other hand, the valence-band structure of the (110)-oriented QW shows anisotropy. In particular, the effective mass of the first sub-band along $k_{y'}$ is observed to be greatly reduced for the (110)-oriented structure. A larger energy spacing between the first two sub-bands (HH1 and SL1) and higher sub-bands are observed for the (001)-oriented structure. The increase in the sub-band energy spacing will reduce the carrier population in the higher sub-bands. However, the energy spacing is gradually reduced with increasing polar angle θ.

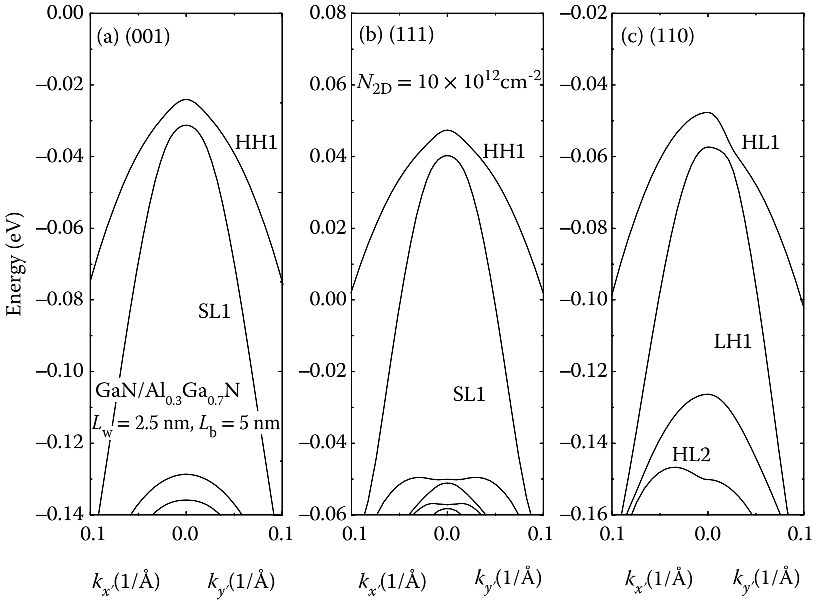


FIGURE 11.3 Valence-band structures along $k_{x'}$ and $k_{y'}$ of (a) (001)-, (b) (111)-, and (c) (110)-oriented zinc-blende GaN/Al$_x$Ga$_{1-x}$N QWs (L_w = 2.5 nm). The sub-bands are labeled HH$_i$ and LS$_i$, where i denotes the sub-band level.

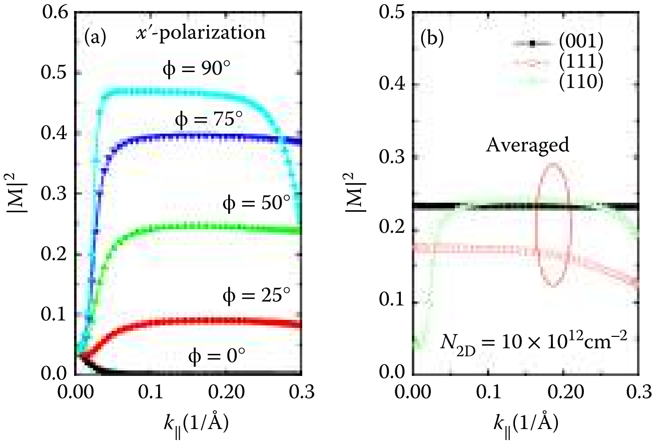


FIGURE 11.4 (a) x'-Polarized normalized optical matrix element as a function of $k_{||}$ for various values of φ for (110)-oriented zinc-blende GaN/Al$_x$Ga$_{1-x}$N quantum wells (QWs) (L_w = 2.5 nm) and (b) averaged optical matrix element as a function of $k_{||}$ for (001)-, (111)-, and (110)-oriented zinc-blende GaN/Al$_x$ Ga$_{1-x}$N QWs (L_w = 2.5 nm).

Figure 11.4 shows the x'-polarized normalized optical matrix element as a function of $k_{||}$ for various values of φ for (110)-oriented zinc-blende GaN/Al$_x$Ga$_{1-x}$N QWs (L_w = 2.5 nm) and (b) the averaged optical matrix element as a function of $k_{||}$ for (001)-, (111)-, and (110)-oriented zinc-blende GaN/Al$_x$Ga$_{1-x}$N QWs (L_w = 2.5 nm). The angle φ is defined by $\tan\phi = k_{y'}/k_{x'}$ and $k_{||} = \sqrt{k_{x'}^2 + k_{y'}^2}$. Here, φ = 0° and 90° mean that optical matrix elements are plotted along $k_{x'}$ and $k_{y'}$. The optical matrix elements of the self-consistent model are calculated at a carrier density of $N_\text{2D} = 10 \times 10^{12}$ cm^{-2}. The optical matrix element increases with increasing angle between $k_{x'}$ and $k_{y'}$ wave vectors, that is, with changing from the x' to y' direction. The optical matrix elements averaged in the $k_{x'} - k_{y'}$ plane show that they significantly depend on the crystal angle. The optical matrix elements of the (001) orientation are nearly independent of the wave vectors. On the other hand, in the case of the (110) orientation, the optical matrix element rapidly increases with increasing wave vectors $k_{||}$ and begins to decrease in a range of large $k_{||}$. The optical matrix elements of the (111)-oriented QW structure slowly decreases with increasing wave vectors $k_{||}$. The (111)-oriented QW structure has a smaller matrix element than the (001)- or the (110)-oriented QW because the spatial separation between the electron and the hole wave functions increases due to the piezoelectric field, which results in a reduction in the transition probability between an electron and a hole.

Figure 11.5 shows x'- and y'-polarized many-body optical gain spectra of (a) (001)-, (b) (111)-, and (c) (110)-oriented zinc-blende GaN/Al$_x$Ga$_{1-x}$N QWs (L_w = 2.5 nm). Optical gains of the self-consistent model are calculated at a carrier density of $N_{2D} = 10 \times 10^{12}$ cm^{-2}. They are also averaged in the $k_{x'} - k_{y'}$ plane because of their anisotropy in the QW plane. The optical gain spectra have peaks corresponding to C1-HH1 and C1-HL1 transitions. The (111)-oriented QW structure shows that y'-polarized optical gain is larger than x'-polarized optical gain. The in-plane optical anisotropy ρ is about −0.17. Here, ρ is defined as $(I_{x'} - I_{y'})/(I_{x'} + I_{y'})$, where $I_i(i = x', y')$ means an intensity of peak optical gain. On the other hand, in the case of the (111)-oriented QW structure, y'-polarized optical gain is shown to be similar to x'-polarized optical gain with $\rho \approx 0$.

For a given carrier density, the (111)-oriented QW structure has a smaller optical gain than the (001)- or the (110)-oriented QW. This can be explained by the fact that the (111)-oriented QW has much smaller matrix elements than the (001)- or the (110)-oriented QW structures, as shown in Figure 11.4b. This is because the spatial separation between the electron and the hole wave functions is increased due to the piezoelectric field, which results in a reduction in the transition probability between electrons and holes.

11.7.2 Wurtzite Structure

Figure 11.6 shows the valence-band structures of wurtzite GaN/Al$_x$Ga$_{1-x}$N QWs (L_w = 2.5 nm) with (a) θ = 0° (*c*-plane), (b) 30°, and (c) 90° (*a*-plane). Here, the naming of the sub-bands follows the dominant composition of the wave function at the Γ point in terms of the $|X'\rangle$, $|Y'\rangle$, and $|Z'\rangle$ bases. The components

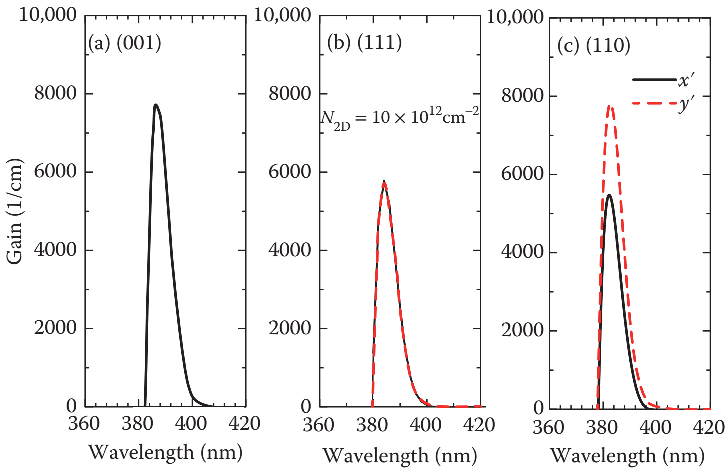


FIGURE 11.5 x'- and y'-polarized many-body optical gain spectra of (a) (001)-, (b) (111)-, and (c) (110)-oriented zinc-blende GaN/Al$_x$Ga$_{1-x}$N quantum wells (QWs) (L_w = 2.5 nm). Optical gains of the self-consistent model are calculated at a carrier density of N_{2D} 10×10^{12} cm^{-2}.

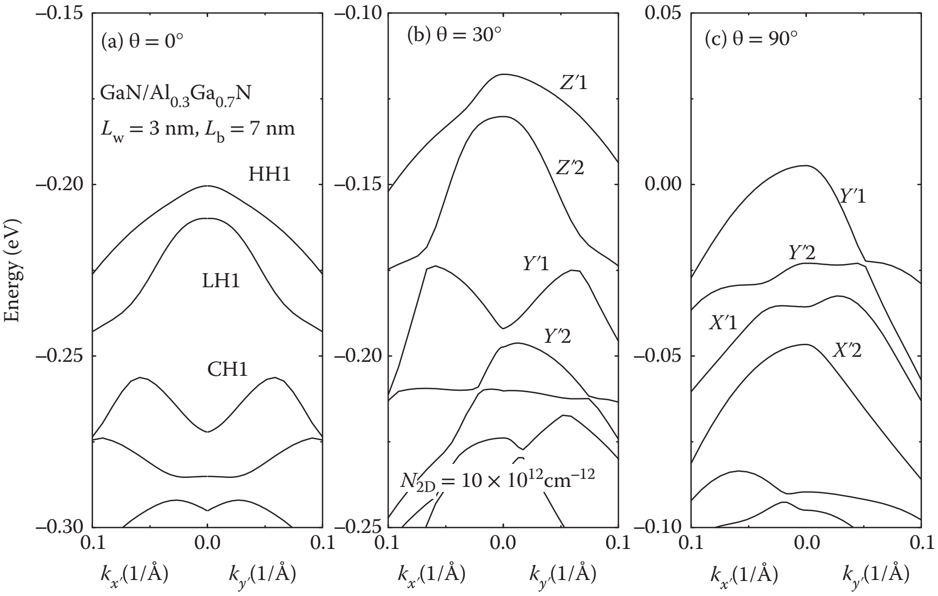


FIGURE 11.6 Valence-band structures of wurtzite GaN/Al$_x$Ga$_{1-x}$N quantum wells (QWs) (L_w = 2.5 nm) with (a) θ = 0° (*c*-plane), (b) 30°, and (c) 90° (*a*-plane).

$P_m^{i(=X',Y',Z')}$ of each wave function are given by

$$P_m^{X'} = \left\langle g_m'^{(3)} \mid g_m'^{(3)} \right\rangle + \left\langle g_m'^{(6)} \mid g_m'^{(6)} \right\rangle$$

$$P_m^{Y'} = \frac{\left\langle g_m'^{(1)} + g_m'^{(2)} \mid g_m'^{(1)} + g_m'^{(2)} \right\rangle + \left\langle g_m'^{(4)} + g_m'^{(5)} \mid g_m'^{(4)} + g_m'^{(5)} \right\rangle}{2}$$

$$P_m^{Z'} = \frac{\left\langle g_m'^{(2)} - g_m'^{(1)} \mid g_m'^{(2)} - g_m'^{(1)} \right\rangle + \left\langle g_m'^{(4)} - g_m'^{(5)} \mid g_m'^{(4)} - g_m'^{(5)} \right\rangle}{2}. \qquad (11.113)$$

The valence-band structures of QW structures with θ = 30° and 90° show anisotropy in the QW plane, unlike the (0001)-oriented structure (*c*-plane). The effective mass of the topmost valence band along $k_{y'}$ is smaller than that along $k_{x'}$. Also, the hole effective mass (1.24 m_0 for *a*-plane) of the topmost valence band along $k_{y'}$ is lower than that (1.56 m_0) of the *c*-plane. However, the hole effective mass (1.39 m_0) averaged in the $k_{x'} - k_{y'}$ QW plane is slightly larger than that along $k_{y'}$. Here, to estimate the magnitude of the hole effective mass, we considered a parabolic band fitted to the lowest sub-band of the exact band structure. The effective mass is determined so that, for a given carrier density and the quasi-Fermi level for holes, the carrier density and the quasi-Fermi level agree with those of the exact band structure. Hence, the effective mass of the fitted parabolic band reflects an averaged density of states.

Figure 11.7 shows the y'-polarized normalized optical matrix element as a function of $k_{||}$ for various values of φ for wurtzite *a*-plane GaN/Al$_x$Ga$_{1-x}$N QWs (L_w = 2.5 nm) and (b) averaged optical matrix element as a function of $k_{||}$ for QW structures with crystal angles of (a) θ = 0° (*c*-plane), (b) 30°, and (c) 90° (*a*-plane). The optical matrix elements of the self-consistent model are calculated at a carrier density of $N_{2D} = 10 \times 10^{12}$ cm^{-2}. The optical matrix element rapidly decreases with increasing $k_{||}$ and greatly depends on angle φ between $k_{x'}$ and $k_{y'}$ wave vectors. In particular, in the case with φ = 90°, the decrease in the optical matrix element is observed to be much more rapid, compared to the other case. The optical

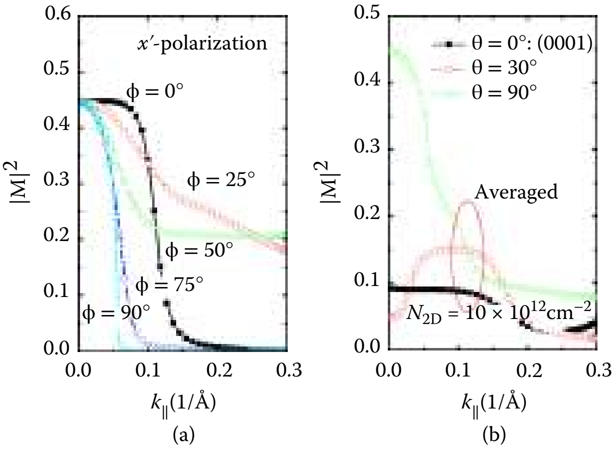


FIGURE 11.7 (a) y'-Polarized normalized optical matrix element as a function of $k_{||}$ for various values of φ for wurtzite *a*-plane GaN/Al$_x$Ga$_{1-x}$N quantum wells (QWs) ($L_w = 2.5$ nm) and (b) averaged optical matrix element as a function of $k_{||}$ for QW structures with crystal angles of $\theta = 0°$ (*c*-plane), 30°, and 90° (*a*-plane).

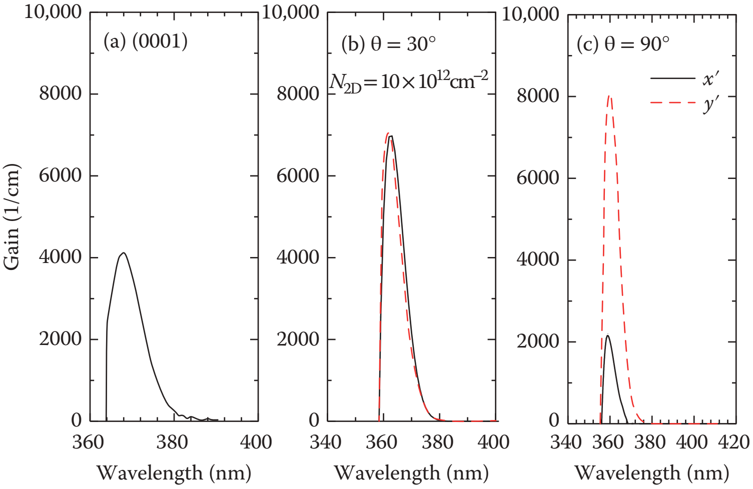


FIGURE 11.8 x'- and y'-polarized, many-body optical gain spectra for quantum well (QW) structures with crystal angles of $\theta = 0°$ (*c*-plane), 30°, and 90° (*a*-plane).

matrix elements averaged in the $k_{x'} - k_{y'}$ plane show that they significantly depend on the crystal angle, as observed for the zinc-blende case of Figure 11.3. The optical matrix element rapidly increases with increasing crystal angle. This can be explained by the fact that the internal field is reduced for the QW structure with larger crystal angle. In particular, in the QW structure with $\theta = 90°$, the internal field is disappeared and the transition probability between an electron and a hole is significantly enhanced due to the reduction in the spatial separation between the electron and the hole wave functions.

Figure 11.8 shows x'- and y'-polarized many-body optical gain spectra for QW structures with crystal angles of (a) $\theta = 0°$ (*c*-plane), (b) 30°, and (c) 90° (*a*-plane). Optical gains of the self-consistent model are calculated at a carrier density of $N_{2D} = 10 \times 10^{12}$ cm^{-2}. They are also averaged in the $k_{x'} - k_{y'}$ plane because of their anisotropy in the QW plane. The (0001)-oriented QW structure shows that y'-polarized optical gain is the same as x'-polarized optical gain because of its isotropy in the QW plane. The QW structure with $\theta = 30°$ also shows that x'- and y'-polarized optical gains are similar to each other. On the other hand, the *a*-plane QW structure shows that y'-polarized optical gain is much larger than x'-polarized optical gain. The in-plane optical anisotropy ρ is about -0.58, which is much larger than that observed for the zinc-blende QW structures.

11.8 Summary

In summary, we reviewed theoretical formalism to obtain eigenvalues and wave functions of (001)-oriented zinc-blende and (0001)-oriented wurtzite QW structures. In addition, we reviewed crystal orientation effects on electronic and optical properties of strained zinc-blende and wurtzite QW structures. As a numerical example, we calculated valence-band structures, optical matrix elements, and optical gains as a function of crystal orientation for zinc-blende and wurtzite AlGaN/GaN QW structures. These results can be used for a design of QW-based optoelectronic devices.

References

1. Corzine, S. W., R. H. Yan, and L. A. Coldren. 1993. Optical gain in III-V bulk and quantum well semiconductors. In *Quantum Well Lasers*, edited by P. S. Zory, Jr. San Diego, CA: Academic Press.
2. Chuang, S. L. 1995. *Physics of Optoelectronic Devices*. New York, NY: Wiley.
3. Piprek, J. 2003. *Semiconductor Optoelectronic Devices: Introduction to Physics and Simulation*. San Diego, CA: Academic Press.
4. Ahn, D. and S.-H. Park. 2011. *Engineering Quantum Mechanics*. Hoboken, NJ: John Wiley & Sons.
5. Chao, C. Y.-P. and S. L. Chuang. 1992. Spin–orbit-coupling effects on the valence-band structure of strained semiconductor quantum wells. *Phys. Rev. B* **46**: 4110–4122.
6. Chuang, S. L. and C. S. Chang. 1996. The k.p method for strained wurtzite semiconductors. *Phys. Rev. B* **54**: 2491–2504.
7. Park, S.-H. and S. L. Chuang. 1999. Crystal-orientation effects on the piezoelectric field and electronic properties of strained wurtzite semiconductors. *Phys. Rev. B* **59**: 4725–4737.
8. Takeuchi, T., H. Amano, and I. Akasaki. 2000. Theoretical study of orientation dependence of piezoelectric effects in wurtzite strained GaInN/GaN heterostructures and quantum wells. *Jpn. J. Appl. Phys.* **39**: 413–416.
9. Park, S.-H. and S. L. Chuang. 2000. Comparison of zinc-blende and wurtzite GaN semiconductors with spontaneous polarization and piezoelectric field effects. *J. Appl. Phys.* **87**: 353–364.
10. Bir, G. L. and G. E. Pikus, 1974. *Symmetry and Strain-Induced Effects in Semiconductors*. New York, NY: Wiley.
11. Park, S.-H., S. L. Chuang, and D. Ahn. 2000. Intraband relaxation time effects on non-Markovian gain with many-body effects and comparison with experiment. *Semicond. Sci. Technol.* **15**: 203–208.
12. Chuang, S. L. 1996. Optical gain of strained wurtzite GaN quantum-well lasers. *IEEE J. Quan. Electron*. **32**: 1791–1800.
13. Ahn, D. 1997. Theory of non-Markovian optical gain in quantum-well lasers. *Prog. Quant. Electron.* **21**: 249–287.
14. Chow, W. W., S. W. Koch, and M. Sergent III. 1994. *Semiconductor-Laser Physics*. Berlin: Springer.
15. Haug, H. and S. W. Koch. 1993. *Quantum Theory of the Optical and Electronic Properties of Semiconductors*. Singapore: World Scientific.
16. Nakamura, S. and G. Fasol. 1997. *The Blue Laser Diode*. Berlin: Springer.
17. Martin, G., A. Botchkarev, A. Rockett, and H. Morkoç. 1996. Valence-band discontinuities of wurtzite GaN, AlN, and InN heterojunctions measured by x-ray photoemission spectroscopy. *Appl. Phys. Lett.* **68**: 2541–2543.

18. Bernardini, F., V. Fiorentini, and D. Vanderbilt. 1997. Spontaneous polarization and piezoelectric constants of III–V nitrides. *Phys. Rev. B*. **56**: 10024–10027.
19. Park, S.-H. and S. L. Chuang. 1998. Piezoelectric effects on electrical and optical properties of wurtzite GaN/AlGaN quantum well lasers. *Appl. Phys. Lett.* **72**: 3103–3105.
20. Ambacher, O., J. Smart, J. R. Shealy, N. G. Weimann, K. Chu, M. Murphy, W. J. Schaff, L. F. Eastman, R. Dimitrov, L. Wittmer, M. Stutzmann, W. Rieger, and J. Hilsenbeck. 1999. Two-dimensional electron gases induced by spontaneous and piezoelectric polarization charges in N- and Ga-face AlGaN/GaN heterostructures. *J. Appl. Phys.* **85**: 3222–3233.
21. Park, S.-H. 2002. Crystal orientation effects on electronic properties of wurtzite InGaN/GaN quantum wells. *J. Appl. Phys.* **91**: 9904–9908.
22. Henderson, R. H. and E. Towe. 1995. Strain and crystallographic orientation effects on interband optical matrix elements and band gaps of [II/]-oriented III–V epilayers. *J. Appl. Phys.* **78**: 2447–2455.
23. Henderson, R. H. and E. Towe. 1996. Effective mass theory for III–V semiconductors on arbitrary (hkl) surfaces. *J. Appl. Phys.* **79**: 2029–2037.
24. Bykhovski, A., B. Gelmont, and S. Shur. 1993. Strain and charge distribution in GaN-AIN-GaN semiconductor–insulator–semiconductor structure for arbitrary growth orientation. *Appl. Phys. Lett.* **63**: 2243–2245.
25. Smith, D. L. and C. Mailhiot. 1998. Piezoelectric effects in strained-layer superlattices. *J. Appl. Phys.* **63**: 2717–2719.
26. Wright, A. F. and J. S. Nelson. 1995. Consistent structural properties for AlN, GaN, and InN. *Phys. Rev. B* **51**: 7866.
27. Rubio, A., J. L. Corkill, M. L. Cohen, E. L. Shirley, and S. G. Louie. 1993. Quasiparticle band structure of AlN and GaN. *Phys. Rev. B*. **48**: 11810.
28. Hellwege, K. H. and O. Madelung (Eds.). 1982. *Physics of Group IV Elements and III–V Compounds*, Landolt-Börnstein, New Series, Group III, Vol. 17, Part A. Berlin: Springer-Verlag.
29. Fan, W. J., M. F. Li, T. C. Chong, and J. B. Xia. 1996. Electronic properties of zinc-blende GaN, AlN, and their alloys $\mathrm{Ga}_{1-x}\mathrm{Al}_x\mathrm{N}$. *J. Appl. Phys.* **79**: 18896.
30. Meney, A. T. and E. P. O'Reilly. 1995. Theory of optical gain in ideal GaN heterostructure lasers. *Appl. Phys. Lett.* **67**: 3013.
31. Kim, K., W. R. L. Lambrecht, and B. Segall. 1996. Elastic constants and related properties of tetrahedrally bonded BN, AlN, GaN, and InN. *Phys. Rev. B* **53**: 16310.
32. Van de Walle, C. G. and J. Neugebauer. 1997. Small valence-band offsets at GaN/InGaN heterojunctions. *Appl. Phys. Lett.* **70**: 2577.
33. Vurgaftman, I. and J. R. Meyer. 2003. Band parameters for nitrogen-containing semiconductors. *J. Appl. Phys.* **94**: 3675–3696.
34. Park, S.-H. 2011. Optical gain characteristics of non-polar Al-rich AlGaN/AlN quantum well structures. *J. Appl. Phys.* **110**: 063105.
35. Park, S.-H., Y.-T. Moon, D.-S. Han, J. S. Park, M.-S. Oh, and D. Ahn. 2011. Light emission enhancement in blue InGaAlN/InGaN quantum well structures. *Appl. Phys. Lett.* **99**: 181101.

12

Nanowires

12.1 Introduction: Semiconductor Nanowires 397
12.2 Basic Charge Confining Mechanisms in NWs........................ 398
Bulk Electronic Properties: Band Offsets and Work Functions • Approaches to Compute Elastic Properties of Nanostructures • Elastic Properties of NW Heterostructures • Piezoelectric and Spontaneous Polarization
12.3 Doping and the Influence of Surfaces................................... 406
Continuous, Homogeneous Doping-Related Background Charge • Random Dopant Fluctuations: The Discrete Nature of Dopants in an NW • Dielectrically Enhanced Ionization Energies of Dopants
12.4 Crystal-Phase Bandgap Engineering.................................... 410
12.5 Optical Anisotropy of NWs—The Antenna Effect 412

Oliver Marquardt

Vladimir M. Kaganer

and

Pierre Corfdir

12.1 Introduction: Semiconductor Nanowires

During the past years, much research effort has been dedicated to explore the properties of semiconductor nanowires (NWs). Correspondingly, remarkable progress has been achieved in design, growth, characterization, and theoretical description of these structures. Many different applications have been suggested, ranging from energy harvesting and storage (Wallentin et al. 2013) to single-photon emitters (Heiss et al. 2013) and general lighting, with some of them poised to enter the commercial market in the near future.

Free-standing semiconductor NWs, sometimes also referred to as *nanorods*, are grown using molecular-beam epitaxy (MBE) or metal-organic chemical vapor deposition (MOCVD) or produced from planar layers by lithography. They have diameters of a few tens to hundreds of nanometers and lengths of a few hundred nanometers to a few micrometers and exhibit some unique properties in comparison to planar semiconductor quantum wells (QWs) or embedded quantum dots (QDs). If the diameter of a NW is small enough such that quantum confinement along the transverse direction is ensured, the NW is called a *quantum wire*. A particular advantage of the NW geometry is that it facilitates a much better strain relaxation, resulting in a better overall material quality. Additionally, the large surface-to-volume ratio of NWs represents an advantage for light emitting and detecting devices.

If heterostructures such as QWs or QDs are incorporated in NWs, the free side facets additionally facilitate elastic relaxation, allowing one to design heterostructures from materials that have a large lattice mismatch with respect to the NW material, which could not be produced in a planar structure. Typical heterostructures in NWs are of either axial or radial character, as depicted in Figure 12.1. The active layer of the heterostructure is then either perpendicular to the growth direction (axial heterostructure) or parallel to it (radial heterostructure). Thickness and material composition of axial and radial layers can be controlled during the growth process, so that the electronic properties of the NW can be tailored to suit specific applications.

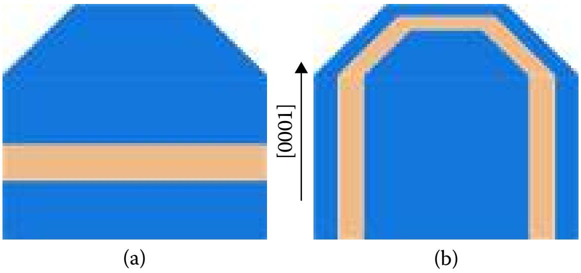


FIGURE 12.1 Cross-section of an axial (a) and a radial (b) heterostructure in a hexagonal nanowire (NW). The active layer is light gray.

Moreover, surface states at the side facets of NWs give rise to additional electrostatic potentials whose impact on the electronic properties needs to be understood. The simulation of semiconductor surfaces is a very challenging field and very few sophisticated approaches exist to model a semiconductor–vacuum interface.

Most III–V semiconductor NWs also show a high density of stacking faults. In some cases, the switching in the crystal structure along the length of an NW can be controlled. This allows one to investigate material properties of semiconductors in crystal structures that do not occur otherwise in the bulk. For example, whereas bulk GaAs exhibits a zinc-blende (ZB) crystal structure, GaAs NWs with the hexagonal wurtzite (WZ) structure have often been reported (Hoang et al. 2009; Heiss et al. 2011; Ahtapodov et al. 2012). Since the electronic properties of ZB and WZ GaAs are different, both phases have to be addressed in a simulation of an NW, with a particular focus on the treatment of their interface.

Finally, the photonic properties of NWs are altered by their peculiar geometry. In particular, the light absorption and extraction efficiencies depend strongly on the polarization of light with respect to the NW axis and it is mandatory to take this effect into account when correlating the degree of polarization of optical transitions with possible valence-band mixing in the NW.

12.2 Basic Charge Confining Mechanisms in NWs

In the following, we provide a brief overview of mechanisms that influence the localization and the energy levels of carriers in NWs. We compare these mechanisms to those that occur in planar layers or QDs and discuss specific features of NWs.

12.2.1 Bulk Electronic Properties: Band Offsets and Work Functions

Similar to planar heterostructures and QDs (see Chapters 11 and 13), the electronic properties of axial or radial semiconductor NW heterostructures are determined by the band offsets between the conduction and valence bands, E_c and E_v, of the materials involved. If we neglect surface states in a first step, the potential barrier arising at the interface between the NW and the surrounding vacuum is best described by the work function of the NW material. For the hole state, the following consideration yields a good description of the confining potential: The minimum energy to create a hole in the valence band is the sum of bandgap (to excite an electron from the valence to the conduction band) and the electron affinity. Therefore, we consider the sum of bandgap and electron affinity as a potential barrier for the valence band at the side facet of the NW. However, as the work function and the electron affinity of most semiconductors is on the order of a few eV and typically much larger than bandgaps and band offsets within an NW heterostructure, a good approximation can already be achieved by assuming an infinite potential at the semiconductor–vacuum interface. This simplification is in particular reasonable for the electron and hole states closest to the conduction and valence-band edges, respectively (Levine 1965).

12.2.2 Approaches to Compute Elastic Properties of Nanostructures

The electronic properties of a semiconductor nanostructure are significantly modified if the system is subject to elastic relaxation. The impact of strain on the electronic properties is commonly taken into account via deformation potentials. For example, within the ZB structure, the lowest conduction band is modified via the deformation potential a_c by the additional contribution $a_c \cdot \mathrm{Tr}(\varepsilon)$. In a WZ system, two deformation potentials a_1 and a_2 are employed such that the conduction band is modified by the term $a_2 \cdot (\varepsilon_{xx} + \varepsilon_{yy}) + a_1 \cdot \varepsilon_{zz}$. The strain-induced modifications of the valence bands are more complex. As an example, the impact of strain on the three highest valence bands can be computed using a six-band Pikus–Bir Hamiltonian (Ghosh et al. 2002). Chapter 1 provides a more detailed discussion on how strain is considered in the simulation of electronic properties of semiconductor nanostructures.

NWs exhibit quite specific strain profiles as the free side facets facilitate elastic relaxation. This feature allows not only the NW itself to relax toward the bulk lattice constants but also a much better elastic relaxation of axial and radial heterostructures incorporated within the wire. The possibility to relax even a large lattice mismatch is a particular strength of NWs in comparison to planar heterostructures, where large differences in the lattice constants lead to defects and crack formation.

To shed some light on the specific elastic properties of semiconductor NWs and their impact on the electronic properties, we discuss different approaches to compute the elastic properties of an NW in the following.

12.2.2.1 Valence Force Field Model

The most complete description of the elastic properties of an NW is achieved using an atomistic model. However, accurate models rely on *ab initio* methods, which are computationally highly expensive thus limiting the simulation domain to a small number of atoms (Yang et al. 2008; Xiang et al. 2008). Given the typical dimensions of semiconductor NWs in the order of tens to hundreds of nanometers, it is clear that such models are not well suited to describe NWs of the commonly observed dimensions.

An alternative approach is the computationally less expensive valence force field (VFF) model. Within the VFF model (Keating 1966), the atomic coordinates $\boldsymbol{r}_i$ of the individual atoms of the NW are the degrees of freedom. For an NW with a ZB crystal structure, the potential energy of the crystal to be minimized can be written as (Keating 1966):

$$E = \sum\sum \frac{3\alpha}{8r_{ij0}^2}\left(r_{ij}^2 - r_{ij0}^2\right)^2 + \sum\sum \frac{3\beta}{8r_{ij0}\cdot r_{ik0}}\left(r_{ij}\cdot r_{ik} - r_{ij0}\cdot r_{ik0}\right)^2, \qquad (12.1)$$

where the equilibrium bond angles satisfy $\boldsymbol{r}_{ij0} \cdot \boldsymbol{r}_{ik0} = -|\boldsymbol{r}_{ij0}||\boldsymbol{r}_{ik0}/3$ and N indicates summation over the nearest neighbors. $\boldsymbol{r}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$ is the distance between atoms i and j and $\boldsymbol{r}_{ij0}$ denotes the value of $\boldsymbol{r}_{ij}$ in a bulk, strain-free crystal. The parameters α and β are coupling constants between neighboring atoms.

However, the basic VFF model is inaccurate for hexagonal crystal structures and thus requires improvements (Rücker and Methfessel 1995; Große and Neugebauer 2001), and it is questionable whether this model is an appropriate choice for the description of semiconductor NWs (Singh et al. 2011). Moreover, the computational effort of the VFF scales nearly linearly with the number of atoms involved. For typical NWs, even the VFF becomes computationally highly expensive if the super cell spans over a large segment of the NW.

12.2.2.2 Linear Continuum Elasticity Theory

As a computationally inexpensive alternative to the atomistic VFF model, the elastic properties of an NW can also be computed using a continuum-based approach. Here, the underlying atomistic lattice is ignored and the strain tensor $\varepsilon_{ij}(\boldsymbol{r})$ is expressed in a continuous manner. The elastic energy to be minimized is then (Landau and Lifshitz 1959) given as follows:

$$E = \frac{1}{2} \int \sum C_{ijkl}(r)\varepsilon_{ij}(r)\varepsilon_{kl}(r)\mathrm{d}V, \tag{12.2}$$

where $C_{ijkl}(\boldsymbol{r})$ denotes the elastic constants of the materials involved and $i, j, k, l = x, y, z$. Note that all information on the underlying crystal structure is contained in the elastic constants. Within the ZB lattice, only the constants $C_{11} = C_{iiii} = C_{xxxx} = C_{yyyy} = C_{zzzz}$, $C_{12} = C_{iijj}$ $(i \neq j)$, and $C_{44} = C_{ijij}$ $(i \neq j)$ are nonzero. The advantage of the continuum elasticity model is of course its computational efficiency, as the discretization of the super cell is not bound to atomistic coordinates. However, the atomistic nature of the system under consideration is fully neglected.

12.2.3 Elastic Properties of NW Heterostructures

In this section, we discuss strain distributions in axial and radial NW heterostructures, as well as their impact on the electronic properties. We also discuss the specific advantages of axial and radial NW heterostructures in terms of elastic relaxation.

12.2.3.1 Radial NWs Heterostructures

In radial semiconductor NW heterostructures—typically referred to as core–shell NW—a core material is surrounded by one or more shells consisting of different materials or having different doping densities. The elastic properties of core and shells are controlled by their respective thicknesses, which finally allows the control of the electronic properties. In a simplified picture, a core–shell NW can be approximated as a cylinder with a lattice mismatched shell. The strain inside such a cylindrical core–shell NW is best described within cylindrical coordinates and the nonzero components of the strain tensor are (Menendez et al. 2011)

$$\begin{aligned}
\varepsilon_{rr} = \varepsilon_{\theta\theta} &= -\varepsilon_0 \left(R_\mathrm{s}^2 - R_\mathrm{c}^2\right) \left[\frac{(\nu+1)(1-2\nu)}{(1-\gamma)(1-2\nu)R_\mathrm{c}^2 - (1-2\nu+\gamma)R_\mathrm{s}^2} + \frac{\nu}{\left(R_\mathrm{s}^2 - R_\mathrm{c}^2\right) + \gamma R_\mathrm{c}^2}\right] \\
\varepsilon_{zz} &= \varepsilon_0 \left[\frac{\left(R_\mathrm{s}^2 - R_\mathrm{c}^2\right)}{\left(R_\mathrm{s}^2 - R_\mathrm{c}^2\right) + \gamma R_\mathrm{c}^2}\right]
\end{aligned} \tag{12.3}$$

in the core and

$$\begin{aligned}
\varepsilon_{rr}(r) &= \frac{\varepsilon_0(\nu+1)\gamma R_\mathrm{c}^2}{(1-\gamma)(1-2\nu)R_\mathrm{c}^2 - (1-2\nu+\gamma)R_\mathrm{s}^2}\left[(1-2\nu) - \frac{R_\mathrm{s}^2}{r^2}\right] + \frac{\nu\gamma R_\mathrm{c}^2\varepsilon_0}{\left(R_\mathrm{s}^2 - R_\mathrm{c}^2\right) + \gamma R_\mathrm{c}^2} \\
\varepsilon_{\theta\theta}(r) &= \frac{\varepsilon_0(\nu+1)\gamma R_\mathrm{c}^2}{(1-\gamma)(1-2\nu)R_\mathrm{c}^2 - (1-2\nu+\gamma)R_\mathrm{s}^2}\left[(1-2\nu) + \frac{R_\mathrm{s}^2}{r^2}\right] + \frac{\nu\gamma R_\mathrm{c}^2\varepsilon_0}{\left(R_\mathrm{s}^2 - R_\mathrm{c}^2\right) + \gamma R_\mathrm{c}^2} \\
\varepsilon_{zz} &= \frac{\gamma R_\mathrm{c}^2\varepsilon_0}{\left(R_\mathrm{s}^2 - R_\mathrm{c}^2\right) + \gamma R_\mathrm{c}^2}
\end{aligned} \tag{12.4}$$

in the shell, where r is the radial position. $\varepsilon_0 = \left(a_\mathrm{s} - a_\mathrm{c}\right)/a_\mathrm{c}$ is the initial misfit for core and shell materials with lattice constants a_c and a_s, respectively. R_c and R_s denote core and shell radii. For the Poisson ratio ν, we use the Voigt average (Hirth and Lothe 1968) and consider this ratio as being constant inside both the core and the shell. Within a cubic crystal, this ratio reads

$\nu^{\mathrm{c,s}} = C_{11}^{\mathrm{c,s}} + 4C_{12}^{\mathrm{c,s}} - 2C_{44}^{\mathrm{c,s}}/4C_{11}^{\mathrm{c,s}} + 6C_{12}^{\mathrm{c,s}} + 2C_{44}^{\mathrm{c,s}}$ where the $C_{ij}^{c,s}$ are the elastic constants of the core and the shell.

$\gamma = E_{\text{core}}/E_{\text{shell}}$ with E denoting Young's modulus. In the case of an NW grown along the [111] axis, this modulus can be determined from the stress–strain relation as (Hirth and Lothe 1968)

$$E_{111} = 3\frac{C_{44}\left(C_{11} + 2C_{12}\right)}{C_{44} + C_{11} + 2C_{12}}.$$

All components of the strain tensor are constants inside the core and depend only on the elastic constants of the materials involved as well as on the ratio between core and shell radii. Within the shell, ε_{zz} is constant. In contrast, $\varepsilon_{rr}(r)$ and $\varepsilon_{\theta\theta}(r)$ are functions of the radial distance from the NW core, and converge for large radii, as shown in Figure 12.2 for a Ge–Si core–shell NW with a core radius of 10 nm and a shell radius of 20 nm. The ratio between the radii of the core and the shell determines the strain state of the NW and is a property that can be easily controlled during the growth process.

For more realistic NW geometries that cannot be described using analytical models, the elastic properties can be computed only numerically using continuum-based or atomistic approaches. A comparison between the analytic solution for the cylindrical NW and a numerical solution using the finite-elements method for hexagonal, [011]-grown Ge–Si NWs can be found in Singh et al. (2011), where the authors report a very good agreement with the analytic model for the ε_{zz} component.

Discrepancies are observed in ε_{rr} and $\varepsilon_{\theta\theta}$ inside the shell toward the corners of the realistic, hexagonal NW, whereas a very reasonable agreement is reported in the vicinity of the core. Note, however, that ε_{rr} and $\varepsilon_{\theta\theta}$ are not uniform within the hexagonal core anymore, as derived analytically for a cylindrical NW. The agreement in ε_{zz} is very good with uniform values in both core and shell (Singh et al. 2011).

A comparative study between a continuum elasticity model and the atomistic VFF model is presented by Grönquist et al. (2009) for GaAs/GaP core–shell NWs with a total diameter of 27.8 nm. The agreement between VFF and continuum elasticity model is very good throughout the NW cross-section and deviations occur only at the free ends of the finite wires. For typical semiconductor NWs, elasticity theory within a continuum picture can thus be considered as a reasonable and computationally inexpensive model.

The elastic relaxation within the model NW has a significant influence on its electronic properties. $\text{Tr}(\varepsilon) > 0$ within the shell leads to a reduction of both the conduction and the valence-band energies. Within the core, the hydrostatic strain is negative and the energies of the conduction band minimum and valence band maximum increase. The impact of strain on the conduction band is larger, such that the bandgap of the core material (GaAs) increases and a red shift of luminescence from the core is predicted. As the magnitude of the strain depends on the thicknesses of the core and the shell, bandgap engineering can be applied in core–shell NWs by a variation of the core and shell thicknesses. An example for such bandgap engineering is presented in Liu et al. (2014), where the bandgap of an InAs shell around an InP core is tuned by up to 210 meV solely by a variation of the core diameter.

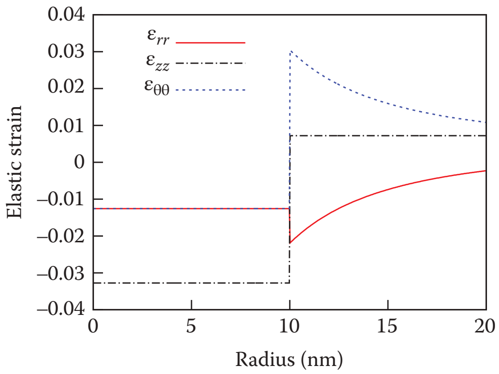


FIGURE 12.2 Nonzero strain components of a Ge–Si core–shell nanowire (NW) in cylindrical coordinates. The radius of the Ge core is 10 nm.

12.2.3.2 Axial NW Heterostructures

Axial NW heterostructures consist, similar to planar heterostructures, of thin layers separated by the matrix material (barriers). In contrast to their planar counterparts, the elastic strain caused by the mismatch between the insertion and the matrix crystal lattices relaxes at the side facets. This relaxation only partially eliminates the strain in the wells and gives rise to a complicated inhomogeneous strain distribution that depends on the relation between the NW diameter, the thickness of the insertion, the barrier thickness, and also the thickness of the whole stack of the insertions. The elastic relaxation facilitated by the free side facets is of particular importance for $In_xGa_{1-x}N$ disks in GaN NWs, as InN and GaN exhibit a lattice mismatch of approximately 10%, making it difficult to produce planar $In_xGa_{1-x}N$/GaN heterostructures with large In content and high crystal quality.

The difference in lattice parameters between the matrix and the inserted layer gives rise to the strain $\varepsilon^0_{zz} = (c - c_0)/c_0$, $\varepsilon^0_{rr} = \varepsilon^0_{\theta\theta} = (a - a_0)/a_0$, which itself does not cause stress and is called the intrinsic strain (the terms self-strain and eigenstrain are also used). a (a_0) and c (c_0) denote the layer (matrix) lattice constants. The elastic strain ε and the stress σ result from the coherency of the crystal lattices of the layer and the matrix. In a planar heterostructure, the layer is clamped by the matrix. Therefore, the components of the total strain in the plane $\varepsilon^T_{rr} = \varepsilon^0_{rr} + \varepsilon_{rr}$ and $\varepsilon^T_{\theta\theta} = \varepsilon^0_{\theta\theta} + \varepsilon_{\theta\theta}$ are equal to zero. The whole structure can freely expand in the direction normal to the layer, so that the stress $\sigma_{zz} = 0$. Using Hooke's law, these conditions lead to the stress in the layer plane $\sigma_{rr} = \sigma_{\theta\theta} = -\sigma^0$, where $\sigma^0 = (C_{11} + C_{12} - 2C^2_{13}/C_{33})\varepsilon^0_{rr}$. In NWs, this stress is relaxed at the free side facets. The solution of the above problem for the NW is, therefore, the sum of two terms: The solution in a planar structure described earlier and the image field due to stress $\sigma^{im}_{rr} = \sigma^0$ applied at the interface of the layer to satisfy the boundary conditions $\sigma_{rr} = 0$, $\sigma_{rz} = 0$ at the side facets. The elastic problem can be solved analytically for a circular cylinder shape of the NW and an arbitrary intrinsic strain distribution $\varepsilon^0(z)$ (Kaganer and Belov 2012). In this case, the image displacements are written as

$$u^{\mathrm{im}}_r(x, \varsigma) = -\frac{R\sigma^0}{c_{44}} \int_0^\infty \left[\frac{1}{1+k_1} \frac{I_1(qx/\nu_1)}{I_1(q/\nu_1)} - \frac{1}{1+k_2} \frac{I_1(qx/\nu_2)}{I_1(q/\nu_2)} \right] \frac{g(q)}{D(q)} \cos q\varsigma \mathrm{d}q,$$

$$u^{\mathrm{im}}_z(x, \varsigma) = -\frac{R\sigma^0}{c_{44}} \int_0^\infty \left[\frac{k_1\nu_1}{1+k_1} \frac{I_0(qx/\nu_1)}{I_1(q/\nu_1)} - \frac{k_2\nu_2}{1+k_2} \frac{I_0(qx/\nu_2)}{I_1(q/\nu_2)} \right] \frac{g(q)}{D(q)} \sin q\varsigma \mathrm{d}q,$$

$$D(q) = \frac{c_{11} - c_{12}}{c_{44}} \frac{k_2 - k_1}{(1+k_1)(1+k_2)} - q\left(\nu_1 \frac{I_0(q/\nu_1)}{I_1(q/\nu_1)} - \nu_2 \frac{I_0(q/\nu_2)}{I_1(q/\nu_2)} \right). \tag{12.5}$$

I_0 and I_1 are modified Bessel functions, $\zeta = z/R$, and $x = r/R$ with R being the radius of the cylinder. The constants k_1, k_2, ν_1, ν_2 are defined by the elastic moduli of the material, which are $k_1 = 0.234$, $k_2 = 4.264$, $\nu_1 = 1.589$, $\nu_2 = 0.623$ for GaN. The function $g(z)$ describes the strain distribution and $g(q)$ its Fourier transform. For a homogeneous disk of thickness $2d$, it is given by $g(q) = 2sin(qd/R)/(\pi q)$.

Figure 12.3 illustrates the strain state in the disk center of a cylindrical NW with an inserted disk. In the limit of large diameters, the strain state approaches to the one of a planar structure. When the diameter is much smaller than the disk thickness, the inserted disk is strain free. As shown in Figure 12.3, the strain changes sign when going from very thin to thick NWs since both the insertion and the matrix are distorted. The insert in Figure 12.3 visualizes the displacements calculated by Equation 12.5 for the cylinder diameter corresponding to the minimum of ε_{zz} (exaggerated for better visibility).

The strain in the inserted disk is compensated by an opposite strain in the matrix. The displacements well above and well below the disk obtained using Equation 12.5 show that the total displacement due to transverse intrinsic strain is zero. In other words, the strain averaged over the whole cylinder is zero. This result has important consequences for multi-QW NW heterostructures since each QW experiences

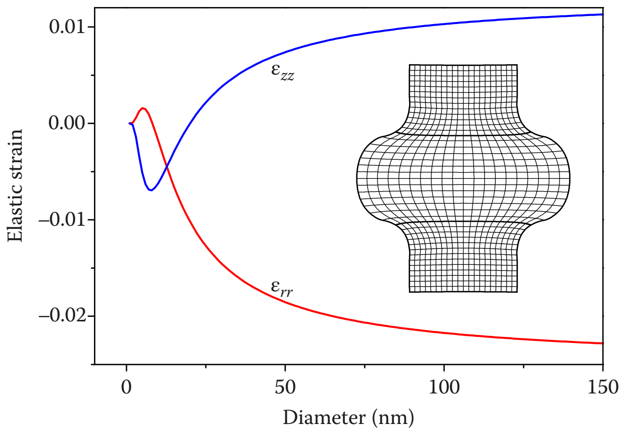


FIGURE 12.3 Elastic strain at the center of a 5-nm-thick $In_{0.25}Ga_{0.75}N$ disk inserted in a GaN cylinder as a function of the diameter of the cylinder. The inset shows the displacements (magnified for better visibility) for a cylinder diameter of 8 nm corresponding to the minimum of ε_{zz}.

strain produced by the other QWs. Variation of the barrier thickness and the NW diameter are therefore additional degrees of freedom for strain engineering of NW heterostructures (Wölz et al. 2013).

Figure 12.4 shows three strain states of lattice-mismatched heterostructures. In the case of a planar heterostructure (cf. Figure 12.4a), the Poisson effect gives rise to a homogeneous strain in the layer. In a multilayer NW heterostructure, the lateral elastic relaxation occurs in both materials. Figure 12.4b shows the displacement (center) and the out-of-plane strain (right) that were obtained for an axial superlattice using Equation 12.5. The strain is nonuniformly distributed. In the center of the cylinder, shown by the red line, the barrier crystal is laterally expanded. The QW is still under lateral compression, but to a smaller extent than in the planar case of Figure 12.4a. Going outwards, the strain distribution remains qualitatively similar to the one in the center up to about $0.8R$, shown by the green line. Toward the surface, the strain profile is more complex as shown by the blue line at $0.95R$.

This complicated strain state can be approximated by the strain state shown in Figure 12.4c, where it is assumed that the in-plane lattice spacing is constant everywhere in the superlattice. The strain can be determined by the condition that the average strain on the side surface is zero (Wölz et al. 2011, 2012). This relaxation reduces the strain in the QW by the factor $d_{\text{barrier}}/(d_{\text{barrier}} + d_{\text{well}})$ compared to the case of a planar heterostructure in Figure 12.4a. This lateral relaxation approximation is a much simpler description of the dominant strain states in the superlattice compared to the full consideration of the inhomogeneous strain resulting in a configuration as shown in Figure 12.4b. A systematic comparison of the full calculation based on Equation 12.5 with this approximation allows one to establish criteria for its validity (Wölz et al. 2013). First, the height of the whole stack of QWs should be larger than the NW diameter. Second, to achieve a large strain in QWs, the NWs should be sufficiently thick.

For more complicated geometries or material compositions, the strain can be determined by finite element calculations. For example, Figure 12.5 shows the strain distribution in an axial $In_xGa_{1-x}N$/GaN NW heterostructure containing two $In_xGa_{1-x}N$ insertions (Knelangen et al. 2011). Each insertion consists of a 2-nm-thick $In_xGa_{1-x}N$ layer with an In content of 19% and a 9-nm-thick layer with an In content of 39%. They are embedded in a GaN NW with a diameter of 30 nm and separated from the side facets by a 7-nm-thick GaN shell. The distance between the two disks is 3 nm. The magnitude of the strain increases toward the side facets, whereas elastic relaxation occurs near the center of the NW. Nevertheless, significant strain persists at the side facets of the NW.

A study on single $In_xGa_{1-x}N$ disks in a GaN NW that extend throughout the whole NW section without a GaN shell also reports persisting strains (Böcklin et al. 2010). An $In_{0.4}Ga_{0.6}N$ layer with a thickness

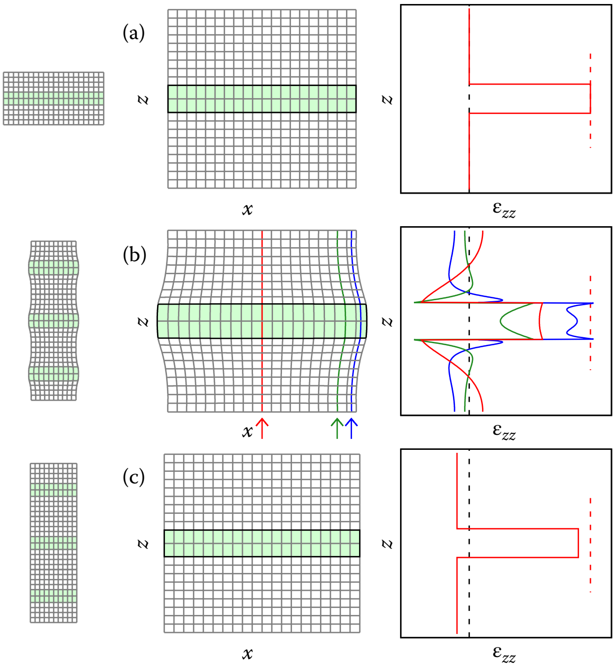


FIGURE 12.4 Strain states of lattice-mismatched heterostructures: sketch of the entire structure (left), atomic displacement (center), and out-of-plane strain ε_{zz} (right) near one quantum well (QW). (a) Planar pseudomorphic epitaxial QW growth, the substrate remains strain free. (b) Axial nanowire (NW) superlattice, nonuniform strain profile calculated with a cylindrical model. The red, green, and blue curves represent the displacement and strain at the radial positions indicated by the arrows (NW center, 80% and 95% of NW radius). (c) Axial NW superlattice, lateral relaxation estimate, QW and barrier assume a common average in-plane lattice parameter. (Reprinted with permission from Wölz et al. (2013). Copyright [2013] **American Chemical Society**.)

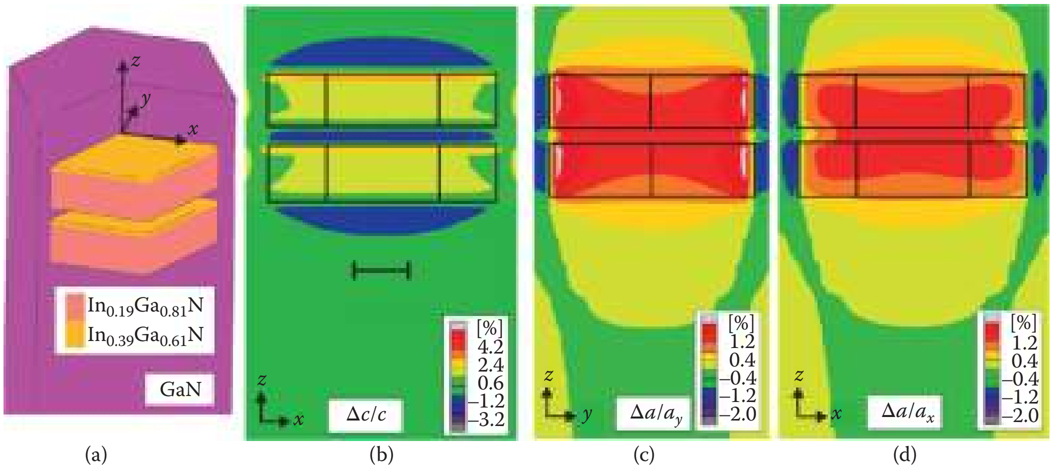


FIGURE 12.5 (a) Perspective view of the GaN nanowire (NW) with two $In_xGa_{1-x}N$ disks. The diagonal components of the strain tensor, ε_{zz}, ε_{yy}, and ε_{xx}, are shown in (b), (c), and (d), respectively. (Reproduced with permission from Knelangen, M. et al., *Nanotechnology*, 2011. © IOP Publishing. All rights reserved.)

of 3 nm within a hexagonal GaN NW with a diameter of 50 nm is highly strained due to the large In content and experiences a significant lateral extension. In the center of the layer, the magnitude of the strain approaches the value predicted in the biaxial limit. As a consequence, the bandgap is significantly lowered at the NW surface, inducing a localization of electrons and holes at the corners and at the side facets, respectively.

12.2.4 Piezoelectric and Spontaneous Polarization

Both GaN and InN exhibit pyroelectric and piezoelectric polarization. Planar heterostructures are commonly grown along the polar [0001] axis, resulting in large internal electrostatic fields in the $In_xGa_{1-x}N$ QWs. These fields lead to a spatial separation of electrons and holes and thus to a considerable reduction of the radiative recombination rate. The use of axial GaN NW heterostructures promises to resolve some of these problems (Kikuchi et al. 2004; Kim et al. 2004; Li and Waag 2012). As discussed in the previous section, the free side facets facilitate the elastic relaxation of strain in the $In_xGa_{1-x}N$ disk, so that higher In contents can be achieved (Björk et al. 2002; Ertekin et al. 2005; Kaganer and Belov 2012). Along with the elastic relaxation, a significant reduction of the built-in piezoelectric potential and hence the quantum confined Stark effect (QCSE) (Miller et al. 1984) is expected. In fact, some researchers report evidence of a vanishing QCSE (Lin et al. 2010; Armitage and Tsubaki 2010; Nguyen et al. 2011; Bardoux et al. 2009), whereas others observe a significant QCSE (Wölz et al. 2013). Typically, the polarization $\boldsymbol{P}$ is computed directly using the strain tensor. For WZ crystals, it reads (Nye 1985):

$$P(r) = \begin{pmatrix} 2e_{15}\varepsilon_{13} \\ 2e_{15}\varepsilon_{23} \\ e_{31}\left(\varepsilon_{11} + \varepsilon_{22}\right) + e_{33}\varepsilon_{33} + P_{\mathrm{sp}} \end{pmatrix}, \tag{12.6}$$

where $e_{ij} = e_{ij}(\boldsymbol{r})$ are the piezoelectric constants and $P_{\mathrm{sp}}(\boldsymbol{r})$ is the spontaneous polarization is the spontaneous polarization: WZ crystals exhibit a charge separation even in the absence of an electric field. The polarization potential $V(\boldsymbol{r})$ is computed solving the Poisson equation.

$$\rho_{\mathrm{p}}(r) = \varepsilon_0 \nabla \cdot \left[\varepsilon_r(r) \nabla V(r)\right] \quad \text{with} \quad \rho_{\mathrm{p}}(r) = -\nabla \cdot P(r). \tag{12.7}$$

ε_0 and ε_r denote the vacuum and the relative dielectric constants, respectively.

The resulting electrostatic potential $V(\boldsymbol{r})$ is used in the model to compute the electronic properties as a potential contribution added to both the conduction and the valence bands.

For the system considered by Böcklin et al. (2010), accounting for the polarization potential leads to a confinement of the hole ground state at the center of the NW, whereas the electron remains at the edges. Depending on size and material composition of NW and disk, both strain and the polarization potential can therefore dominate the localization of carriers in a NW.

For an $In_xGa_{1-x}N$ disk in a circular cylinder GaN NW, the Poisson equation can be solved using Equation 12.5 and the electrostatic potential is expressed in similar integrals (Kaganer et al. 2016). Figure 12.6 displays the electrostatic potential V along the axis of the cylinder ($r = 0$) for an $In_{0.25}Ga_{0.75}N$ disk of thickness 5 nm inserted into GaN cylinders of different diameters. The potential for a film (i.e., a cylinder of an infinite diameter) is included for comparison. Already for very thick NWs with a diameter of 150 nm, the electrostatic potential of the disk along the central NW axis is visibly reduced and this reduction increases for decreasing diameters. However, even for a very thin NW with a diameter of 15 nm, a potential difference of approximately 300 meV persists.

Although the polarization potential is significantly reduced compared to the one of planar heterostructures of similar growth direction, thickness, and In content, a very small NW diameter ($d < 30$ nm) is required to achieve a substantial reduction of the piezoelectric potential for the typical disk thicknesses of only a few nanometers. Even in such systems, a nonvanishing built-in potential remains. The magnitude of

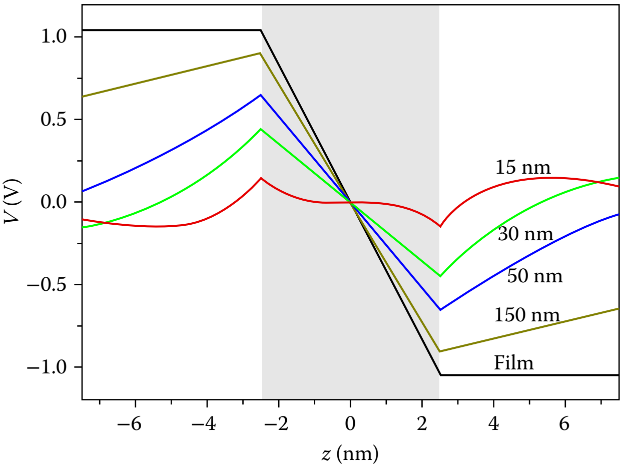


FIGURE 12.6 The electrostatic potential V along the axis of the cylinder ($r = 0$) for a 5-nm-thick $In_{0.25}Ga_{0.75}N$ disk inserted into GaN cylinders of different diameters, as indicated in the figure. The interior of the disk is highlighted. (Reproduced with permission from Kaganer, V. M. et al., *Nanotechnology*, 2016. ©IOP Publishing. All rights reserved.)

the polarization potential decreases with larger aspect ratio and the extrema of the potential shift from the central axis of the NW toward the side facets (Kaganer et al. 2016). The polarization potential therefore not only affects the transition energies but also the confinement of electrons and holes and thus the recombination rates. Strain relaxation in axial NW heterostructures represents an additional degree of freedom to control the emission wavelength and the electron–hole overlap in comparison to their planar counterparts. However, electrostatic potentials remain and exert an important influence on the electronic properties of the system.

12.3 Doping and the Influence of Surfaces

The electrical conductivity of semiconductor materials can be tuned over many orders of magnitude by the intentional incorporation of impurities (Schubert 2005). The ability to dope semiconductors in a controlled fashion is a key feature of all semiconductor-based technology (Sze 1985; Seeger 1991). For a theoretical description, doping poses a formidable problem, as dopants occupy random positions in the host lattice. Established solid-state simulation methods exploiting the periodicity of the system cannot be employed (Shklovskii and Efros 1984). For device simulations, the discrete nature of the dopant positions is therefore commonly ignored altogether, and a homogeneous, continuous distribution of charge is assumed (Keyes 1975).

In semiconductor NWs, intentional or unintentional doping is of particular importance due to the vicinity of free surfaces. In the case of an $In_xGa_{1-x}N$ based NW, the residual doping concentration of O and Si atoms is $\rho_D = 10^{16} - 10^{17}$ cm^{-3}. These donors transfer their extra electrons to surface states, leaving the donor ionized and the NW depleted at all temperatures, as long as the radius of the NW does not exceed the typical thickness of the depletion region in the bulk material. In the following, we discuss concepts to simulate the impact of doping on the electronic structure of semiconductor NWs.

12.3.1 Continuous, Homogeneous Doping-Related Background Charge

In a simple picture, the potential arising from ionized donors is treated as an electrostatic potential obtained from a homogeneous background charge density by solving the Poisson equation. In the case of a semiconductor NW with doping density ρ_D and with a cylindrical shape, the surface potential V arising from a homogeneous background charge density is

$$V(r) = E_\mathrm{F} - \frac{e \cdot \rho_\mathrm{D}}{4\varepsilon_0 \cdot \varepsilon_\mathrm{r}} R^2 \left[1 - \left(\frac{r}{R}\right)^2\right], \tag{12.8}$$

with R being the radius of the NW and E_F being the surface Fermi level pinning. For a cylindrical GaN NW with $R = 40$ nm and a doping density of 10^{17} cm^{-3}, the potential difference between the center and the side facets of the NW is 80 mV. If the residual doping consists of Si and O, which is typical for GaN, the potential is positive at the side facets, that is, attractive for holes.

Figure 12.7 shows the electron (red) and hole (blue) ground-state-charge density in an axial $\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}$/GaN NW heterostructure for various values of x and accounting for the influence of a surface potential, computed using an eight-band **k·p** model (see Chapter 1 for details). It can be seen that the surface potential gives rise to a behavior specific to NWs. As long as the thickness of the $\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}$ disk is small compared to the NW diameter, the strain-induced piezoelectric potential exhibits extrema at the top and bottom surface of the $\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}$ disk along the central axis of the NW (see discussion in Section 12.2.4). For the hole state, this potential is attractive at the bottom center of the disk. A surface potential arising from an unintentional background doping, on the contrary, localizes hole states at the side facets of the NW. The localization of the hole state therefore depends on the magnitude of surface and polarization potentials: For large In contents, the polarization potential increases and induces a localization of holes along the central axis of the NW. For small In contents, the localization of the hole state is dominated by the surface potential and the hole is confined near the side facets of the NW (Marquardt et al. 2013).

12.3.2 Random Dopant Fluctuations: The Discrete Nature of Dopants in an NW

In a typical axial $\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}$/GaN NW heterostructure, the number of dopants is commonly not large enough to justify the picture of a homogeneous doping-related background charge (Corfdir et al. 2014). With a diameter of 80 nm, a doping concentration of 10^{17} cm^{-3} corresponds on average to only 8.3 charges within a segment of 20 nm length. Therefore, individual, randomly distributed dopants have to be considered in NWs. Of course, the ideal approach to deal with influences of individual atoms is an atomistic picture, such as empirical tight binding or pseudopotential methods or density functional theory (for more details on these models see Chapter 1). At the same time, experimentally relevant semiconductor

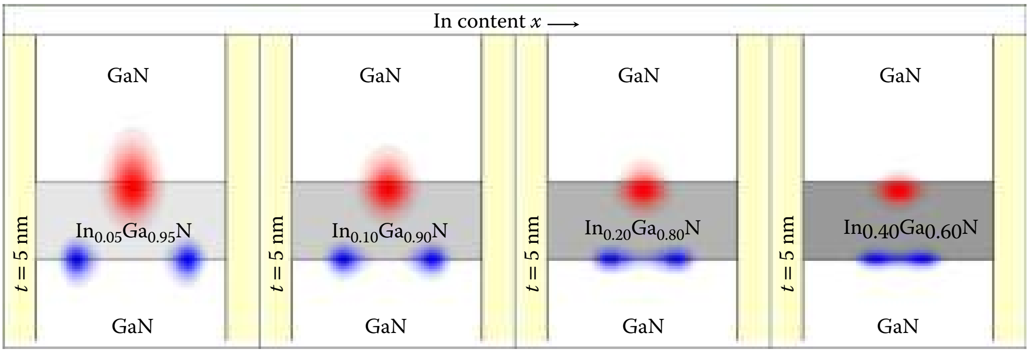


FIGURE 12.7 Side view of the electron (red) and hole (blue) ground-state charge density in an axial $\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}$/GaN nanowire (NW) heterostructure for $x = 0.05, 0.1, 0.2$, and 0.4 (left to right). The NW diameter and disk thickness are not to scale. (Reprinted with permission from Marquardt et al. (2013). Copyright 2013 American Chemical Society.)

NWs exhibit typical diameters of a few tens to hundreds of nanometers. Therefore, a reasonable simulation domain commonly contains millions of atoms. Random dopant fluctuations (RDFs) induce variations of the electronic potential landscape between individual NWs. A meaningful study of the impact of RDFs on the electronic structure of NWs will thus require a statistical evaluation of these variations so that a correspondingly large number of simulations needs to be performed. Within an atomistic picture, this represents a massive computational effort. However, individual dopants can also be included within the computationally inexpensive continuum approaches, for example, effective mass or multiband **k·p** models (see Chapter 1 for more details) under certain conditions.

If an impurity can be considered a hydrogenic dopant, its main influence on the electronic structure is a long-range Coulomb potential. These shallow donors and acceptors induce only a small deformation to the host lattice and require only small energies (less than 3 $k_B T$ ~75 meV at room temperature) to be ionized. Some examples for shallow donors and acceptors in different host materials are listed in Table 12.1.

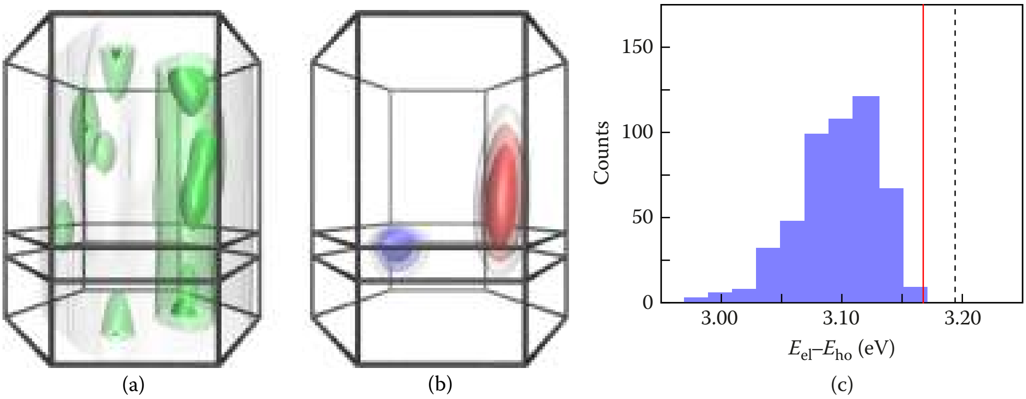
A shallow impurity can be modeled as a point charge. The respective Coulomb potential obtained from the Poisson equation is then simply added to the potential landscape within the simulation cell. Note that a shallow impurity in a binary material is not necessarily a shallow impurity in a ternary system. For example, O is a shallow donor in GaN but its behavior in $Al_xGa_{1-x}N$ alloys is still a matter of controversy. Some studies report a shallow–deep transition (Park and Chadi 1997; McCluskey et al. 1998), whereas others claim O to remain a shallow donor in $Al_xGa_{1-x}N$ even for large Al contents (Kakanakova-Georgieva et al. 2013). For deep dopants, the model of a simple point charge for an ionized dopant becomes inaccurate.

Figure 12.8a shows an example of a Coulomb potential resulting from RDFs in an axial $In_xGa_{1-x}N$/GaN NW heterostructure. The respective electron (red) and hole (blue) ground-state charge densities are shown in the center of Figure 12.8, computed using an eight-band **k·p** model (Chuang and Chang 1996). To investigate the variations of electron and hole wave functions as well as binding energies in an ensemble of otherwise identical NWs, one can perform a statistical study by employing a larger number of random donor distributions. The electron–hole ground-state recombination energy resulting from such a study is shown in Figure 12.8c. Additionally, the energy values obtained by assuming a continuous, homogeneous

TABLE 12.1 Activation Energies of Shallow Impurities in Different Host Materials

Host Material	Donor	$E_c - E_d$ (meV)	Acceptor	$E_a - E_v$ (meV)
GaAs	S	6	Mg	28
	Se	6	Zn	31
	Te	30	Cd	35
	Si	58	Si	26
GaN	O	33.21 (Freitas et al. 2003)		
	Si	30.19 (Freitas et al. 2003)		
	C	34.0 (Wang and Chen 2000)		
	Ge	31.1 (Wang and Chen 2000)		
	S	29.5 (Wang and Chen 2000)		
	Se	29.5 (Wang and Chen 2000)		
AlN	Si	60 (Zeisel et al. 2000)		
ZnO	Al	53 (Meyer et al. 2005)		
	Ga	54.5 (Meyer et al. 2005)		
	In	63.2 (Meyer et al. 2005)		
Si	Sb	39	B	45
	P	45	Al	67
	As	54	Ga	73

Source: If not indicated differently, taken from Whitaker (2005). Freitas, J. A. et al., *Phys. Stat. Solidi B,* 240, 330, 2003; Wangm, H. and A.-B. Chen, *J. Appl. Phys.,* 87, 7859, 2000; Zeisel, R. et al. *Phys. Rev. B*, 61, R16283, 2000; Meyer, B. K. et al., *Semicond. Sci. Technol.*, 20, S62, 2005.



(a) (b) (c)

FIGURE 12.8 (a) Coulomb potential isosurfaces for a random distribution of nine individual dopants in an axial $In_xGa_{1-x}N$/GaN nanowire (NW) heterostructure. The active layer is marked in the lower part of the NW. (b) Electron (red) and hole (blue) ground-state charge density. (c) Histogram of the electron–hole ground-state recombination energy for an ensemble of 500 different random dopant distributions with an average of 8.3 charges in the vicinity of the active layer. For the dimensions of the NW segment under consideration, this number corresponds to a doping density of 10^{17} cm^{-3}.

doping-related background charge (solid red line) as well as the one of an ideal, undoped NW (dashed black line) are shown. It is seen that both of these energies are larger than the ones obtained assuming RDFs, which arises from the fact that RDFs induce much deeper local potentials than the surface potential arising from a homogeneous doping-related background charge. The ground-state recombination energy of the undoped NW represents an upper limit for the other two models (Marquardt et al. 2015).

In contrast to shallow impurities, an atomistic treatment as described in Chapter 1 is necessary for deep donors or acceptors, that is, impurities that exhibit a large chemical shift or induce a large deformation to the host lattice.

12.3.3 Dielectrically Enhanced Ionization Energies of Dopants

In the previous sections, we have neglected the mismatch in dielectric constants at the surface of the NWs. Still, the discontinuity of the dielectric constant at the surface of NWs strongly affects the properties of dopants in these nanostructures. In particular, it has been demonstrated that the Coulomb interaction is enhanced in nanostructures surrounded by a barrier material with a lower dielectric constant (Keldysh 1979). This so-called dielectric confinement is a consequence of the electrostatic potential that is set up by the presence of a charged particle in a nanostructure and that can be calculated using the image charge method (Kumagai and Takagahara 1989).

Whereas the mismatch in dielectric constants at the interface of heterostructures such as GaN (Al,Ga)N or GaAs/(Al,Ga)As is small and induces only minor changes in the binding energy of excitons and in the ionization energy of dopants (Andreani and Pasquarello 1990), the mismatch is maximum for NWs in air or vacuum, and the increase in the dopant ionization energies due to the dielectric confinement is considerable (Shik 1993; Diarra et al. 2007). Although quantum confinement leads as well to increased ionization energies for impurities in NWs with a radius smaller than the Bohr radius of donors in the bulk, a_B (Bryant 1984), the dielectric confinement in semiconductor NWs in vacuum sets in already for NW radii smaller than $5a_B$ (Diarra et al. 2007). This finding has been verified experimentally by Björk et al. (2009) who measured a significant increase in resistivity in n-doped Si NWs with a diameter smaller than 15 nm.

The impact of image charges on donors or excitons in low dimensional systems has been studied using envelope function calculations (Kumagai and Takagahara 1989), tight binding methods (Allan et al. 1995; Diarra et al. 2007), or density functional calculations (Chan et al. 2008). All of these simulations predict an increase in ionization energy for donors located at the center of a nanostructure as a result of the dielectric confinement (Shik 1993; Diarra et al. 2007; Chan et al. 2008; Corfdir and Lefebvre 2012a). The situation is more complicated for donors located off the axis of NWs. Early effective-mass simulations neglecting the dielectric confinement have predicted that the infinite potential barrier at the surface leads to a change in the symmetry of the wave function of the donor ground-state electron and to a monotonic decrease in donor ionization energy when getting closer to the surface (Levine 1965; Satpathy 1983). In particular, the binding energy of an electron to a donor located at the surface of a semi-infinite layer is one quarter of that of a bulk donor (Levine 1965). However, accounting for the combination of the surface potential barrier with the image charges, one obtains that the ionization energy of a donor in a thin slab bounded by vacuum exhibits a nonmonotonic dependence on the donor site: It first increases and then decreases when the donor moves from the center to the surface of the slab (Corfdir and Lefebvre 2012a). This position-dependent binding energy of electrons to donors has been used to explain the inhomogeneous broadening of the photoluminescence lines related to donor bound excitons in unintentionally doped and strain free GaN NWs (Corfdir et al. 2014). However, using scanning tunneling microscopy, Wijnheimer et al. measured a gradual increase in the ionization of Si over the last 1.2 nm below the (110) surface of a thick GaAs layer (Wijnheijmer et al. 2009). This increase, not reproduced by the effective mass theory, was tentatively explained by the authors in terms of broken symmetry at the surface.

The increase in dopant ionization energies in thin NWs due to the dielectric confinement leads to a decrease in conductivity, which may hinder the realization of efficient electronic and optoelectronic devices based on these structures (Björk et al. 2009). In addition, theoretical and experimental studies have shown that the deactivation of dopants is not homogeneous across the NW section (Wijnheijmer et al. 2009). This finding, together with the nonuniform distribution of dopants in NWs discussed above and the modification of the dielectric environment by electrical contacts or gates (Diarra et al. 2008), renders the simulation of NW-based devices challenging.

12.4 Crystal-Phase Bandgap Engineering

An additional peculiarity of III–V NWs with respect to their planar counterpart is their pronounced polytypism, that is, the crystal structure along the NW length alternates between WZ and ZB. Although III-nitride NWs usually grow within the WZ structure with only a few stacking faults, III-arsenides and III-phosphides NWs show frequent alternation between ZB and WZ segments of various lengths along their axis (Figure 12.9a; Algra et al. 2008; Bao et al. 2008; Caroff et al. 2009; Jacopin et al. 2011; Heiss et al. 2011; Graham et al. 2013). This polytypism is usually regarded as detrimental for the electronic properties of NWs: Alternation between WZ and ZB crystal structures efficiently scatters electrons (Konar et al. 2011) and leads to a significant increase in the resistivity of NWs (Thelander et al. 2011). In addition, the presence of stacking faults along the length of NWs leads to localization of electrons and holes, increasing the ionization energies of dopants in these nanostructures (Corfdir et al. 2009).

Although considerable effort has been devoted to the elimination of stacking defects in NWs, some degree of control on the switching between WZ and ZB during the NW growth has been demonstrated (Algra et al. 2008; Caroff et al. 2009). This control opens the possibility of realizing crystal-phase bandgap engineering in NWs, where carrier confinement is obtained by alternation of the crystal structure and not by the composition (Algra et al. 2008; Caroff et al. 2009; Akopian et al. 2010; Corfdir et al. 2013).

Such crystal-phase quantum structures exhibit atomically flat interfaces (Bolinsson et al. 2011) and are free of alloy disorder. Thanks to the high quality of their interfaces, crystal-phase quantum structures exhibit a sub-meV photoluminescence line width at cryogenic temperatures (Figure 12.9b; Akopian et al. 2010; Jacopin et al. 2011; Graham et al. 2013) and they can be of interest for quantum optics applications (Akopian et al. 2010; Castelletto et al. 2014).

The band alignment between the WZ and the ZB phases of a III–V semiconductor is usually considered to be of type II, with the electron and the hole localized in the ZB and WZ phases, respectively

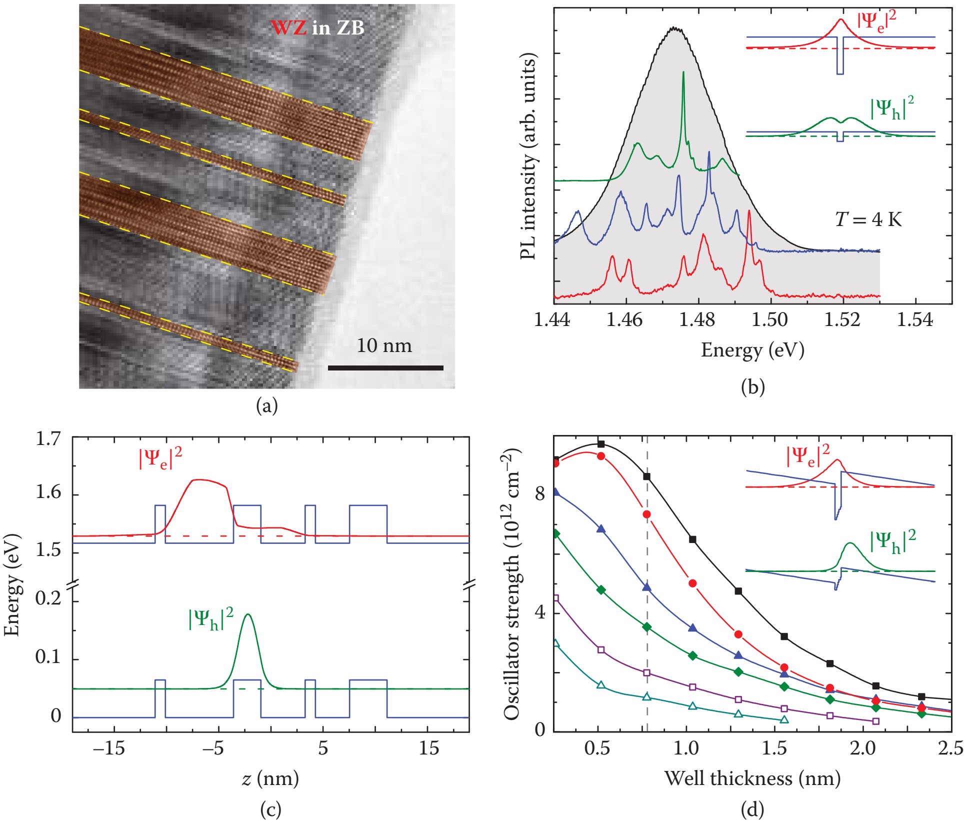


FIGURE 12.9 (a) High-resolution transmission electron micrograph of a polytype GaAs nanowire (NW). Wurtzite (WZ) segments are highlighted in red. (Adapted with permission from Corfdir et al. (2013). Copyright 2013 American Chemical Society.) (b) Photoluminescence spectra at 4 K from single NWs (red, blue, green) and from an ensemble (gray) of polytype GaAs NWs. The signal between 1.46 and 1.50 eV arises from recombination at crystal-phase quantum disks. The inset depicts schematically the band profile for a zinc-blende (ZB) inclusion in WZ GaAs (blue). The energy and the wave function of the electron (red) and the hole (green) are shown with solid and dashed lines, respectively. (Adapted with permission from Graham et al. (2013). Copyright 2013 by the American Physical Society.) (c) Band profiles and electron and hole energy and wave function for the NW segment shown in (a). (Adapted with permission from Corfdir et al. (2013). Copyright 2013 American Chemical Society.) (d) Oscillator strength for an exciton in a QW of ZB GaN in WZ GaN as a function of the well thickness and in the presence of built-in electric fields. The lines from top to bottom correspond to built-in electric fields of 0, 0.1, 0.5, 1, 2, and 3 MV/cm, respectively. The built-in electric fields along the confinement axis lead to a spatial separation of the electron and hole wave functions (inset). (Adapted with permission from Corfdir and Lefebvre (2012a). Copyright 2013, American Institute of Physics.)

(Figure 12.9b; Murayama and Nakayama 1994). *Ab initio* calculations have predicted small band offsets and small bandgap differences between the WZ and the ZB phases (Bechstedt and Belabbes 2013). Therefore, although long radiative lifetimes have been calculated for thick crystal-phase QWs in NWs (Zhang et al. 2010), the electron and hole wave functions for thin ZB insertions in the WZ show significant spreading into the WZ and the ZB, respectively. Their squared overlap is large (Corfdir and Lefebvre 2012b; Corfdir et al. 2013) and thin crystal-phase QWs and QDs in NWs act as efficient light emitters, despite their type-II band alignment.

Effective-mass calculations have been used to calculate the photoluminescence energy associated to single crystal-phase QWs and to polytype semiconductor NW segments (Bao et al. 2008; Jacopin et al. 2011; Heiss et al. 2011; Corfdir and Lefebvre 2012b; Lähnemann et al. 2012; Corfdir et al. 2013). For instance, such calculations have been combined with photoluminescence and transmission electron microscopy experiments to obtain information on the band structure of WZ GaAs (Figure 12.9c; Corfdir et al. 2013). However, the conclusions obtained from those studies remain only qualitative since many parameters required for the calculations (e.g., the conduction and valence-band offsets) are poorly known. The situation is particularly complex for GaAs NWs. Despite a large number of experimental and theoretical studies, there is still no agreement on the value of the bandgap of WZ GaAs; values larger and smaller than the bandgap of ZB GaAs have been reported, and the symmetry of the lowest energy conduction band of this material remains under debate (Martelli et al. 2015). Note also that, due to the small conduction and valence-band offsets between the WZ and the ZB phases of a semiconductor, the energy computed for an exciton in crystal-phase quantum structures of a thickness of a few monolayers is not sensitive to the exact values of the band offsets. As shown in Lähnemann et al. (2012), the photoluminescence energy for three-to-five-monolayers-thick crystal-phase QWs in GaN NWs can in fact be reproduced using either a type-I or a type-II band alignment between WZ and ZB GaN.

The connection of bands at the interface of different crystal phases is of particular importance for the modeling of polytype NWs. In the case of GaAs, the lowest conduction band in the ZB phase has a Γ_6 symmetry and is well separated in energy from any remote conduction bands (Belabbes et al. 2012). The equivalent of the Γ_6 band in the WZ structure is the Γ_7 band. However, compounds with a WZ structure show an additional band with a Γ_8 symmetry that results from the zone folding due to doubling of the WZ unit cell along the [0001] direction when compared to the ZB unit cell. For WZ GaAs, recent studies (De and Pryor 2010; Heiss et al. 2011; Belabbes et al. 2012; Graham et al. 2013) suggest that this Γ_8 band is very close in energy to the Γ_7 band. It is, however, not clear yet which of these two bands exhibits the lowest energy. This point is, however, essential since (1) these two bands are expected to exhibit very different effective masses and a weak coupling with the fundamental Γ_7 and Γ_9 valence bands, and (2) confinement in crystal-phase quantum disks could give rise to some conduction band mixing.

Finally, although strain-free compounds with a ZB structure do not exhibit any spontaneous polarization, WZ III–V semiconductors are polar materials. Differential phase contrast microscopy and photoluminescence spectroscopy on polytype NWs have shown that the spontaneous polarization in WZ GaAs and GaN is aligned along the *c*-axis of the WZ and has a strength equal to 2.7×10^{-3} and -2.2×10^{-2} C/m^2, respectively (Bauer et al. 2014; Lähnemann et al. 2012). Since the axis of most WZ NWs is parallel to the *c* direction, the spontaneous polarization field in polytype NWs shows discontinuities at the interfaces between WZ and ZB segments. These discontinuities give rise to built-in electric fields parallel to the axis of polytype NWs with an intensity on the order of MV/cm in WZ/ZB GaN quantum structures (Corfdir and Lefebvre 2012b; Lähnemann et al. 2012). The resulting quantum confined Stark effect strongly affects the optical properties of thick crystal-phase QWs. The photoluminescence transitions are redshifted and may be centered even at energies below the band gap of ZB GaN, as observed for ZB insertions thicker than six monolayers (Jacopin et al. 2011), and the oscillator strength of the exciton is decreased (Figure 12.9d; Corfdir and Lefebvre 2012b).

12.5 Optical Anisotropy of NWs—The Antenna Effect

Semiconductors in the form of NWs not only show altered electrical characteristics but also strongly modified optical properties. As shown by Wang et al. (2001), the electric field at the sidewalls of NWs with a diameter much smaller than the wavelength is discontinuous (continuous) for light polarized perpendicular (parallel) to the NW axis. For an NW considered as an infinite dielectric cylinder, the electric field is $E_i = 2\kappa_0/\left(\kappa_r + \kappa_0\right) E_0$ for light polarized perpendicular to the NW axis, where E_0 is the electric field outside the NW and κ_0 and κ_r are the dielectric constants of vacuum and of the NW, respectively. Since $\kappa_r \approx 10$ in III–V semiconductors, this *antenna effect* leads to significant polarization anisotropies of interband transitions in NWs. In particular, the antenna effect may affect the emission intensity ratio between different interband transitions. For instance, while heavy holes have dipoles oriented perpendicular to the axis of GaAs NWs, light holes exhibit a large on-axis dipole. Optical transitions related to light–hole excitons are thus promoted by the antenna effect (Spirkoska et al. 2012). Complete reversal of the polarization anisotropy with respect to the bulk has even been predicted for WZ NWs with crystal field and spin–orbit splittings such that the A and B valence bands are close in energy (Efros and Lambrecht 2014).

The situation is more complicated for NWs with diameters on the order of the wavelength, as they support guided modes (Ruppin 2002). The latter gives rise to oscillations in the degree of polarization of the NW luminescence with varying wavelength or NW diameter (Giblin et al. 2009; Corfdir et al. 2015). NW ensembles usually exhibit a broad diameter distribution, and the polarization response may vary from one NW to another (Corfdir et al. 2015).

The polarization properties of interband transitions have been commonly used to extract the strain state of planar structures. Comparison between the measured optical anisotropy and the one computed using $\mathbf{k}\cdot\mathbf{p}$ calculations has made it possible to estimate the strain state of group-III-nitrides layers grown along a nonpolar or a semipolar axis (Ghosh et al. 2002; Funato et al. 2013). The analysis is more complicated for NWs since the combination of strain, dielectric contrast and possible surface-induced wave function distortion (see Section 12.3.3) may result in a complex polarization anisotropy of the interband transitions in NWs. For instance, the degree of polarization of single InP/InAs NWs exhibiting valence-band mixing and dispersed on a carbon grid had to be calculated using detailed finite difference time domain calculations (Anufriev et al. 2015).

Acknowledgments

The authors would like to thank Oliver Brandt and Lutz Geelhaar as well as present and former colleagues at Paul-Drude-Institut für Festkörperelektronik, Berlin, Germany, and at the Cavendish Laboratory, Cambridge, UK, for stimulating and fruitful discussions. We also thank Lutz Schrottke for critical reading of the manuscript. Pierre Corfdir acknowledges funding from the Fonds National Suisse de la Recherche Scientifique through project 161032.

References

Ahtapodov, L., J. Todorovic, P. Olk, T. Mjåland, P. Slåttnes, D. L. Dheeraj, A. T. J. Van Helvoort, B. O. Fimland, and H. Weman. 2012. A story told by a single nanowire: Optical properties of wurtzite GaAs. *Nano Lett.* 12:6090.

Akopian, N., G. Patriarche, L. Liu, J.-C. Harmand, and V. Zwiller. 2010. Crystal phase quantum dots. *Nano Lett.* 10:1198.

Algra, R. E., M. A. Verheijen, M. T. Borgström, L.-F. Feiner, G. Immink, W. J. P. van Enckevort, E. Vlieg, and E. P. A. M. Bakkers. 2008. Twinning superlattices in indium phosphide nanowires. *Nature* 456:369.

Allan, G., C. Delerue, M. Lannoo, and E. Martin. 1995. Hydrogenic impurity levels, dielectric constant, and Coulomb charging effects in silicon crystallites. *Phys. Rev. B* 52:11982.

Andreani, L. C., and A. Pasquarello. 1990. Accurate theory of excitons in GaAs-$Ga_{1-x}Al_xAs$ quantum wells. *Phys. Rev. B* 42:8928.

Anufriev, R., J.-B. Barakat, G. Patriarche, X. Letartre, C. Bru-Chevallier, J.-C. Harmand, M. Gendry, and N. Chauvin. 2015. Optical polarization properties of InAs/InP quantum dot and quantum rod nanowires. *Nanotechnology* 26:395701.

Armitage, R., and K. Tsubaki. 2010. Multicolour luminescence from InGaN quantum wells grown over GaN nanowire arrays by molecular-beam epitaxy. *Nanotechnology* 21:195202.

Bao, J., D. C. Bell, F. Capasso, J. B. Wagner, T. Mårtensson, J. Trägårdh, and L. Samuelson. 2008. Optical properties of rotationally twinned InP nanowire heterostructures. *Nano Lett.* 8:836.

Bardoux, R., A. Kaneta, M. Funato, Y. Kawakami, A. Kikuchi, and K. Kishino. 2009. Positive binding energy of a biexciton confined in a localization center formed in a single $In_xGa_{1-x}N$/GaN quantum disk. *Phys. Rev. B* 79:155307.

Bauer, B., J. Hubmann, M. Lohr, E. Reiger, D. Bougeard, and J. Zweck. 2014. Direct detection of spontaneous polarization in wurtzite GaAs nanowires. *Appl. Phys. Lett.* 104:211902.

Bechstedt, F., and A. Belabbes. 2013. Structure, energetic, and electronic states of III–V compound types. *J. Phys. Condens. Matter* 25:273201.

Belabbes, A., C. Panse, J. Furthmüller, and F. Bechstedt. Electronic bands of III–V semiconductor polytypes and their alignment. 2012. *Phys. Rev. B* 86:075208.

Björk, M. T., B. J. Ohlsson, T. Sass, A. I. Persson, C. Thelander, M. H. Magnusson, K. Deppert, L. R. Wallenberg, and L. Samuelson. 2002. One-dimensional heterostructures in semiconductor whiskers. *Appl. Phys. Lett.* 80:1058.

Björk, M. T., H. Schmid, J. Knoch, H. Riel, and W. Riess. 2009. Donor deactivation in silicon nanostructures. *Nat. Nanotechnol.* 4:103.

Böcklin, C., R. G. Veprek, S. Steiger, and B. Witzigmann. 2010. Computational study of an InGaN/GaN nanocolumn light-emitting diode. *Phys. Rev. B* 81:155306.

Bolinsson, J., P. Caroff, B. Mandl, and K. A. Dick. 2011. Wurtzite-zincblende superlattices in InAs nanowires using a supply interruption method. *Nanotechnology* 22:265606.

Bryant, G. W. 1984. Hydrogenic impurity states in quantum-well wires. *Phys. Rev. B* 29:6632.

Caroff, P. K., A. Dick, J. Johansson, M. E. Messing, K. Deppert, and L. Samuelson. 2009. Controlled polytypic and twin-plane superlattices in III–V nanowires. *Nat. Nanotechnol.* 4:50.

Castelletto, S., Z. Bodrog, A. P. Magyar, A. Gentle, A. Gali, and I. Aharonovich. 2014. Quantum-confined single photon emission at room temperature from SiC tetrapods. *Nanoscale* 6:10027.

Chan, T. L., M. L. Tiago, E. Kaxiras, and J. R. Chelikowsky. 2008. Size limits on doping phosphorus into silicon nanocrystals. *Nano Lett.* 8:596.

Chuang, S. L., and C. S. Chang. 1996. $\mathbf{k}\cdot\mathbf{p}$ method for strained wurtzite semiconductors. *Phys. Rev. B* 54, 2491.

Corfdir, P., P. Lefebvre, J. Ristić, J.-D. Ganière, and B. Deveaud-Plédran. 2009. Electron localization by a donor in the vicinity of a basal stacking fault in GaN. *Phys. Rev. B* 80:153309.

Corfdir, P., and P. Lefebvre. 2012a. Role of the dielectric mismatch on the properties of donors in semiconductor nanostructures bounded by air. *J. Appl. Phys.* 112:106104.

Corfdir, P., and P. Lefebvre. 2012b. Importance of excitonic effects and the question of internal electric fields in stacking faults and crystal phase quantum discs: The model-case of GaN. *J. Appl. Phys.* 112:053512.

Corfdir, P., B. Van Hattem, E. Uccelli, S. Conesa-Boj, P. Lefebvre, A. Fontcuberta i Morral, and R. T. Phillips. 2013. Three-dimensional magneto-photoluminescence as a probe of the electronic properties of crystal-phase quantum disks in GaAs nanowires. *Nano Lett.* 13:5303.

Corfdir, P., J. K. Zettler, C. Hauswald, S. Férnandez-Garrido, O. Brandt, and P. Lefebvre. 2014. Sub-meV linewidth in GaN nanowire ensembles: Absence of surface excitons due to the field ionization of donors. *Phys. Rev. B* 90:205301.

Corfdir, P., F. Feix, J. K. Zettler, S. Fernández-Garrido, and O. Brandt. 2015. Importance of the dielectric contrast for the polarization of excitonic transitions in single GaN nanowires. *New J. Phys.* 17:033040.

De, A., and C. Pryor. 2010. Predicted band structures of III–V semiconductors in the wurtzite phase. *Phys. Rev. B* 81:155210.

Diarra, M., Y.-M. Niquet, C. Delerue, and G. Allan. 2007. Ionization energy of donor and acceptor impurities in semiconductor nanowires: Importance of dielectric confinement. *Phys. Rev. B* 75:045301.

Diarra, M., C. Delerue, Y.-M. Niquet, and G. Allan. 2008. Screening and polaronic effects induced by a metallic gate and a surrounding gate oxide on donor and acceptor impurities in silicon nanowires. *J. Appl. Phys.* 103:073703.

Efros, A. L., and W. R. L. Lambrecht. 2014. Theory of light emission polarization reversal in zinc-blende and wurtzite nanowires. *Phys. Rev. B* 89:035304.

Ertekin, E., P. A. Greaney, D. C. Chrzan, and T. D. Sands. 2005. Equilibrium limits of coherency in strained nanowire heterostructures. *J. Appl. Phys.* 97:114325.

Freitas, J. A., W. J. Moore, B. V, Shanabrook, G. C. B. Braga, D. D. Koleske, S. K. Lee, S. S. Park, and J. Y. Han. 2003. Shallow donors in GaN. *Phys. Stat. Solidi B* 240:330.

Funato, M., K. Matsuda, R. G. Banal, R. Ishii, and Y. Kawakami. 2013. Strong optical polarization in nonpolar (1–100) $Al_xGa_{1-x}N$/AlN quantum wells. *Phys. Rev. B* 87:041306(R).

Ghosh, S., P. Waltereit, O. Brandt, H. T. Grahn, and K. H. Ploog. 2002. Electronic band structure of wurtzite GaN under biaxial strain in the M plane investigated with photoreflectance spectroscopy. *Phys. Rev. B* 65: 075202.

Giblin, J., V. Protasenko, and M. Kuno. 2009. Wavelength sensitivity of single nanowire excitation polarization anisotropies explained through a generalized treatment of their linear absorption. *ACS Nano* 3:1979.

Graham, A. M., P. Corfdir, M. Heiss, S. Conesa-Boj, E. Uccelli, A. Fontcuberta, I. Morral, and R. T. Phillips. 2013. Exciton localization mechanisms in wurtzite/zinc-blende GaAs nanowires. *Phys. Rev. B* 87:125304.

Grönquist, J., N. Søndergaard, F. Boxberg, T. Guhr, S. Åberg, and H. Q. Xu. 2009. Strain in semiconductor core–shell nanowires. *J. Appl. Phys.* 106:053508.

Große, F., and J. Neugebauer. 2001. Limits and accuracy of valence force field models for $In_xGa_{1-x}N$ alloys. *Phys. Rev. B* 63:085207.

Heiss, M., S. Conesa-Boj, J. Ren, H. H. Tseng, A. Gali, A. Rudolph, E. Uccelli, F. Peiró, J. R. Morante, D. Schuh, E. Reiger, E. Kaxiras, J. Arbiol, and A. Fontcuberta i Morral. 2011. Direct correlation of crystal structure and optical properties in wurtzite/zinc-blende GaAs nanowire heterostructures. *Phys. Rev. B* 83:045303.

Heiss, M., Y. Fontana, A. Gustafsson, G. Wüst, C. Magen, D. D. O'Regan, J. W. Luo, B. Ketterer, S. Conesa-Boj, A. V. Kuhlmann, J. Houel, E. Russo-Averchi, J. R. Morante, M. Cantoni, N. Marzari, J. Arbiol, A. Zunger, R. J. Warburton, and A. Fontcuberta i Morral. 2013. Self-assembled quantum dots in a nanowire system for quantum photonics. *Nat. Materials* 12:439.

Hirth, J.P., and J. Lothe. 1968. *Theory of Dislocations*. New York: McGraw-Hill.

Hoang, T. B., A. F. Moses, H. L. Zhou, D. L. Dheeraj, B. O. Fimland, and H. Weman. 2009. Observation of free exciton photoluminescence emission from single wurtzite GaAs nanowires. *Appl. Phys. Lett.* 94:133105.

Jacopin, G., L. Rigutti, L. Largeau, F. Fortuna, F. Furtmayr, F. H. Julien, M. Eickhoff, and M. Tchernycheva. 2011. Optical properties of wurtzite/zincblende heterostructures in GaN nanowires. *J. Appl. Phys.* 110:064313.

Kaganer, V. M., and A. Yu. Belov. 2012. Strain and x-ray diffraction from axial nanowire heterostructures. *Phys. Rev. B* 85:125402.

Kaganer, V. M., O. Marquardt, and O. Brandt. 2016. Piezoelectric potential in axial (In,Ga)N/GaN nanowire heterostructures. *Nanotechnology* 27(16):165201.

Kakanakova-Georgieva, A., D. Nilsson, X. T. Trinh, U. Forsberg, N. T. Son, and E. Janzén. 2013. The complex impact of silicon and oxygen on the n-type conductivity of high-Al-content AlGaN. *Appl. Phys. Lett.* 102:132113.

Keating, P. N. 1966. Effect of invariance requirements on the elastic strain energy of crystals with application to the diamond structure. *Phys. Rev.* 145:637.

Keldysh, L. 1979. Coulomb interaction in thin semiconductor and semimetal films. *JETP Lett.* 29:658.

Keyes, W. R. 1975. The effect of randomness in the distribution of impurity atoms on FET thresholds. *Appl. Phys.* 8:251.

Kikuchi, A., M. Kawai, M. Tada, and K. Kishino. 2004. InGaN/GaN multiple quantum disk nanocolumn light-emitting diodes grown on (111) Si substrate. *Jpn. J. Appl. Phys.* 43:1524.

Kim, H., Y. Cho, H. Lee, S. Kim, S. Ryu, D. Kim, T. Kang, and K. Chung. 2004. High-brightness light emitting diodes using dislocation-free indium gallium nitride/gallium nitride multiquantum-well nanorod arrays. *Nano Lett.* 4:1059.

Knelangen, M., M. Hanke, E. Luna, L. Schrottke, O. Brandt, and A. Trampert. 2011. Monodisperse (In,Ga)N insertions in catalyst-free-grown GaN(0001) nanowires. *Nanotechnology* 22:365703.

Konar, A., T. Fang, N. Sun, and D. Jena. 2011. Charged basal stacking fault scattering in nitride semiconductors. *Appl. Phys. Lett.* 98:022109.

Kumagai, M., and T. Takagahara. 1989. Excitonic and nonlinear-optical properties of dielectric quantum-well structures. *Phys. Rev. B* 40:12359.

Lähnemann, J., O. Brandt, U. Jahn, C. Pfüller, C. Roder, P. Dogan, F. Grosse, A. Belabbes, F. Bechstedt, A. Trampert, and L. Geelhaar. 2012. Direct experimental determination of the spontaneous polarization of GaN. *Phys. Rev. B* 86:081302(R).

Landau, L. D., and E. M. Lifshitz. 1959. *Theory of Elasticity*. London: Pergamon.

Levine, J. 1965. Nodal hydrogenic wave functions of donors on semiconductor surfaces. *Phys. Rev.* 140:A586.

Li, S., and A. Waag. 2012. GaN-based nanorods for solid state lighting. *J. Appl. Phys.* 111:071101.

Lin, H. W., Y. J. Lu, H. Y. Chen, H. M. Lee, and S. Gwo. 2010. InGaN/GaN nanorod array white light-emitting diode. *Appl. Phys. Lett.* 97:073101.

Liu, P., H. Huang, X. Liu, M. Bai, D. Zhao, Z. Tang, X. Huang, J.-Y. Kim, and J. Guo. 2014. Core–shell nanowire diode based on strain-engineered bandgap. *Phys. Stat. Solidi A* 212:617.

Marquardt, O., C. Hauswald, M. Wölz, L. Geelhaar, and O. Brandt. 2013. Luminous efficiency of axial $In_xGa_{1-x}N$/GaN nanowire heterostructures: Interplay of polarization and surface potentials. *Nano Lett.* 13:3298.

Marquardt, O., L. Geelhaar, and O. Brandt. 2015. Impact of random dopant fluctuations on the electronic properties of $In_xGa_{1-x}N$/GaN axial nanowire heterostructures. *Nano Lett.* 15:4289.

Martelli, F., G. Priante, and S. Rubini. 2015. Photoluminescence of GaAs nanowires at an energy larger than the ZB band-gap: Dependence on growth parameters. *Semicond. Sci. Tech.* 30:055020.

McCluskey, M. D., N. M. Johnson, C. G. Van de Walle, D. P. Bour, M. Kneissl, and W. Walukiewicz. 1998. Metastability of oxygen donors in AlGaN. *Phys. Rev. Lett.* 80:4008.

Menéndez, J., R. Singh, and J. Drucker. 2011. Theory of strain effects on the Raman spectrum of Si–Ge core–shell nanowires. *Ann. Phys.* 523:145.

Meyer, B. K., J. Sann, D. M. Hofmann, C. Neumann, and A. Zeuner. 2005. Shallow donors and acceptors in ZnO. *Semicond. Sci. Technol.* 20:S62.

Miller, D. A. B., D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus. 1984. Band-edge electroabsorption in quantum well structures: The quantum-confined stark effect. *Phys. Rev. Lett.* 53:2173.

Murayama, M., and T. Nakayama. 1994. Chemical trend of band offsets at wurtzite/zinc-blende heterocrystalline semiconductor interfaces. *Phys. Rev. B* 49:4710–4724.

Nguyen, H. P. T., S. Zhang, K. Cui, X. Han, S. Fathololoumi, M. Couillard, G. A. Botto, and Z. Mi. 2011. p-Type modulation doped InGaN/GaN dot-in-a-wire white-light-emitting diodes monolithically grown on Si(111). *Nano Lett.* 11:1919.

Nye, J. F. 1985. *Physical Properties of Crystals—Their Representation by Tensors and Matrices*. Clarendon: Oxford.

Park, C. H., and D. J. Chadi. 1997. Stability of deep donor and acceptor centers in GaN, AlN, and BN. *Phys. Rev. B* 55:12995.

Rücker, H., and M. Methfessel. 1995. Anharmonic Keating model for group-IV semiconductors with application to the lattice dynamics in alloys of Si, Ge, and C. *Phys. Rev. B* 52:11059.

Ruppin, R. 2002. Extinction by a circular cylinder in an absorbing medium. *Optics Commun.* 211:335.

Satpathy, S. 1983. Eigenstates of Wannier excitons near a semiconductor surface. *Phys. Rev. B* 28:4585.

Schubert, E. F. 2005. *Doping in III–V Semiconductors*. Cambridge: Cambridge University Press.

Seeger, K. 1991. *Semiconductor Physics*. Berlin: Springer.

Shik, A. 1993. Excitons and impurity centers in thin wires and in porous silicon. *J. Appl. Phys.* 74:2951.

Shklovskii, B. I., and A. L. Efros. 1984. *Electronic Properties of Doped Semiconductors*. Berlin: Springer.

Singh, R., E. J. Dailey, J. Drucker, and J. Menéndez. 2011. Raman scattering from Ge–Si core–shell nanowires: Validity of analytical strain models. *J. Appl. Phys.* 110:124305.

Spirkoska, D., A. L. Efros, W. R. L. Lambrecht, T. Cheiwchanchamnangij, A. Fontcuberta i Morral, and G. Abstreiter. 2012. Valence band structure of polytypic zinc-blende/wurtzite GaAs nanowires probed by polarization-dependent photoluminescence. *Phys. Rev. B* 85:045309.

Sze, S. M. 1985. *Semiconductor Devices: Physics and Technology*. New York: Wiley.

Thelander, C., P. Caroff, S. Plissard, A. W. Dey, and K. A. Dick. 2011. Effects of crystal phase mixing on the electrical properties of InAs nanowires. *Nano Lett.* 11:2424.

Wallentin, J., N. Anttu, D. Asoli, M. Huffman, I. Åberg, M. H. Magnusson, G. Siefer, P. Fuss-Kailuweit, F. Dimroth, B. Witzigmann, H. Q. Xu, L. Samuelson, K. Deppert, and M. T. Borgström. 2013. InP nanowire array solar cells achieving 13.8% efficiency by exceeding the ray optics limit. *Science* 339:1057.

Wang, H., and A.-B. Chen. 2000. Calculation of shallow donor levels in GaN. *J. Appl. Phys.* 87:7859.

Wang, J., M. S. Gudiksen, X. Duan, Y. Cui, and C. M. Lieber. 2001. Highly polarized photoluminescence and photodetection from single indium phosphide nanowires. *Science* 293:1455.

Whitaker, J. C. 2005. *Microelectronics*. Boca Raton, FL: Taylor & Francis.

Wijnheijmer, A. P., J. K. Garleff, K. Teichmann, M. Wenderoth, S. Loth, R. G. Ulbrich, P. A. Maksym, M. Roy, and P. M. Koenraad. 2009. Enhanced donor binding energy close to a semiconductor surface. *Phys. Rev. Lett.* 102:166101.

Wölz, M., V. M. Kaganer, O. Brandt, L. Geelhaar, and H. Riechert. 2011. Analyzing the growth of $In_xGa_{1-x}N$/GaN superlattices in self-induced GaN nanowires by x-ray diffraction. *Appl. Phys. Lett.* 98:261907.

Wölz, M., V. M. Kaganer, O. Brandt, L. Geelhaar, and H. Riechert. 2012. Erratum: Analyzing the growth of $In_xGa_{1-x}N$/GaN superlattices in self-induced GaN nanowires by x-ray diffraction. *Appl. Phys. Lett.* 100:179902.

Wölz, M., M. Ramsteiner, V. M. Kaganer, O. Brandt, L. Geelhaar, and H. Riechert. 2013. Strain engineering of nanowire multi-quantum well demonstrated by Raman spectroscopy. *Nano Lett.* 13:4053.

Xiang, H. J., S.-H. Wei, J. L. F. Da Silva, and J. Li. 2008. Strain relaxation and band-gap tunability in ternary $In_xGa_{1-x}N$ nanowires. *Phys. Rev. B* 77:193301.

Yang, L., R. N. Musin, X.-Q. Wang, and M. Y. Chou. 2008. Quantum confinement effect in Si/Ge core–shell nanowires: First-principles calculations. *Phys. Rev. B* 77:195325.

Zeisel, R., M. W. Bayerl, S. T. B. Goennenwein, R. Dimitrov, O. Ambacher, M. S. Brandt, and M. Stutzmann. 2000. DX-behavior of Si in AlN. *Phys. Rev. B* 61:R16283.

Zhang, L., J.-W. Luo, A. Zunger, N. Akopian, V. Zwiller, and J.-C. Harmand. 2010. Wide InP nanowires with wurtzite/zinc-blende superlattice segments are type-II whereas narrower nanowires become type-I: An atomistic pseudopotential calculation. *Nano Lett.* 10:4055.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

13

Quantum Dots

13.1 Introduction 419
13.2 The Core/Shell QD Structure 420
13.3 The **k·p** Method for SP Electronic Structure 421
13.4 PW Implementation of the **k·p** Hamiltonian and Coulomb Integrals for QD Structures 423
13.5 The Effect of Dielectric Confinement 424
13.6 The Configuration Interaction Hamiltonian for NQD 426
13.7 Transport Properties of QD Nanocrystal Arrays 430
13.8 Practical Examples: Cd-Based Chalcogenide QDs 434
Comparison with Experiment: Absorption Edge Wavelength of CdSe/CdTe Type II CQDs • Radiative Lifetimes in CdSe/CdTe CQDs • Correlation Energy • Absorption Spectra of Equivalent CdTe and CdTe/CdSe QD • Charge Carrier Mobility in CdSe QD Arrays
13.9 Conclusion 443

Stanko Tomić
and
Nenad Vukmirović

13.1 Introduction

Semiconductor nanocrystals or quantum dots (QDs) are the subject of intensive research due to a number of novel properties, which make them attractive for both fundamental studies and technological applications [1–6]. QDs are of particular interest for solar cell applications due to their ability to increase efficiency via the generation of multiexcitons from a single photon [7–9]. QDs can be synthesized with a high degree of control using colloidal chemistry [10,11]. Much research effort has been directed toward studying QDs grown from more than one semiconductor, e.g., core/shell heterostructures [12–14]. Such core/shell nanostructures provide means to control the optical properties by tuning the electron–hole wave function overlap, which is affected by the alignment of the conduction band (CB) and valence band (VB) edges, as well as the QD shape and size [15–17]. In addition, such core/shell structures can provide for type-II alignments with staggered CB and VB edges so the lowest energy states for electrons and holes lie in different spatial regions, leading to charge separation between the carriers. Such staggered band alignments have several useful physical consequences, including longer radiative recombination times for more efficient charge extraction in photovoltaic applications [18,19], optical gaps that can be made smaller than the bulk values of constituent materials [12,20,21] and control of the electron–hole wave function overlap that determines the exchange interaction energy [22]. Charge separation in type-II QDs can also be used to increase the repulsion between like-sign charges in biexciton states [23,24], leading to the possibility of lasing in the single exciton regime [6,25,26].

To determine the energetics of many-body states in QDs both the confinement potential and many-body interactions between the carriers need to be taken into account. Many-body interactions lead to Coulomb (charge) and Fermi (spin) correlation. Coulomb correlation arises from the electrostatic interaction of charge carriers in the many-body complex, while spin correlation occurs due to the fermionic character of charge carriers (i.e., the Pauli exclusion principle) [27]. Correlated many-body states may be calculated with the configuration interaction (CI) method, which can be used in the framework of continuum or atomistic descriptions of single-particle (SP) states. [28–38].

Colloidal QDs are usually embedded or dispersed in media [39] of lower dielectric constant than the semiconductor itself—this dielectric confinement leads to a modification of the Coulomb interaction, which can be described using classical image charge theory. While atomistic calculations [40] showed that dielectric confinement significantly affects the charging energies of QDs, in single-material spherical QDs the similar electron and hole charge distributions lead to a weakened dielectric confinement effect [41] on exciton states, which mainly increases the binding energy [28,42]. It is, therefore, natural to wonder if the optical properties of spherical type-II core/shell QDs can be significantly affected by the dielectric environment.

The effect of dielectric confinement and many-electron correlation means that the single-particle picture is not good enough to faithfully predict exciton energetics or wave functions in colloidal QDs. The proper treatment of charges requires a many-electron description that goes beyond mean-field theories, and the CI method is the most appropriate approach. However, the full configuration method becomes progressively computationally expensive as the number of states increases. Luckily, however, the interpretation of physical experiments often requires detailed knowledge of just a few excitons of particular symmetry. This allows us to evaluate the effect of correlation on exciton energies and dipole matrix elements for arbitrary QD geometries.

In this chapter, we review the theoretical methods for the description of the electronic structure and transport properties of colloidal nanocrystal quantum dots (NQD). The chapter is organized as follows: after introducing the core/shell morphology in Section 13.2, in Section 13.3, we review the widely used **k·p** theory for electronic structure of semiconductor nanostructures, both the 8-band and 14-band version, taking into account effects like band-mixing, strain, piezoelectric field, and spin–orbit interaction. In Section 13.4, we outline the plane wave (PW) basis set implementation of the **k·p** Hamiltonian and necessary modifications for QDs, i.e., three-dimensional (3D) confined systems. In Section 13.5, we discuss the effect of dielectric confinement, polarization, and self-polarization due to large contrast and spatial variation of the dielectric constant in colloidal NQD. In Section 13.6, we derive the CI Hamiltonian for the description of many-electron states in colloidal NQD. In Section 13.7, we present the basic theory of the electron transport through NQD networks. In Section 13.8, we present our results on the ground state exciton absorption edge, radiative times of the ground state exciton, and correlation energies of the ground state exciton, as well as electronic charge transport in NQD solids.

13.2 The Core/Shell QD Structure

The core/shell QD structures are comprised of a single-crystal "core" region enveloped by a second "shell" region of a different semiconductor material. The shell region usually possesses a larger band gap than the core region. In such a configuration, with appropriate materials combinations and alloys, such QD can exhibit a type-II band alignment, thus offering an extra degree of freedom in *quantum engineering* of nanostructures for specific purposes. The addition of the shell has two main effects. First, the shell serves to passivate dangling bonds at the surface of the dot (surface ligands), which arise from the lack of coordination in atoms at the dot surface, vastly reducing the number of trap states near the surface. This reduces the effect of the nonradiative processes on device operations. Second, the energy offsets forming the type-II heterostructure result in confinement of the charge carriers into separate regions of the QD: electrons are confined to the core, while holes are confined to the shell region, or vice versa, depending on the relative band alignment between the two regions. Such staggered band alignments have several useful physical consequences, including longer radiative recombination times for more efficient charge extraction in photovoltaic applications and optical gaps that can be made smaller than the bulk values of the constituent materials. To illustrate the problem and introduce basic parameters, in Figure 13.1 we show

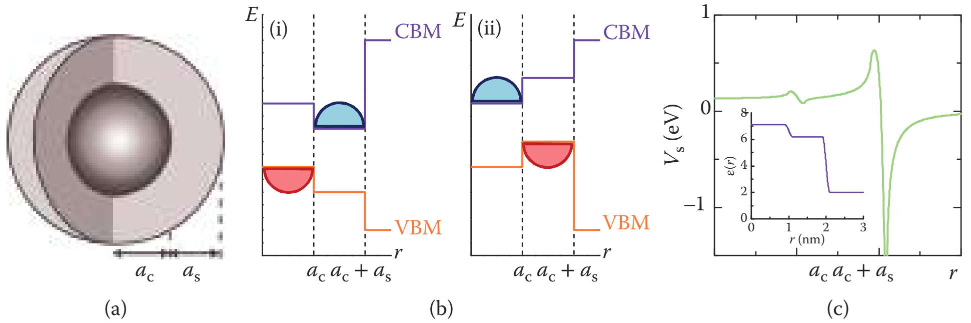


FIGURE 13.1 (a) Cutaway view of a spherical core/shell quantum dot (QD) characterized by core radius a_c and shell thickness a_s; (b) conduction band minimum (CBM) and the valence band maximum (VBM) corresponding to type-II alignment, in (i) an h/e CdTe/CdSe QD and (ii) an e/h CdSe/CdTe QD.

(a) a schematic of a spherical core/shell QD characterized by core radius a_c and shell thickness a_s, the staggered alignments of the CB minimum (CBM) and VB maximum (VBM) in type-II structures where (b,i) holes (h) are confined in the core and electrons (e) are confined in the shell, as in CdTe/CdSe QDs and (b,ii) electrons are confined in the core and holes are confined in the shell, as in CdSe/CdTe QDs.

13.3 The k•p Method for SP Electronic Structure

The quantum mechanical description of electrons in any material requires detailed knowledge of their wave functions, $\psi_{n,\mathbf{k}}(\mathbf{r})$, which are found by solving the Schrödinger equation (in the single-electron approximation):

$$H_0\psi_{n,\mathbf{k}}(\mathbf{r}) = E_{n,\mathbf{k}}\psi_{n,\mathbf{k}}(\mathbf{r}). \tag{13.1}$$

The Hamiltonian in Equation 13.1, $H_0 = p^2/2m_0 + V(\mathbf{r})$, is the function of the quantum mechanical momentum operator, $\mathbf{p} = -i\hbar\nabla$, and the crystal potential experienced by electrons, $V(\mathbf{r}) = V(\mathbf{r}+\mathbf{R})$, which is a periodic function, with the periodicity of the crystal lattice, $\mathbf{R}$, and m_0 is the free electron mass. According to Bloch's theorem, the solutions to this Schrödinger equation can be written as $\psi_{n,\mathbf{k}}(\mathbf{r}) = e^{i\mathbf{k}\cdot\mathbf{r}}u_{n,\mathbf{k}}(\mathbf{r})$ where $\mathbf{k}$ is the electron wave vector, n is the band index, and $u_{n,\mathbf{k}}$ is the *cell-periodic* function, with the same periodicity as the crystal lattice. The cell-periodic function, $u_{n,\mathbf{k}}$, satisfies the equation

$$H_{\mathbf{k}}u_{n,\mathbf{k}} = E_{n,\mathbf{k}}u_{n,\mathbf{k}}, \tag{13.2}$$

where the Hamiltonian

$$H_{\mathbf{k}} = H_0 + H'_{\mathbf{k}} = \frac{p^2}{2m_0} + V + \frac{\hbar^2k^2}{2m_0} + \frac{\hbar\mathbf{k}\cdot\mathbf{p}}{m_0} \tag{13.3}$$

is given as a sum of two terms: the unperturbed, H_0, which in fact equals the exact Hamiltonian at $\mathbf{k} = 0$ (i.e., at the Γ point in the Brillouin zone) and the "perturbation," $H'_{\mathbf{k}}$. The Hamiltonian given by Equation 13.3 is called the **k**·**p** Hamiltonian [43–45]. If the eigenvalues are nondegenerate, the first-order energy correction is given by $E'_{n,\mathbf{k}} \approx \langle u_{n,0}|H'_{\mathbf{k}}|u_{n,0}\rangle$ and there is no correction (to first order, in the absence of nondiagonal matrix elements) in the eigenfunctions. The second-order correction arises from nondiagonal terms; the energy correction is given by $E''_{n,\mathbf{k}} \approx \sum_{n'\neq n}|\langle u_{n,0}|H'_{\mathbf{k}}|u_{n',0}\rangle|^2/(E_{n,0} - E_{n',0})$. The "perturbation" term $H'_{\mathbf{k}}$ gets progressively smaller as $\mathbf{k}$ approaches zero. Therefore, the **k**·**p** perturbation theory is most accurate for small values of **k**. However, if enough terms are included in the perturbative expansion, then the theory can in fact be reasonably accurate for any value of **k** in the entire Brillouin zone [46,47]. For a band n, with an extremum at $\mathbf{k} = 0$, and with no spin–orbit coupling, the result of **k·p** perturbation theory is (to the lowest nontrivial order):

$$u_{n,\mathbf{k}} = u_{n,0} + \frac{\hbar}{m_0} \sum_{n' \neq n} \frac{\langle u_{n,0} | \mathbf{k} \cdot \mathbf{p} | u_{n',0} \rangle}{E_{n,0} - E_{n',0}} u_{n',0} \tag{13.4}$$

$$E_{n,\mathbf{k}} = E_{n,0} + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar^2}{m_0^2} \sum_{n' \neq n} \frac{|\langle u_{n,0} | \mathbf{k} \cdot \mathbf{p} | u_{n',0} \rangle|^2}{E_{n,0} - E_{n',0}}. \tag{13.5}$$

The parameters required to do these calculations, the band edge energies, $E_{n,0}$, and the optical matrix elements, $\langle u_{n,0} | \mathbf{p} | u_{n',0} \rangle$, are typically inferred from experimental data or detailed atomistic-based theories.

A particular strength of the **k·p** theory is a straightforward inclusion of the spin–orbit interaction and of the strain effects on the band structure via deformation potential theory [48].

Relativistic effects in the **k·p** method are included also perturbatively via the spin–orbit (SO) interaction Hamiltonian, $H_{SO} = \frac{\hbar}{4m_0^2 c^2} \left[\nabla V(\mathbf{r}) \times \mathbf{p} \right] \cdot \vec{\sigma}$, where $\vec{\sigma} = (\sigma_x, \sigma_y, \sigma_z)$ is a vector consisting of the three Pauli spin matrices. Finally, the **k·p** Hamiltonian becomes

$$H_{\mathbf{k}} = \frac{p^2}{2m_0} + V + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar \mathbf{k} \cdot \mathbf{p}}{m_0} + \frac{\hbar}{4m_0^2 c^2} (\vec{\sigma} \times \nabla V) \cdot (\hbar \mathbf{k} + \mathbf{p}). \tag{13.6}$$

In a strained system, the lattice constants are stretched or compressed [48]. This affects the SP potential experienced by electrons and consequently yields an additional term in the **k·p** Hamiltonian. Taking into account only the effects up to first order in strain tensor ϵ, the strained **k·p** Hamiltonian becomes

$$H_{\mathbf{k}} = \frac{p^2}{2m_0} + V + \frac{\hbar^2 k^2}{2m_0} + \frac{\hbar \mathbf{k} \cdot \mathbf{p}}{m_0} + \frac{1}{4m_0^2 c^2} (\vec{\sigma} \times \nabla V) \cdot (\mathbf{k} + \mathbf{p}) + \mathbf{D}^{\epsilon} \cdot \boldsymbol{\epsilon}, \tag{13.7}$$

where $\mathbf{D}^{\epsilon}$ is the deformation potential operator. This Hamiltonian can be subjected to the same sort of perturbation theory analysis as above [49].

The most widely used version of the **k·p** Hamiltonian is the 8-band model [50], in which the operator given in Equation 13.7 is represented in the basis of s-bonding and p-antibonding states around the energy gap:

$\lvert u_1\rangle = \lvert \frac{1}{2}, +\frac{1}{2}\rangle = \lvert s; \uparrow\rangle$	$\lvert u_5\rangle = \lvert \frac{1}{2}, -\frac{1}{2}\rangle = -\lvert s; \downarrow\rangle$
$\lvert u_2\rangle = \lvert \frac{3}{2}, +\frac{3}{2}\rangle = \frac{i}{\sqrt{2}}[\lvert x; \uparrow\rangle + i\lvert y; \uparrow\rangle]$	$\lvert u_6\rangle = \lvert \frac{3}{2}, -\frac{3}{2}\rangle = -\frac{i}{\sqrt{2}}[\lvert x; \downarrow\rangle - i\lvert y; \downarrow\rangle]$
$\lvert u_3\rangle = \lvert \frac{3}{2}, +\frac{1}{2}\rangle = \frac{i}{\sqrt{6}}[\lvert x; \downarrow\rangle + i\lvert y; \downarrow\rangle - 2\lvert z; \uparrow\rangle]$	$\lvert u_7\rangle = \lvert \frac{3}{2}, -\frac{1}{2}\rangle = +\frac{i}{\sqrt{6}}[\lvert x; \uparrow\rangle - i\lvert y; \uparrow\rangle + 2\lvert z; \downarrow\rangle]$
$\lvert u_4\rangle = \lvert \frac{1}{2}, +\frac{1}{2}\rangle = \frac{i}{\sqrt{3}}[\lvert x; \downarrow\rangle + i\lvert y; \downarrow\rangle + \lvert z; \uparrow\rangle]$	$\lvert u_8\rangle = \lvert \frac{1}{2}, -\frac{1}{2}\rangle = +\frac{i}{\sqrt{3}}[\lvert x; \uparrow\rangle - i\lvert y; \uparrow\rangle - \lvert z; \downarrow\rangle]$

Occasionally, the 14-band **k·p** Hamiltonian is also used, in which one exploits the basis of p-bonding, s-bonding, and p-antibonding states around the energy gap [51]:

$\lvert u_1\rangle = \lvert \frac{3}{2},+\frac{3}{2}\rangle = \frac{i}{\sqrt{2}}\lvert(x_a+iy_a)\uparrow\rangle$	$\lvert u_8\rangle = \lvert \frac{3}{2},-\frac{3}{2}\rangle = -\frac{i}{\sqrt{2}}\lvert(x_a-iy_a)\downarrow\rangle$
$\lvert u_2\rangle = \lvert \frac{3}{2},+\frac{1}{2}\rangle = \frac{i}{\sqrt{6}}[\lvert(x_a+iy_a)\downarrow\rangle - 2\lvert z_a\uparrow\rangle]$	$\lvert u_9\rangle = \lvert \frac{3}{2},-\frac{1}{2}\rangle = +\frac{i}{\sqrt{6}}[\lvert(x_a-iy_a)\uparrow\rangle + 2\lvert z_a\downarrow\rangle]$
$\lvert u_3\rangle = \lvert \frac{1}{2},+\frac{1}{2}\rangle = \frac{i}{\sqrt{3}}[\lvert(x_a+iy_a)\downarrow\rangle + \lvert z_a\uparrow\rangle]$	$\lvert u_{10}\rangle = \lvert \frac{1}{2},-\frac{1}{2}\rangle = +\frac{i}{\sqrt{3}}[\lvert(x_a-iy_a)\uparrow\rangle - \lvert z_a\downarrow\rangle]$
$\lvert u_4\rangle = \lvert \frac{1}{2},+\frac{1}{2}\rangle = \lvert s_a\uparrow\rangle$	$\lvert u_{11}\rangle = \lvert \frac{1}{2},-\frac{1}{2}\rangle = -\lvert s_a\downarrow\rangle$
$\lvert u_5\rangle = \lvert \frac{3}{2},+\frac{3}{2}\rangle = \frac{i}{\sqrt{2}}\lvert(x_b+iy_b)\uparrow\rangle$	$\lvert u_{12}\rangle = \lvert \frac{3}{2},-\frac{3}{2}\rangle = -\frac{i}{\sqrt{2}}\lvert(x_b-iy_b)\downarrow\rangle$
$\lvert u_6\rangle = \lvert \frac{3}{2},+\frac{1}{2}\rangle = \frac{i}{\sqrt{6}}[\lvert(x_b+iy_b)\downarrow\rangle - 2\lvert z_b\uparrow\rangle]$	$\lvert u_{13}\rangle = \lvert \frac{3}{2},-\frac{1}{2}\rangle = +\frac{i}{\sqrt{6}}[\lvert(x_b-iy_b)\uparrow\rangle + 2\lvert z_b\downarrow\rangle]$
$\lvert u_7\rangle = \lvert \frac{1}{2},+\frac{1}{2}\rangle = \frac{i}{\sqrt{3}}[\lvert(x_b+iy_b)\downarrow\rangle + \lvert z_b\uparrow\rangle]$	$\lvert u_{14}\rangle = \lvert \frac{1}{2},-\frac{1}{2}\rangle = +\frac{i}{\sqrt{3}}[\lvert(x_b-iy_b)\uparrow\rangle - \lvert z_b\downarrow\rangle]$

13.4 PW Implementation of the k•p Hamiltonian and Coulomb Integrals for QD Structures

The QD as a three-dimensional (3D) object breaks the translational symmetry of the bulk material along all three Cartesian directions implying operator replacement $k_\nu \rightarrow -i\partial/\partial\nu$ in Equation 13.6, where $\nu = (x, y, z)$. To solve the multiband system of Schrödinger equations, Equation 13.6, the PW methodology is employed as an expansion method [33,52–56]. In the PW representation, the eigenvalues (E_i) and the coefficients $[A^{(i)}_{n,\mathbf{k}}]$ of the ith eigenvector $[\psi^{(i)}_n(\mathbf{r}) = \sum_\mathbf{k} A^{(i)}_{n,\mathbf{k}} e^{i\mathbf{kr}}]$, are linked by the relation

$$\sum_{n,\mathbf{k}'} h_{m,n}(\mathbf{k},\mathbf{k}')A^{(i)}_{n,\mathbf{k}'} = E_\mathrm{i}A^{(i)}_{m,\mathbf{k}}, \tag{13.8}$$

where $h_{m,n}(\mathbf{k},\mathbf{k}')$ are the Hamiltonian matrix elements in the PW basis, and $m, n \in \{1, ..., 8\}$ or $m, n \in \{1, ..., 14\}$ are the band indices in the **k·p** Hamiltonian. All the elements in the Hamiltonian matrix, Equation 13.8, can be expressed as a linear combination of different kinetic and strain related terms and their convolution with the characteristic function of the QD shape, $\chi_\mathrm{qd}(\mathbf{k})$ [33,57]. The whole **k**-space is discretized by embedding the QD in a rectangular box of dimensions L_x, L_y, and L_z and volume $\Omega = L_x \times L_y \times L_z$ and choosing the **k**-vectors in the form $\mathbf{k} = 2\pi(n_x/L_x, n_y/L_y, n_z/L_z)$, where n_x, n_y, and n_z are integers. Maximal absolute values of these integers control the accuracy of the method.

Having determined PW expansion coefficients, $A^{(i)}_{n,\mathbf{k}}$, from Equation 13.8, we can use the single-particle wave functions, $\psi^{(i)}_n(\mathbf{r})$, to calculate the Coulomb integrals, V_{ijkl}, relevant for many-electron processes that occur in QDs. The Coulomb integral among states i, j, k, and l is defined as

$$V_{ijkl} = \sum_{b=1}^{N_b}\sum_{b'=1}^{N_b}\int_\Omega d^3\mathbf{r}\int_\Omega d^3\mathbf{r}'\psi^{(i)}_b(\mathbf{r})^*\psi^{(j)}_b(\mathbf{r})V(|\mathbf{r}-\mathbf{r}'|)\psi^{(k)}_{b'}(\mathbf{r}')^*\psi^{(l)}_{b'}(\mathbf{r}'), \tag{13.9}$$

where $V(\mathbf{u}) = e^2/4\pi\epsilon\mathbf{u}$, with ϵ being the static dielectric constant, b is the band index, and N_b is the number of bands, i.e., 8 or 14 in the **k·p** Hamiltonian. The integral 13.9 can be rewritten as

$$V_{ijkl} = \int_\Omega d^3\mathbf{r}\int_\Omega d^3\mathbf{r}' B_{ij}(\mathbf{r})V(|\mathbf{r}-\mathbf{r}'|)B_{kl}(\mathbf{r}'), \tag{13.10}$$

where we introduce

$$B_{ij}(\mathbf{r}) = \sum_{b=1}^{N_b}\psi^{(i)}_b(\mathbf{r})^*\psi^{(j)}_b(\mathbf{r}). \tag{13.11}$$

The integral 13.10 can be calculated by performing a six-dimensional integration in real space, which is, however, numerically very demanding, especially when a large number of integrals has to be calculated, as in the CI approach [33,58]. One of the advantages of the PW method is that relevant physical quantities can be expressed in terms of the coefficients in the PW expansion, i.e., by the Fourier transform (FT). The last statement holds in principle when the evaluation of Coulomb integrals is concerned, since the FT of the Coulomb potential is fully analytical. This can be shown by the following set of transformations leading to Equation 13.16. Defining the PW expansion of $B_{ij}(\mathbf{r})$

$$B_{ij}(\mathbf{r}) = \sum_{\mathbf{k}} B_{ij}(\mathbf{k}) e^{i\mathbf{k}\cdot\mathbf{r}} \tag{13.12}$$

and putting the last expression into Equation 13.10 one obtains

$$V_{ijkl} = \sum_{\mathbf{k}} B_{ij}(\mathbf{k}) \sum_{\mathbf{k}'} B_{kl}(\mathbf{k}') \int_{\Omega} d^3\mathbf{r} \int_{\Omega} d^3\mathbf{r}' e^{i\mathbf{k}\cdot\mathbf{r}} V(|\mathbf{r}-\mathbf{r}'|) e^{i\mathbf{k}'\cdot\mathbf{r}'}. \tag{13.13}$$

Using the properties of convolution and Parseval's theorem in inverse space, the $B_{ij}(\mathbf{k})$ can be expressed in terms in the coefficients in the PW expansion of the envelope functions as

$$B_{ij}(\mathbf{k}) = \sum_{b=1}^{N_b} \sum_{\mathbf{q}} A^{(i)*}_{b,\mathbf{q}} A^{(j)}_{b,\mathbf{q}+\mathbf{k}}. \tag{13.14}$$

By introducing the approximation changing the domain of integration in one of the integrals in Equation 13.13 from Ω to the whole space (which is valid when Ω is large enough) one gets after the replacement of variables from $\mathbf{r}$ and $\mathbf{r}'$, to $\mathbf{r}$ and $\mathbf{u} = \mathbf{r} - \mathbf{r}'$

$$V_{ijkl} = \sum_{\mathbf{k}} B_{ij}(\mathbf{k}) \sum_{\mathbf{k}'} B_{kl}(\mathbf{k}') [\int_{\Omega} d^3\mathbf{r} e^{i\mathbf{k}\cdot\mathbf{r}} e^{i\mathbf{k}'\cdot\mathbf{r}}][\int d^3\mathbf{u} V(|\mathbf{u}|) e^{-i\mathbf{k}'\cdot\mathbf{u}}]. \tag{13.15}$$

Using the relations

$$\int d^3\mathbf{u} e^{-i\mathbf{k}'\mathbf{u}} V(|\mathbf{u}|) = \frac{e^2}{\epsilon k'^2}$$

and

$$\frac{1}{\Omega} \int_{\Omega} d^3\mathbf{r} e^{i(\mathbf{k}+\mathbf{k}')\mathbf{r}} = \delta_{\mathbf{k}+\mathbf{k}',0}$$

one obtains

$$V_{ijkl} = \Omega \sum_{\substack{\mathbf{k} \\ \mathbf{k}\neq 0}} B_{ij}(\mathbf{k}) B_{kl}(-\mathbf{k}) \frac{e^2}{\epsilon k^2}. \tag{13.16}$$

13.5 The Effect of Dielectric Confinement

For colloidal QDs, the dielectric constant ϵ of the QD material is typically much larger than that of the surrounding medium, and the spatial variation $\epsilon = \epsilon(\mathbf{r})$ cannot be ignored. Such dielectric contrast means that any free charge in the QD induces polarization charge in the QD and its surroundings. As a consequence, the Coulomb energy for the electron–hole pair for a system with a spatially varying dielectric constant is

$$V(\mathbf{r}_e, \mathbf{r}_h) = V_{\mathrm{c}}(\mathbf{r}_e, \mathbf{r}_h) + V_{\mathrm{s}}(\mathbf{r}_e) + V_{\mathrm{s}}(\mathbf{r}_h), \tag{13.17}$$

where V_{c} is the interparticle Coulomb potential and V_{s} is the self-polarization potential. The potential, $V_{\mathrm{c}}(\mathbf{r}_e, \mathbf{r}_h)$, encompasses both the "direct" interparticle Coulomb interaction and the interface polarization potential due to the interaction between a real particle and the induced charge of the other particle. The self-polarization potential $V_{\mathrm{s}}(\mathbf{r})$ arises from the interaction of a particle and its own induced charge.

The interparticle Coulomb potential in real space:

$$V_{\mathrm{c}}(\mathbf{r}_e, \mathbf{r}_h) = \frac{1}{\epsilon(\mathbf{r}_e, \mathbf{r}_h)} \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}_e - \mathbf{r}_h|}, \tag{13.18}$$

under the assumption, $\epsilon(\mathbf{r}_e, \mathbf{r}_h) \approx \epsilon(\mathbf{r}_e - \mathbf{r}_h) = \epsilon(\mathbf{r})$ [59], can be recast in the form of product of two functions

$$V_{\mathrm{c}}(\mathbf{r}) = \frac{1}{\epsilon(\mathbf{r})} \frac{e^2}{4\pi\epsilon_0 |\mathbf{r}|} = \frac{1}{\epsilon(\mathbf{r})} U(\mathbf{r}), \tag{13.19}$$

where $U(\mathbf{r}) = e^2/4\pi\epsilon_0 \mathbf{r}$ is the bare Coulomb potential due to charge e, with the well-known FT in the inverse space: $\mathcal{F}[U(\mathbf{r})] = \tilde{U}(\mathbf{k}) = e^2/\epsilon_0 k^2$. Considering stepwise uniform regions where $\epsilon(r) = \epsilon_m =$ const., and m is the index of the region, and spherical QD shape, the general expression, Equation 13.18, acquires the simple analytical form:

$$\tilde{V}_{\mathrm{c}}(\mathbf{k}) = \frac{e^2}{\epsilon_0 k^2} \left[\frac{1}{\epsilon_{\mathrm{c}}} + \left(\frac{1}{\epsilon_{\mathrm{s}}} - \frac{1}{\epsilon_{\mathrm{c}}} \right) \cos(kR_{\mathrm{c}}) + \left(\frac{1}{\epsilon_{\mathrm{coll.}}} - \frac{1}{\epsilon_{\mathrm{s}}} \right) \cos(kR_{\mathrm{s}}) \right], \tag{13.20}$$

where ϵ_{c}, ϵ_{s}, and $\epsilon_{\mathrm{coll.}}$ are the dielectric constants of the QD's core, shell, and surrounding colloid respectively, and R_{c} and R_{s} are the radius of core and outer radius of the shell, respectively.

Calculation of the self-polarization potential for the arbitrary QD shape in the inverse space is a bit involved. Alternatively, for several characteristic QD shapes, like spherical QD [60], spherical core/shell QD [61,62], elliptical QD [63], elliptical core/shell QD [64–66], and cuboidal QD [67], there are available expressions for $V_{\mathrm{s}}(\mathbf{r})$ in real space. Here, we list a general expression for the self-polarization potential of spherical QD with an arbitrary number, m, of core/shells:

$$V_{\mathrm{s}}^{m}(r) = \frac{e^2}{8\pi\epsilon_0\epsilon_m} \sum_{l=0}^{\infty} \frac{1}{1 - p_{m,l}q_{m,l}} \left[p_{m,l} r^{2l} + p_{m,l}q_{m,l} r^{-1} + q_{m,l} r^{-2(l+1)} \right], \tag{13.21}$$

where $p_{m,l}$ and $q_{m,l}$ are recursive coefficients that depend on the dielectric constants, ϵ_m, and radii, R_m, in the region m, of the core/shell structures (see Ref. [61]). Such expressions are then simple to FT into $\tilde{V}_{\mathrm{s}}(\mathbf{k})$ and are used when calculating the Coulomb matrix elements, V_{ijkl} for the CI Hamiltonian; see Section 13.6. The expression for the contribution of the direct Coulomb, Equation 13.20, and self-polarization energy, Equation 13.21, to the total Coulomb integral reads as

$$V_{ijkl}^{(c)} = \Omega \sum_{\substack{\mathbf{k} \\ \mathbf{k} \neq 0}} B_{ij}(\mathbf{k}) B_{kl}(-\mathbf{k}) \tilde{V}_{\mathrm{c}}(\mathbf{k}) \tag{13.22}$$

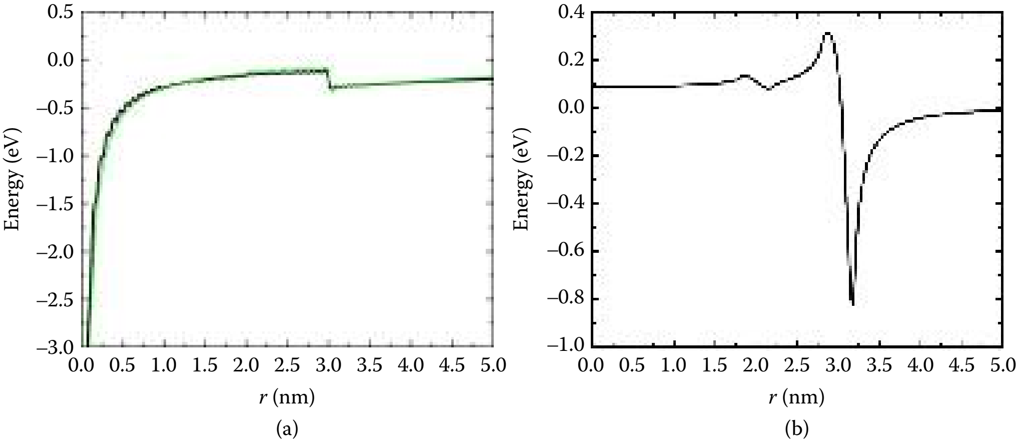


FIGURE 13.2 Variation of the interparticle Coulomb potential (a), $V_c(\mathbf{r})$, and self-polarization potential (b), $V_s(\mathbf{r})$, of CdSe/CdTe core shell quantum dot (QD) with $\epsilon_c = 7.1$, $\epsilon_s = 6.2$, $\epsilon_{\text{coll.}} = 2$, and $R_c = 2$ nm and $R_s = 3$ nm. In (a), the tiny bright line represents analytic $V_c(r)$ obtained by Equation 13.19, while solid thick line represents the same $V_c(r)$ recovered from the analytical expression in the inverse space, Equation 13.20.

and

$$V_{ijkl}^{(s)} = \Omega \sum_{\mathbf{k}} [B_{ij}(\mathbf{k})\delta_{kl} + \delta_{ij}B_{kl}(\mathbf{k})]\tilde{V}_s(\mathbf{k}). \tag{13.23}$$

In colloidal core/shell QDs, the self-polarization potential is characterized by a small peak and well near the core/shell interface due to the small dielectric mismatch between the core and shell materials. However, a much larger peak just inside $r = R_c + R_s$ and a deep well slightly outside the QD are due to the far greater dielectric mismatch of the shell and matrix material. In Figure 13.2, we present the radial variation of $V_c(\mathbf{r})$ and $V_s(\mathbf{r})$ of CdSe/CdTe core shell QD with $\epsilon_c = 7.1$, $\epsilon_s = 6.2$, $\epsilon_{\text{coll.}} = 2$, and $R_c = 2$ nm and $R_s = 3$ nm.

13.6 The Configuration Interaction Hamiltonian for NQD

In this section, we derive the Hamiltonian describing interacting electrons and holes in QDs. This Hamiltonian can be used to calculate neutral and charged excitonic and multiexcitonic states that are populated upon photoexcitation of the QD.

We start with a general Hamiltonian describing interacting electrons in the solid

$$H = H_s + H_{\text{int}}, \tag{13.24}$$

where the noninteracting part of the Hamiltonian is

$$H_s = \int d^3x\, \psi^\dagger(x) \left[-\frac{\hbar^2}{2m_0}\nabla^2 + V(x)\right] \psi(x)\, d^3x \tag{13.25}$$

and the interaction part is

$$H_{\text{int}} = \frac{1}{2}\int d^3x\, d^3y\, \psi^\dagger(x)\psi^\dagger(y)\, V_{\text{int}}(x,y)\, \psi(y)\, \psi(x). \tag{13.26}$$

In previous equations, $\psi(x)$ is the fermionic field operator, $\psi^{\dagger}(x)$ is its adjoint operator, V is the potential of atomic cores in the solid, and $V_{\text{int}}(x,y)$ is the interaction potential between the particles at x and y, which for the bare Coulomb interaction reads $V_{\text{int}}(x,y) = \frac{e^2}{4\pi\epsilon_0|x-y|}$.

Next, we express the field operator in terms of annihilation operators as

$$\psi(x) = \sum_{\text{v}\in\text{VB}} a_{\text{v}}\phi_{\text{v}}(x) + \sum_{\text{c}\in\text{CB}} a_{\text{c}}\phi_{\text{c}}(x), \tag{13.27}$$

where $\phi_{\text{c}}(x)$ and $\phi_{\text{v}}(x)$ are the SP wave function for states in the CB and VB, while a_{c} and a_{v} are the corresponding electron annihilation operators. The operator a_{c} acts on arbitrary many-body state of the system by destroying an electron in state c and a_{v} acts by destroying an electron in state v. Their adjoint operators $a_{\text{c}}^{\dagger}$ and $a_{\text{v}}^{\dagger}$ act by creating an electron in the corresponding state. After substitution of Equation 13.27 into Equations 13.25 and 13.26, one arrives at the expressions

$$H_{\text{s}} = \sum_{j_1 j_2} a_{j_1}^{\dagger} a_{j_2} \int d^3x\, \phi_{j_1}^{*}(x)\left[-\frac{\hbar^2}{2m_0}\nabla^2 + V(x)\right]\phi_{j_2}(x) \tag{13.28}$$

$$H_{\text{int}} = \frac{1}{2}\sum_{j_1 j_2 j_3 j_4} a_{j_1}^{\dagger} a_{j_2}^{\dagger} a_{j_3} a_{j_4} \int d^3x\, d^3y\, \phi_{j_1}^{*}(x)\,\phi_{j_4}(x)\, V_{\text{int}}(x,y)\,\phi_{j_2}^{*}(y)\,\phi_{j_3}(y), \tag{13.29}$$

i.e.,

$$H = \sum_{j_1 j_2} v_{j_1 j_2} a_{j_1}^{\dagger} a_{j_2} + \frac{1}{2}\sum_{j_1 j_2 j_3 j_4} V_{j_1 j_4 j_2 j_3} a_{j_1}^{\dagger} a_{j_2}^{\dagger} a_{j_3} a_{j_4}, \tag{13.30}$$

where

$$v_{j_1 j_2} = \int d^3x\, \phi_{j_1}^{*}(x)\left[-\frac{\hbar^2}{2m_0}\nabla^2 + V(x)\right]\phi_{j_2}(x) \tag{13.31}$$

$$V_{j_1 j_4 j_2 j_3} = \int d^3x\, d^3y\, \phi_{j_1}^{*}(x)\,\phi_{j_4}(x)\, V_{\text{int}}(x,y)\,\phi_{j_2}^{*}(y)\,\phi_{j_3}(y). \tag{13.32}$$

The summations in Equation 13.30 should in principle include all SP states in the valence and the CB. However, we are interested mainly in the lowest energy excitations of the system where electrons at the bottom of the CB and holes at the top of the VB are created. To study these excitations, it is more convenient to use electron and hole creation and annihilation operators $e_j(e_j^{\dagger})$ and $h_j(h_j^{\dagger})$ defined as $e_j = a_j$, when $j \in \text{CB}$ and $h_j = a_j^{\dagger}$ when $j \in \text{VB}$. These operators satisfy the following fermionic commutation rules: $\{e_{\text{i}}, e_j^{\dagger}\} = \delta_{ij}$, $\{h_{\text{i}}, h_j^{\dagger}\} = \delta_{ij}$, $\{e_{\text{i}}, e_j\} = 0$, $\{h_{\text{i}}, h_j\} = 0$, $\{e_{\text{i}}^{\dagger}, h_j\} = 0$, $\{e_{\text{i}}, h_j^{\dagger}\} = 0$, where $\{a, b\} = ab + ba$ denotes the anticommutator of two operators. The H_{s} term then reads

$$H_{\text{s}} = \sum_{j_1 j_2 \in \text{CB}} v_{j_1 j_2} e_{j_1}^{\dagger} e_{j_2} - \sum_{j_1 j_2 \in \text{VB}} v_{j_2 j_1} h_{j_1}^{\dagger} h_{j_2} + \sum_{j \in \text{VB}} v_{jj}, \tag{13.33}$$

where the last term contributes an irrelevant constant shift and can be omitted from further consideration. To obtain the H_{int} term, we first introduce an approximation to consider only the terms in H_{int} that contain the same number of electron creation and annihilation operators, as well as the same number of hole creation and annihilation operators. The other terms that do not conserve the number of pair excitations give rise to the polarization of the electronic orbitals, and their main effects can be taken into account by a dielectric constant in the Coulomb potential, which can be inserted into the final result [68].

Under such an assumption the H_{int} term consists of three sets of terms $H_{\text{int}} = H_{\text{int}}^{(1)} + H_{\text{int}}^{(2)} + H_{\text{int}}^{(3)}$. The first set of these terms (when $j_1, j_2, j_3, j_4 \in \text{CB}$) reads:

$$H_{\text{int}}^{(1)} = \frac{1}{2} \sum_{j_1 j_2 j_3 j_4} V_{j_1 j_4 j_2 j_3} e_{j_1}^{\dagger} e_{j_2}^{\dagger} e_{j_3} e_{j_4}. \tag{13.34}$$

The second set of these terms $H_{\text{int}}^{(2)}$ that is obtained when $j_1, j_2, j_3, j_4 \in \text{VB}$ contains the term $h_{j_1} h_{j_2} h_{j_3}^{\dagger} h_{j_4}^{\dagger}$. By exploiting the fermionic commutation relations, this term can be rewritten as

$$h_{j_1} h_{j_2} h_{j_3}^{\dagger} h_{j_4}^{\dagger} = \delta_{j_2 j_3} \delta_{j_1 j_4} - \delta_{j_1 j_3} \delta_{j_2 j_4} - \delta_{j_2 j_3} h_{j_4}^{\dagger} h_{j_1} - \delta_{j_1 j_4} h_{j_3}^{\dagger} h_{j_2} + \delta_{j_1 j_3} h_{j_4}^{\dagger} h_{j_2} + \delta_{j_2 j_4} h_{j_3}^{\dagger} h_{j_1} + h_{j_3}^{\dagger} h_{j_4}^{\dagger} h_{j_1} h_{j_2}. \tag{13.35}$$

The $H_{\text{int}}^{(2)}$ term is then given as

$$\begin{aligned} H_{\text{int}}^{(2)} = {} & \frac{1}{2} \sum_{j_1 j_2} V_{j_1 j_1 j_2 j_2} - \frac{1}{2} \sum_{j_1 j_2} V_{j_1 j_2 j_2 j_1} - \frac{1}{2} \sum_{j_1 j_2 j_4} V_{j_1 j_4 j_2 j_2} h_{j_4}^{\dagger} h_{j_1} - \frac{1}{2} \sum_{j_1 j_2 j_3} V_{j_1 j_1 j_2 j_3} h_{j_3}^{\dagger} h_{j_2} \\ & + \frac{1}{2} \sum_{j_1 j_2 j_4} V_{j_1 j_4 j_2 j_1} h_{j_4}^{\dagger} h_{j_2} + \frac{1}{2} \sum_{j_1 j_2 j_3} V_{j_1 j_2 j_2 j_3} h_{j_3}^{\dagger} h_{j_1} + \frac{1}{2} \sum_{j_1 j_2 j_3 j_4} V_{j_1 j_4 j_2 j_3} h_{j_3}^{\dagger} h_{j_4}^{\dagger} h_{j_1} h_{j_2}. \end{aligned} \tag{13.36}$$

In Equation 13.36, the first and the second terms describe the Coulomb direct and exchange interactions between the carriers in fully occupied VB. These terms lead to an irrelevant shift of the total energy and can be omitted from the Hamiltonian. The third and the fourth terms describe the direct Coulomb interaction of holes with fully occupied VB, while the fifth and the sixth terms describe the exchange interaction of holes with fully occupied VB. The third, fourth, fifth, and sixth terms are of the same mathematical form as the second term in H_s [Equation 13.33] and can be considered to originate from the effective potential that describes the interaction of holes with occupied VB. Consequently, these terms can be incorporated into H_s. Finally, the last term in Equation 13.36 describes the hole–hole interaction.

The third set of terms, $H_{\text{int}}^{(3)}$, includes the terms that contain two electron and two hole operators. Among these terms, the terms that contain $h_{j_1} e_{j_2}^{\dagger} e_{j_3} h_{j_4}^{\dagger}$ and $e_{j_1}^{\dagger} h_{j_2} h_{j_3}^{\dagger} e_{j_4}$ contribute equally to $H_{\text{int}}^{(3)}$. The same holds for the terms that contain $h_{j_1} e_{j_2}^{\dagger} h_{j_3}^{\dagger} e_{j_4}$ and $e_{j_1}^{\dagger} h_{j_2} e_{j_3} h_{j_4}^{\dagger}$. By exploiting the identities

$$h_{j_1} e_{j_2}^{\dagger} e_{j_3} h_{j_4}^{\dagger} = e_{j_2}^{\dagger} e_{j_3} h_{j_1} h_{j_4}^{\dagger} = \delta_{j_1 j_4} e_{j_2}^{\dagger} e_{j_3} + e_{j_2}^{\dagger} h_{j_4}^{\dagger} e_{j_3} h_{j_1} \tag{13.37}$$

and

$$h_{j_1} e_{j_2}^{\dagger} h_{j_3}^{\dagger} e_{j_4} = -e_{j_2}^{\dagger} e_{j_4} h_{j_1} h_{j_3}^{\dagger} = -\delta_{j_1 j_3} e_{j_2}^{\dagger} e_{j_4} - e_{j_2}^{\dagger} h_{j_3}^{\dagger} e_{j_4} h_{j_1} \tag{13.38}$$

one obtains

$$\begin{aligned} H_{\text{int}}^{(3)} = {} & \sum_{j_1 j_2 j_3 j_4} V_{j_1 j_4 j_2 j_3} e_{j_2}^{\dagger} h_{j_4}^{\dagger} e_{j_3} h_{j_1} - \sum_{j_1 j_2 j_3 j_4} V_{j_1 j_4 j_2 j_3} e_{j_2}^{\dagger} h_{j_4}^{\dagger} e_{j_3} h_{j_1} \\ & + \sum_{j_1 j_2 j_3} V_{j_1 j_1 j_2 j_3} e_{j_2}^{\dagger} e_{j_3} - \sum_{j_1 j_2 j_4} V_{j_1 j_4 j_2 j_1} e_{j_2}^{\dagger} e_{j_4}. \end{aligned} \tag{13.39}$$

The last two terms in Equation 13.39 describe direct and exchange Coulomb interaction of electrons with fully occupied VB. In the same way as similar terms for holes these can be incorporated into H_s. One thus obtains the Hamiltonian

$$\begin{aligned} H = &\sum_{j_1 j_2 \in \mathrm{CB}} \tilde{v}_{j_1 j_2} e^{\dagger}_{j_1} e_{j_2} - \sum_{j_1 j_2 \in \mathrm{VB}} \tilde{v}_{j_2 j_1} h^{\dagger}_{j_1} h_{j_2} \\ &+ \frac{1}{2} \sum_{ijkl} V_{iljk} e^{\dagger}_{\mathrm{i}} e^{\dagger}_{j} e_k e_l + \frac{1}{2} \sum_{ijkl} V_{likj} h^{\dagger}_{\mathrm{i}} h^{\dagger}_{j} h_k h_l \\ &- \sum_{ijkl} e^{\dagger}_{\mathrm{i}} h^{\dagger}_{j} h_k e_l \left(V_{ilkj} - V_{ijkl} \right). \end{aligned} \tag{13.40}$$

So far, we have not specified the SP states $\phi_c(x)$ and $\phi_v(x)$. If we choose these states to be eigenstates of the effective SP Hamiltonian $\tilde{v}_{ij}$, Equation 13.40 reduces to

$$\begin{aligned} H = &\sum_{j \in \mathrm{CB}} \varepsilon_j e^{\dagger}_j e_j - \sum_{j \in \mathrm{VB}} \varepsilon_j h^{\dagger}_j h_j \\ &+ \frac{1}{2} \sum_{ijkl} V_{iljk} e^{\dagger}_{\mathrm{i}} e^{\dagger}_{j} e_k e_l + \frac{1}{2} \sum_{ijkl} V_{likj} h^{\dagger}_{\mathrm{i}} h^{\dagger}_{j} h_k h_l \\ &- \sum_{ijkl} e^{\dagger}_{\mathrm{i}} h^{\dagger}_{j} h_k e_l \left(V_{ilkj} - V_{ijkl} \right). \end{aligned} \tag{13.41}$$

To obtain the energies of exciton states, one needs to diagonalize the Hamiltonian given in Equation 13.41. Since all terms in the Hamiltonian contain only the operators that conserve the number of electrons and the number of holes, one can separately find exciton states (one electron and one hole), biexciton states (two electrons and two holes), negative trions (two electrons and one hole), etc. For example, to find the exciton states one has to diagonalize the Hamiltonian in the Hilbert space spanned by the vectors $e^{\dagger}_a h^{\dagger}_b |G\rangle$, where $|G\rangle$ is the ground state of the system (fully occupied VB and empty CB), $a \in$ CB and $b \in$ VB. Each element of the Hamiltonian matrix depends on Coulomb integrals defined in Equation 13.32. When the **k**·**p** method is used to evaluate the SP wave functions, these are expressed as

$$\phi_j(x) = \sum_b \psi^{(b)}_j(x)\, u_b(x), \tag{13.42}$$

where b is the band index, $u_b(x)$ is the Bloch function of underlying crystal, and $\psi^{(b)}_j$ is the slowly varying envelope function. The Coulomb integral (Equation 13.32) then reads

$$\begin{aligned} V_{j_1 j_4 j_2 j_3} = \sum_{b_1 b_2 b_3 b_4} \int d^3x\, d^3y\, \psi^{(j_1)}_{b_1}(x)^{*}\, \psi^{(j_4)}_{b_4}(x)\, V_{\mathrm{int}}(x, y)\, \psi^{(j_2)}_{b_2}(y)^{*}\, \psi^{(j_3)}_{b_3}(y) \times \\ \times u^{*}_{b_1}(x)\, u_{b_4}(x)\, u^{*}_{b_2}(y)\, u_{b_3}(y). \end{aligned} \tag{13.43}$$

Next, we exploit the fact that for the slowly varying function $f(\mathbf{u})$ and rapidly varying periodic function $g(\mathbf{u})$, the following approximation holds

$$\int d\mathbf{u}\, f(\mathbf{u})\, g(\mathbf{u}) \approx \int d\mathbf{u}\, f(\mathbf{u}) \langle g(\mathbf{u}) \rangle_T, \tag{13.44}$$

where $\langle g(\mathbf{u})\rangle_T$ is the average of the function $g(\mathbf{u})$ over its period. In Equation 13.43, the terms in the first line define a slowly varying function, while the terms in the second line define a rapidly varying function whose average over the period is

$$\left\langle u^*_{b_1}(x)\, u_{b_4}(x)\, u^*_{b_2}(y)\, u_{b_3}(y)\right\rangle_T = \delta_{b_1 b_4}\delta_{b_2 b_3}, \tag{13.45}$$

due to orthonormality of the Bloch functions. Consequently, the Coulomb integral 13.43 can be expressed only in terms of envelope functions as

$$V_{j_1 j_4 j_2 j_3} = \sum_{b_1 b_2} \int d^3x\, d^3y\, \psi^{(j_1)}_{b_1}(x)^*\, \psi^{(j_4)}_{b_1}(x)\, V_{\text{int}}(x,y)\, \psi^{(j_2)}_{b_2}(y)^*\, \psi^{(j_3)}_{b_2}(y)\,. \tag{13.46}$$

These integrals can be evaluated in the PW representation using the expression given by Equation 13.16 in Section 13.4. More details on the implementation of a numerically efficient scheme for evaluation of Coulomb integrals when the PW basis is used to represent the envelope functions is given in Ref. [33].

13.7 Transport Properties of QD Nanocrystal Arrays

In this section, we discuss electrical transport properties of colloidal QD nanocrystal arrays and we derive the formula for the charge carrier hopping probability from one dot to another.

QDs obtained by colloidal chemistry typically contain ligand molecules that passivate the surface and eliminate unwanted surface trap states. However, the ligands also increase the separation between the neighboring dots and consequently decrease the wave function overlap between the electron wave functions in neighboring dots (and their electronic coupling). The presence of ligands, therefore, significantly inhibits charge transport between neighboring dots. The approaches where short ligands are used enable the improvement of electrical transport between the dots.

To engineer the transport in QD arrays, it is highly desirable to have a theoretical approach for calculating the electrical transport properties of these systems. Since electronic coupling between the dots is weak, it can be considered a perturbation that leads to occasional carrier hopping from one dot to another. To evaluate the hopping rate of charge carrier between the dots, we model the system using the following Hamiltonian

$$H = H_1 + H_2 + V. \tag{13.47}$$

In the previous equation, H_1 is the Hamiltonian of the first dot, H_2 is the Hamiltonian of the second dot, and V is the part of the Hamiltonian that describes the electronic coupling between two dots. For notational simplicity, we will assume that each of the dots accommodates only one electronic energy level that is coupled to all vibrational modes of that dot. The Hamiltonian H_1 is then given as

$$H_1 = \varepsilon_1 a_1^\dagger a_1 + \sum_s \hbar\omega_{1s} b_{1s}^\dagger b_{1s} + \sum_s g_{1s}\hbar\omega_{1s} a_1^\dagger a_1 \left(b_{1s}^\dagger + b_{1s}\right). \tag{13.48}$$

In the previous equation, a_1 is the electron annihilation operator that annihilates an electron in state of energy ε_1, g_{1s} is the dimensionless coupling constant describing the coupling of electron in dot 1 to the phonon mode s of dot 1, and $\hbar\omega_{1s}$ is the energy of that phonon mode, while b_{1s} is the corresponding phonon annihilation operator. Electronic operators satisfy fermionic commutation relations, while phonon operators satisfy bosonic commutation relations. Analogously, the Hamiltonian H_2 reads

$$H_2 = \varepsilon_2 a_2^\dagger a_2 + \sum_s \hbar\omega_{2s} b_{2s}^\dagger b_{2s} + \sum_s g_{2s}\hbar\omega_{2s} a_2^\dagger a_2 \left(b_{2s}^\dagger + b_{2s}\right), \tag{13.49}$$

where the notation is analogous to that in the Hamiltonian H_1. The last term in Equation 13.47 reads

$$V = -t\left(a_1^\dagger a_2 + a_2^\dagger a_1\right), \tag{13.50}$$

where t is the electronic coupling parameter between the two electronic states in dots 1 and 2.

We are interested in determining the transition rate for the process in which an electron that is initially in dot 1 makes a transition to dot 2. Possible initial states for this process are $|i\rangle = a_1^\dagger \left|i_{\mathrm{ph}}\right\rangle$, where $\left|i_{\mathrm{ph}}\right\rangle$ denotes a state of the phonon subsystem. The final state is given as $|f\rangle = a_2^\dagger \left|f_{\mathrm{ph}}\right\rangle$, where $\left|f_{\mathrm{ph}}\right\rangle$ is the final state of the phonon subsystem. The transition rate is then given by Fermi's Golden Rule rule expression

$$W = \frac{2\pi}{\hbar}\sum_i p_i \sum_f \left|V_{if}\right|^2 \delta\left(E_{\mathrm{i}} - E_{\mathrm{f}}\right), \tag{13.51}$$

where p_{i} is the probability that the initial state is $|i\rangle$ $V_{if} = \langle i|V|f\rangle$ is the matrix element of the operator V, while E_{i} and E_{f} are energies of the unperturbed system in the initial and final states. We will transform this conventional expression to a form that is more convenient for our analysis. Using the identities

$$\delta\left(E_{\mathrm{i}} - E_{\mathrm{f}}\right) = \frac{1}{2\pi\hbar}\int_{-\infty}^{\infty} dt \exp\left[\frac{i}{\hbar}\left(E_{\mathrm{i}} - E_{\mathrm{f}}\right)t\right], \tag{13.52}$$

$$\left|V_{if}\right|^2 = \langle i|V|f\rangle\langle f|V|i\rangle, \tag{13.53}$$

and

$$\langle i|V|f\rangle \exp\left[\frac{i}{\hbar}\left(E_{\mathrm{i}} - E_{\mathrm{f}}\right)t\right] = \left\langle i\left|\exp\left[\frac{i}{\hbar}H_0 t\right] V \exp\left[-\frac{i}{\hbar}H_0 t\right]\right|f\right\rangle \tag{13.54}$$

where $H_0 = H_1 + H_2$, we arrive at the expression for transition rates that is convenient for our analysis

$$W = \frac{1}{\hbar^2}\int_{-\infty}^{\infty} dt \sum_i p_i \sum_f \left\langle i\left|V_I(t)\right|f\right\rangle \left\langle f\left|V_I(0)\right|i\right\rangle, \tag{13.55}$$

where $V_I(t) = \exp\left[\frac{i}{\hbar}H_0 t\right] V \exp\left[-\frac{i}{\hbar}H_0 t\right]$.

The Hamiltonians H_1 and H_2 can be diagonalized exactly using the unitary transformation

$$U = \exp\left[a_1^\dagger a_1 \sum_s g_{1s}\left(b_{1s}^\dagger - b_{1s}\right)\right] \exp\left[a_2^\dagger a_2 \sum_s g_{2s}\left(b_{2s}^\dagger - b_{2s}\right)\right]. \tag{13.56}$$

This transformation acts on electron and phonon operators as

$$Ua_\mathrm{i}U^\dagger = a_\mathrm{i}\exp\left[\sum_\mathrm{s} g_{is}\left(b_{is}^\dagger - b_{is}\right)\right] \tag{13.57}$$

and

$$Ub_{is}U^\dagger = b_{is} - g_{is}a_\mathrm{i}^\dagger a_\mathrm{i}. \tag{13.58}$$

The transformed Hamiltonian then reads

$$\tilde{H} = UHU^\dagger = \tilde{H}_1 + \tilde{H}_2 + \tilde{V}, \tag{13.59}$$

with

$$\tilde{H}_1 = \left(\varepsilon_1 - \sum_\mathrm{s} g_{1s}^2\hbar\omega_\mathrm{s}\right)a_1^\dagger a_1 + \sum_\mathrm{s}\hbar\omega_{1s}b_{1s}^\dagger b_{1s}, \tag{13.60}$$

$$\tilde{H}_2 = \left(\varepsilon_2 - \sum_\mathrm{s} g_{2s}^2\hbar\omega_\mathrm{s}\right)a_2^\dagger a_2 + \sum_\mathrm{s}\hbar\omega_{2s}b_{2s}^\dagger b_{2s}, \tag{13.61}$$

and

$$\tilde{V} = -t\left[a_1^\dagger a_2\exp\left[-\sum_\mathrm{s} g_{1s}\left(b_{1s} - b_{1s}^\dagger\right)\right]\exp\left[\sum_\mathrm{s} g_{2s}\left(b_{2s} - b_{2s}^\dagger\right)\right] + c.c.\right]. \tag{13.62}$$

To obtain the expression for transition probability given by Equation 13.55, we first express the matrix elements as

$$\left\langle i\left|V_I(t)\right|f\right\rangle = \left\langle\Psi_\mathrm{i}\right|\exp\left(\frac{i}{\hbar}\tilde{H}_0 t\right)\tilde{V}\exp\left(-\frac{i}{\hbar}\tilde{H}_0 t\right)\left|\Psi_\mathrm{f}\right\rangle. \tag{13.63}$$

In the last expression $\left|\Psi_\mathrm{i}\right\rangle = U\left|i\right\rangle$ and $\left|\Psi_\mathrm{f}\right\rangle = U\left|f\right\rangle$. By noting that $\tilde{H}_0 = \tilde{H}_0^\mathrm{e} + \tilde{H}_0^\mathrm{ph}$ and $\exp\left[-\frac{i}{\hbar}\tilde{H}_0^\mathrm{e}t\right]\left|\Psi_\mathrm{f}\right\rangle = \exp\left[-\frac{i}{\hbar}\varepsilon_\mathrm{f}t\right]\left|\Psi_\mathrm{f}\right\rangle$, using Equations 13.62 and 13.55 one arrives at

$$\begin{aligned}W = {} & \frac{t^2}{\hbar^2}\int_{-\infty}^{\infty} du\exp\left[\frac{i}{\hbar}\left(\varepsilon_\mathrm{i} - \varepsilon_\mathrm{f}\right)u\right]\sum_{i_\mathrm{ph}} p_{i_\mathrm{ph}} \\ & \times\sum_{f_\mathrm{ph}}\left\langle i_\mathrm{ph}\right|\exp\left(\frac{i}{\hbar}\tilde{H}_0^\mathrm{ph}u\right)\exp\left[-\sum_\mathrm{s} g_{1s}\left(b_{1s} - b_{1s}^\dagger\right)\right]\exp\left[\sum_\mathrm{s} g_{2s}\left(b_{2s} - b_{2s}^\dagger\right)\right]\exp\left(-\frac{i}{\hbar}\tilde{H}_0^\mathrm{ph}u\right)\left|f_\mathrm{ph}\right\rangle \\ & \times\left\langle f_\mathrm{ph}\right|\exp\left[\sum_\mathrm{s} g_{1s}\left(b_{1s} - b_{1s}^\dagger\right)\right]\exp\left[-\sum_\mathrm{s} g_{2s}\left(b_{2s} - b_{2s}^\dagger\right)\right]\left|i_\mathrm{ph}\right\rangle.\end{aligned} \tag{13.64}$$

Taking into account that $\sum_{f_{\mathrm{ph}}} |f_{\mathrm{ph}}\rangle\langle f_{\mathrm{ph}}| = 1$ and $\sum_{i_{\mathrm{ph}}} p_{i_{\mathrm{ph}}} \langle i_{\mathrm{ph}}|X|i_{\mathrm{ph}}\rangle = \langle X\rangle_{\mathrm{ph}}$, where $\langle X\rangle_{\mathrm{ph}}$ denotes the expectation value of the operator X, we obtain

$$\begin{aligned} W =& \frac{t^2}{\hbar^2}\int_{-\infty}^{\infty} du \exp\left[\frac{i}{\hbar}\left(\varepsilon_{\mathrm{i}}-\varepsilon_{\mathrm{f}}\right)u\right] \\ &\times\left\langle \exp\left\{-\sum_s g_{1s}\left[b_{1s}(u)-b_{1s}^{\dagger}(u)\right]\right\}\exp\left\{\sum_s g_{1s}\left[b_{1s}-b_{1s}^{\dagger}\right]\right\}\right\rangle \\ &\times\left\langle \exp\left\{\sum_s g_{2s}\left[b_{2s}(u)-b_{2s}^{\dagger}(u)\right]\right\}\exp\left\{-\sum_s g_{2s}\left[b_{2s}-b_{2s}^{\dagger}\right]\right\}\right\rangle . \end{aligned} \tag{13.65}$$

Next, we exploit the identity

$$\begin{aligned} &\left\langle \exp\left\{-g_{is}\left[b_{is}(u)-b_{is}^{\dagger}(u)\right]\right\}\exp\left\{g_{is}\left[b_{is}-b_{is}^{\dagger}\right]\right\}\right\rangle = \\ &\exp\left\{-g_{is}^2\left[2n_{is}+1-(n_{is}+1)e^{-i\omega_{is}u}-n_{is}e^{i\omega_{is}u}\right]\right\}, \end{aligned} \tag{13.66}$$

where g_{is}, n_{is}, and $\hbar\omega_{is}$ are, respectively, the electron–phonon coupling constant, the phonon occupation number and the phonon energy for the sth phonon mode at site i. We finally obtain the transition rate as [69–71]

$$\begin{aligned} W =& \frac{t^2}{\hbar^2}\int_{-\infty}^{\infty} du \exp\left[\frac{i}{\hbar}\left(\varepsilon_{\mathrm{i}}-\varepsilon_{\mathrm{f}}\right)u\right]\times \\ &\prod_s \exp\left\{-g_{1s}^2\left[2n_{1s}+1-(n_{1s}+1)e^{-i\omega_{1s}u}-n_{1s}e^{i\omega_{1s}u}\right]\right\}\times \\ &\prod_s \exp\left\{-g_{2s}^2\left[2n_{2s}+1-(n_{2s}+1)e^{-i\omega_{2s}u}-n_{2s}e^{i\omega_{2s}u}\right]\right\}. \end{aligned} \tag{13.67}$$

Equation 13.67 can be used directly to calculate the transition rate from one dot to another. In some cases, a simpler Marcus formula is often used instead of Equation 13.67. In what follows, we introduce additional approximations that lead to the reduction of Equation 13.67 to the Marcus formula. However, we alert the reader that the conditions for the validity of the Marcus formula should be carefully checked before that formula is used.

A detailed analysis of the exponential terms in Equation 13.67 leads to the conclusion that when electron–phonon interaction is strong ($\sum_s g_{is}^2 \gg 1$) exponential terms decay very quickly when $\left|\omega_{is}u\right| \gg 1$. Therefore, only the terms with $\left|\omega_{is}u\right| \ll 1$ contribute significantly to the integral. By keeping terms up to the quadratic in the Taylor expansion $e^{i\omega_{is}t} = 1 + i\omega_{is}t - \frac{1}{2}\omega_{is}^2 t^2$, one obtains

$$\begin{aligned} &\prod_s \exp\left\{-g_{1s}^2\left[2n_{1s}+1-(n_{1s}+1)e^{-i\omega_{1s}u}-n_{1s}e^{i\omega_{1s}u}\right]\right\} \approx \\ &\prod_s \exp\left(i\omega_{1s}g_{1s}^2 u\right)\exp\left[-\frac{1}{2}g_{1s}^2\omega_{1s}^2u^2\left(2n_{1s}+1\right)\right]. \end{aligned} \tag{13.68}$$

When the temperature is high, $k_BT \gg \hbar\omega_{is}$, the $\left(2n_{is}+1\right)$ term reduces to $\frac{2k_BT}{\hbar\omega_{is}}$. Using that approximation one obtains

$$W \approx \frac{t^2}{\hbar^2}\int_{-\infty}^{\infty} du \exp\left[\frac{i}{\hbar}\left(\varepsilon_{\mathrm{f}}-\varepsilon_{\mathrm{i}}+\lambda_1+\lambda_2\right)u\right]\exp\left[-\frac{k_BT}{\hbar^2}\left(\lambda_1+\lambda_2\right)u^2\right], \tag{13.69}$$

where $\lambda_i = \sum_s \hbar\omega_{is}\, g_{is}^2$. The solution of the last integral gives the Marcus formula

$$W \approx \frac{t^2}{\hbar}\sqrt{\frac{\pi}{k_B T\lambda}}\exp\left[-\frac{\left(\lambda + \varepsilon_f - \varepsilon_i\right)^2}{4k_B T\lambda}\right], \tag{13.70}$$

where $\lambda = \lambda_1 + \lambda_2$.

13.8 Practical Examples: Cd-Based Chalcogenide QDs

13.8.1 Comparison with Experiment: Absorption Edge Wavelength of CdSe/CdTe Type II CQDs

In this section, we apply the theoretical approach described in previous sections to calculate the absorption wavelength in CdSe/CdTe core/shell QDs. We analyze the influence of CdSe/CdTe core/shell colloidal QD (CQD) morphology on the variation of the first exciton peak in the absorption spectra, which corresponds to excitation of the $1S^e_{1/2}1S^h_{3/2}$ state. As a first step, we validate our theoretical methodology, Sections 13.3–13.6, based on the combination of **k·p** and CI Hamiltonians, against available experimental and theoretical results on CdSe and CdTe core-only CQD, as they are the constituent materials of our core/shell structure. Figure 13.3 shows the variation of the $1S^e_{1/2}1S^h_{3/2}$ exciton energy as a function of CdSe and CdTe QD radius, a_c, respectively. Solid symbols represent experimental results from several independent measurements from different groups and the line is an empirical inverse polynomial fitting curve, $E_X^{1S_{1/2}1S_{3/2}}(a_c) = E_g^{\text{bulk}} + (Aa_c^2 + Ba_c + C)^{-1}$, to this experimental data proposed by de Mello Donega in Ref. [72], where A, B, and C are fitting parameters. Results of $1S^e_{1/2}1S^h_{3/2}$ exciton energies predicted by our calculation are shown on the same figures and exhibit excellent agreement with the experimental measurements [73–82].

Having validated the methodology for core-only CQD structures, we now employ our method to describe CdSe/CdTe core/shell type-II CQD structures. The remaining question when modeling the system of core/shell structures is the parametrization of the valence band offset (VBO) between core and shell materials, in our case CdSe and CdTe, respectively. From our previous analysis [58] of the variation of $1S^e_{1/2}1S^h_{3/2}$ and $1S^e_{1/2}2S^h_{3/2}$ excitonic energies with the shell thickness in CdTe/CdSe core/shell CQD, we have estimated the $\text{VBO}_{\text{CdTe/CdSe}} = 0.4$ eV. We have assumed the same VBO for inverse structures, i.e., for CdSe/CdTe core/shell CQD considered here.

The $1S^{(e)}_{1/2}nS^{(h)}_{3/2}$ $(n = 1, 2)$ states are the two lowest-energy excitons observed in the absorption spectra of colloidal CdTe/CdSe NCs [23,83,84], making them the most important for understanding the near band-edge absorption characteristics of such nanoparticles. Figure 13.3 shows the $1S^{(e)}_{1/2}1S^{(h)}_{3/2}$ and $1S^{(e)}_{1/2}2S^{(h)}_{3/2}$ exciton energies (solid lines) calculated by the CI as a function of shell thickness for CdTe/CdSe QDs with (a) $a_c = 1.7$ nm, (b) $a_c = 1.72$ nm, (c) $a_c = 1.75$ nm, and (d) $a_c = 1.95$ nm. Dashed lines show upper and lower limits on the exciton energies resulting from an uncertainty of 1 monolayer (ML) in the displayed core radii ($\sim \pm 0.3$ nm). Filled circles show exciton energies taken from the first and second absorption peak positions in absorption spectra measured by (a) Gong et al., [84], (b) Ma et al., [85], (c) Cai et al. [86], and (d) Oron et al. [23]. We see good quantitative agreement between the calculated exciton energies and the experimental data, with the data lying in the channels defined by an uncertainty of ±1 ML width in the core size. It should be noted that the results of Oron et al. [23] were obtained on zinc-blende NC structures, in addition to those of Cai et al. [86] The papers by Gong et al. [84] and Ma et al. [85] do not explicitly state the crystal structures of the core/shell nanoparticles, although Ma et al. [85] note that their core/shell NCs gave very similar absorption and photoluminescence spectra to those of Cai and coworkers [86]. Our calculations accurately reproduce the 0.25 eV energy separation between the $1S^{(e)}_{1/2}1S^{(h)}_{3/2}$ and $1S^{(e)}_{1/2}2S^{(h)}_{3/2}$ excitons that is nearly independent of shell thickness [84]. This constant energy separation is

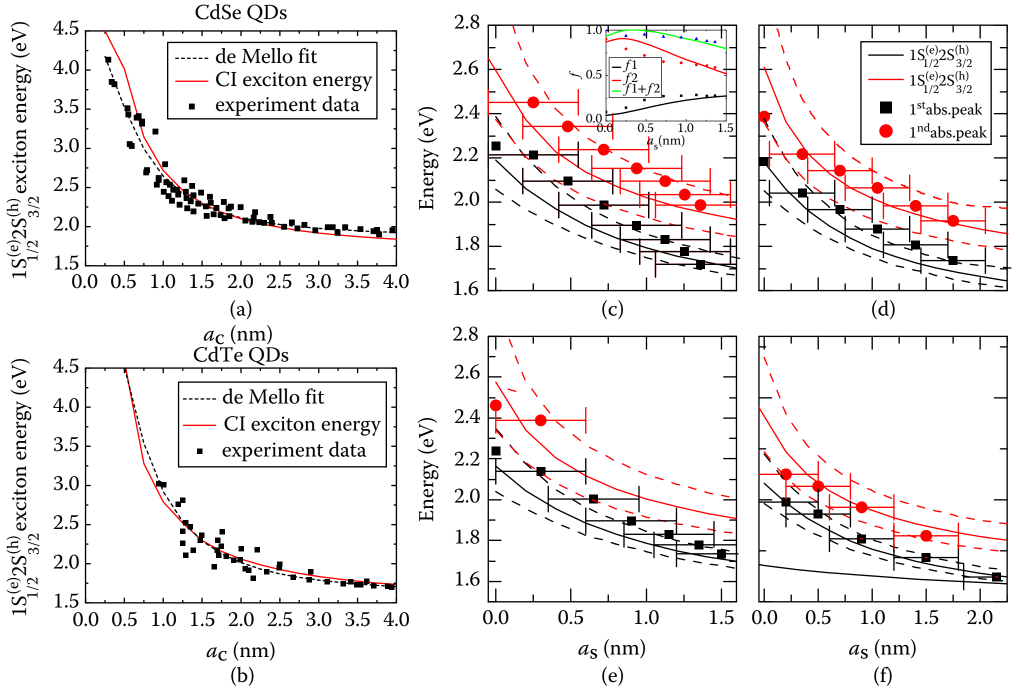


FIGURE 13.3 Variation of the $1S^e_{1/2}1S^h_{3/2}$ exciton energy as a function of colloquial quantum dot (CQD) core radius, a_c, in (a) CdSe and (b) CdTe CQDs: experimental results (square symbols), de Mello Donega's empirical fit to the experimental data (dotted line) [72], and theoretical model (solid line). Calculated (lines) energies of the $1S^{(e)}_{1/2}nS^{(h)}_{3/2}$ ($n = 1, 2$) excitons for CdTe/CdSe QDs with shell thickness a_s and core radius of (c) $a_c = 1.7$ nm, (d) $a_c = 1.72$ nm, (e) $a_c = 1.75$ nm, and (f) $a_c = 1.95$ nm; experimental data taken from Refs. [84], [85],[86], and [23] are shown as filled symbols. Error bars represent an uncertainty of 1 mono-layer (ML) ($\approx \pm 0.3$ nm) in the shell thickness. Dashed lines show upper and lower limits on the exciton energies resulting from an uncertainty of 1 ML in the nominal core radii.

characteristic of changing electron confinement but approximately constant hole confinement in the h/e heterostructure. We also find good agreement between the oscillator strength obtained by Gong et al. [84] from the absorption spectra and our calculations. Calculating the oscillator strength f_n of the $1S^{(e)}_{1/2}nS^{(h)}_{3/2}$ ($n = 1, 2$) excitons as $f_n = 2P^2_X/m_0E_X$ we find that $f_1 + f_2 \sim$ constant (inset Figure 13.3), confirming the validity of relevant excitonic wave functions too.

In contrast to core-only CQDs, the size dependence of $1S^e_{1/2}1S^h_{3/2}$ exciton energies in core/shell CQDs is more complex and it is not possible to capture its trend by a simple polynomial fitting curve. Figure 13.5 shows the shell thickness, a_s, dependence of the $1S^e_{1/2}1S^h_{3/2}$ absorption wavelengths for CdSe/CdTe core/shell CQDs for different core radii, a_c, ranging from a_c=1.5 nm to 2.5 nm. Also shown in Figure 13.5 are experimental data for a CdSe core of 1.7-nm radius, as measured by transmission electron microscopy, and different CdTe shell thicknesses [87,88]. Good agreement is seen between the calculated energy of the $1S^e_{1/2}1S^h_{3/2}$ exciton and the spectral position observed for the first absorption peak, further validating our theoretical approach.

To illustrate our theoretical model outlined in Section 13.6, in Figure 13.4, we show the excitonic spectra of an $a_c = 2$ nm CdSe core-only QD ($a_s = 0$ case) and a set of CdSe/CdTe core/shell QDs with shell thicknesses varying from $a_s = 0.5$ to 3.0 nm. By changing the shell thickness from $a_s = 0$ (no shell) to $a_s = 3$ nm the excitonic ground state energy, E_{X0}, is changed from 2.30 eV to 1.49 eV. It is also interesting to note that for certain shell thicknesses, i.e., between $a_s = 0.5$ and 1 nm, the character of the ground state exciton changes from $1S_{1/2}1S_{3/2}$ to $1S_{1/2}1P_{3/2}$. It can be observed that the density of excitonic states

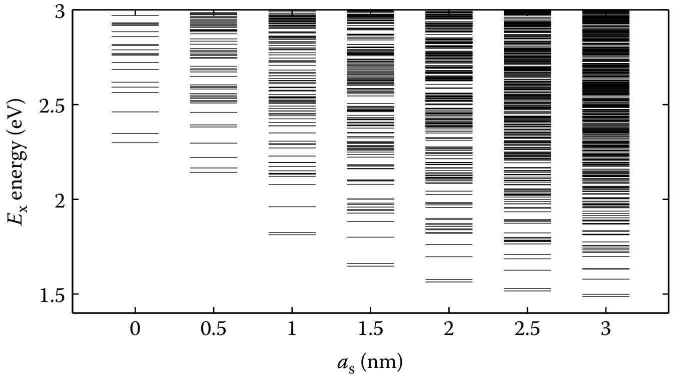


FIGURE 13.4 Excitonic spectra of CdSe and CdSe/CdTe quantum dots (QDs) as a function of the shell thickness, a_s.

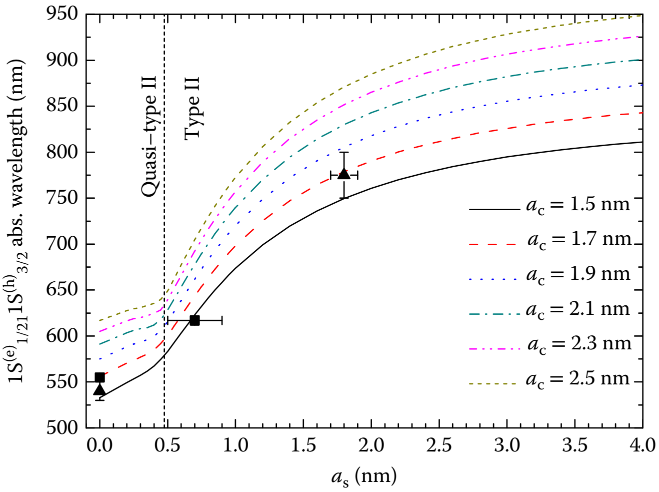


FIGURE 13.5 Core radius, a_c, and shell thickness, a_s, dependence of the $1S^e_{1/2}1S^h_{3/2}$ exciton peak in the absorption spectra of CdSe/CdTe core/shell quantum dots (QDs). Also shown for comparison are experimental data from Ref. [87] (squares) and Ref. [88] (triangles).

increases with shell thickness: this suggests that phononic gaps are more likely to be found in core-only or core/shell structures with thin shells [16].

In Figure 13.5, we can clearly distinguish between two trends in the $1S^e_{1/2}1S^h_{3/2}$ absorption wavelengths in the region of shell thicknesses <0.5 nm and for shell thicknesses >0.5 nm. For the QDs with shells of thickness approximately or less than 0.5 nm, we can see that (1) the $1S^e_{1/2}1S^h_{3/2}$ absorption wavelengths are approximately linearly dependent on a_s, suggesting strong influence of the dielectric confinement on $1S^e_{1/2}1S^h_{3/2}$ excitons and their wavelengths in this region and (2) as the core size, a_c, increases the gradient of the $1S^e_{1/2}1S^h_{3/2}$ absorption wavelength versus a_s curves decreases. We explain such nonmonotonic behavior in terms of the changing localization regime of the $1s^h_{3/2}$ SP hole, in the correlated $1S^e_{1/2}1S^h_{3/2}$ exciton, as the QD dimensions change [58]. The lower gradient of the curves for shell widths <0.5 nm is due to the fact that the $1s^h_{3/2}$ hole is in the delocalized regime, i.e., its probability density is spread over the whole heterostructure and its energy is mainly determined by the global confinement provided by the QD potential well of radius $a_c + a_s$. In this regime, the size dependence of the hole confinement is closer to that of one confined in a core-only QD. For the QDs with shells of width greater than 0.5 nm, we can see that the $1S^e_{1/2}1S^h_{3/2}$ absorption wavelengths again acquire a trend that can be described by a quadratic polynomial function. As a_s increases, for a particular core size, the $1s^h_{3/2}$ SP hole localizes in the shell fully so the absorption wavelength is again more strongly affected by the effect of shell thickness on the confinement energy of $1s^h_{3/2}$. This behavior is closer to the SCR in a core/shell heterostructure. Maps of the absorption wavelength against QD core/shell dimensions as depicted in Figure 13.5 can only be obtained numerically and should be of use to experimentalists studying such systems.

13.8.2 Radiative Lifetimes in CdSe/CdTe CQDs

Next, we analyze the radiative lifetimes in core/shell CdSe/CdTe CQDs [17]. In order to assess the variation of the radiative times, $\tau_{\text{rad.}}$, with the core size, a_c, and the shell thickness, a_s, we use the following expression:

$$\frac{1}{\tau_{\text{rad.}}} = \frac{1}{3}\frac{F^2\bar{n}e^2}{\pi\epsilon_0 c^3\hbar^4}E_X\frac{1}{d_X}\sum_{i=1}^{d_X}|\mathbf{P}_{X_i}|^2, \tag{13.71}$$

where $\bar{n}$ is the refractive index of the colloidal material, e is the electron charge, ϵ_0 is the permittivity of free space, m_0 is the electron rest mass, c is the speed of light, and $\hbar$ is the reduced Planck constant. In the expression above, $|\mathbf{P}_X|$ is the modulus of excitonic dipole matrix element, given in (eVÅ), and E_X is the excitonic energy, both obtained from the CI calculation, and d_X is the integer number representing the degree of degeneracy of a particular excitonic state. For the $1S^e_{1/2}1S^h_{3/2}$ exciton considered below the degeneracy is $d_X = 8$. The expression for dielectric screening of spherical core/shell QDs is given as $F = 9\epsilon_s\epsilon_{\text{coll.}}/(\epsilon_s\epsilon_a + 2\epsilon_{\text{coll.}}\epsilon_b)$, where $\epsilon_a = \epsilon_c(3 - 2\Omega_s/\Omega_{qd}) + 2\epsilon_s\Omega_s/\Omega_{qd}$, $\epsilon_b = \epsilon_c\Omega_s/\Omega_{qd} + \epsilon_s(3 - \Omega_s/\Omega_{qd})$, and Ω_{qd} and Ω_s are the volume of QD and shell, respectively [89].

Figure 13.6 shows the mean radiative lifetime of the $1S^e_{1/2}1S^h_{3/2}$ exciton in CdSe/CdTe CQDs as a function of shell thickness a_s for several different core radii. The most noticeable feature is the sudden increase in $\tau_{\text{rad.}}$ at around $a_s \sim 0.5$ nm. In comparison to core-only CQDs, in which the value of $1S^e_{1/2}1S^h_{3/2}$ excitonic radiative time τ_{rad} changes relatively weakly with the CQD size [35,90], in core/shell structures it is

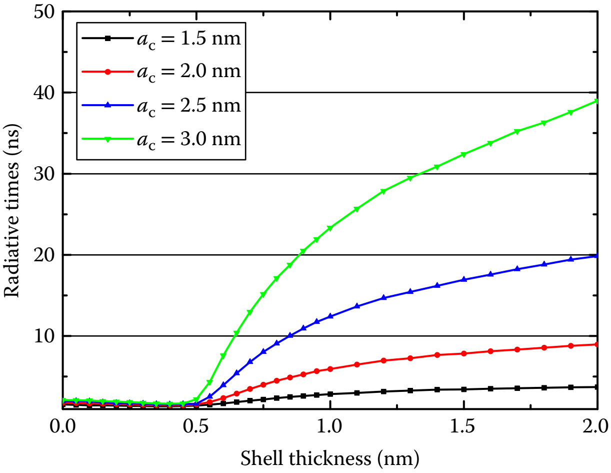


FIGURE 13.6 Plot of the $1S^e_{1/2}1S^h_{3/2}$ exciton radiative lifetime as a function of the shell thickness for different cores: $a_c = 1.5, 2.0, 2.5$, and 3.0 nm in CdSe/CdTe colloquial quantum dots (CQDs).

possible to increase this radiative time over one order of magnitude with suitable change in the core and shell sizes. Several observations regarding the trend in τ_{rad} in CdSe/CdTe core/shell type II CQDs can be made here: (1) again as in the case of the absorption wavelength, shown in Figure 13.5, we can distinguish clearly between two different regions of the shell thicknesses $a_{\text{s}} < 0.5$ nm and $a_{\text{s}} > 0.5$ nm. In the region of $a_{\text{s}} < 0.5$ nm, the whole CQD system behaves similar to core-only CQDs, since both electron and hole correlated charge densities are either largely confined in the core region or just starting to delocalize over the whole CQD structure. In this region, the optical dipole matrix element, $|\mathbf{P}_{\text{X}}|$, is strong while the magnitude of both $|\mathbf{P}_{\text{X}}|$ and E_{X} changes very little with the overall sizes of QDs, which explains the almost constant τ_{rad}. (2) As shell thickness increases, and in particular beyond $a_{\text{s}} > 1$ nm, the hole states become strongly confined in the shell region. Here, dielectric confinement does not have enough strength to overcome the confinement imposed by the type II aligned VB edge of CdTe, so that the exciton reaches the SCR. Consequently, the electron–hole wave function overlap is dramatically reduced and E_{X} is mainly now determined by the variation of the $1s_{3/2}^{(h)}$ hole confinement with a_{s}. For $a_{\text{s}} > 1$ nm, $\tau_{\text{rad.}}$ continues to increase, but more slowly, reaching values that are about one order of magnitude greater than those for core-only CQDs. For all shell thicknesses, a monotonic trend to larger $\tau_{\text{rad.}}$ with increasing core size is also observed. It is interesting to observe that for CdSe/CdTe type II CQDs, the value of $\tau_{\text{rad.}}$ increases over "only" one order of magnitude, while in other systems, such as epitaxially grown type II InAs/GaAs/GaAsSb structure, the radiative times can increase over three orders of magnitude [91,92]. This difference in behavior is attributed to the absence of significant dielectric confinement in epitaxially grown QDs, compared to CQDs.

13.8.3 Correlation Energy

The correlation energy of the exciton can be defined as

$$E_{\text{corr}} = E_{\text{X,CI}} - E_{\text{X}}, \tag{13.72}$$

where E_{X} is the exciton energy calculated according to first-order perturbation theory inside the strong confinement approximation for the exciton wave function [42,61,83,93], and $E_{\text{X,CI}}$ is the excitonic energy calculated using the CI method.

13.8.3.1 CdTe/CdSe QD: Effect of Electron Shell Localization

In Figure 13.7a, we present E_{corr} for the $1S_{1/2}^{(e)}1S_{3/2}^{(h)}$ exciton as a function of core radius for fixed shell $a_{\text{s}} = 2$ nm CdTe/CdSe QDs. We see that in the presence of dielectric confinement $|E_{\text{corr}}| \lesssim 20$ meV and that E_{corr} exhibits at least one minimum as a function of a_{c} in the $\epsilon = \text{const.}$ and $\epsilon = \epsilon(r)$ case. $|E_{\text{corr}}|$ is up to four times greater in the presence of dielectric confinement, $\epsilon = \epsilon(r)$, compared to the $\epsilon = \text{const.}$ case. This result highlights the importance of a proper treatment of the dielectric environment in such nanostructures. The minimum in E_{corr} for the $1S_{1/2}^{(e)}1S_{3/2}^{(h)}$ exciton, Figure 13.7a, is a consequence of two competing effects: proximity of the self-polarization potential peak, which tends to reduce the electron–hole separation and the effect of the type-II confinement profile, which tends to separate the carriers as a_{c} increases.

13.8.3.2 CdSe/CdTe QD: Effect of Hole Shell Localization

In Figure 13.7b, we present E_{corr} for the $1S_{1/2}^{(e)}1S_{3/2}^{(h)}$ exciton as a function of shell thickness for an $a_{\text{c}} = 3.5$ nm fixed core in CdSe/CdTe QDs. We observe the largest size correlation energy of the considered excitons for the $1S_{1/2}^{(e)}1S_{3/2}^{(h)}$ state in CdSe/CdTe QDs when $\epsilon = \epsilon(r)$, with E_{corr} reaching −62 meV for an $a_{\text{c}} = 3.5$ nm, $a_{\text{s}} = 0.6$ nm QD, Figure 13.7b. This value is more than six times larger than the corresponding value for the $\epsilon = \text{const.}$ case, highlighting the particularly strong effect of the dielectric environment on this exciton. It can be observed from Figure 13.7b that the effect of dielectric mismatch on E_{corr} is strongest

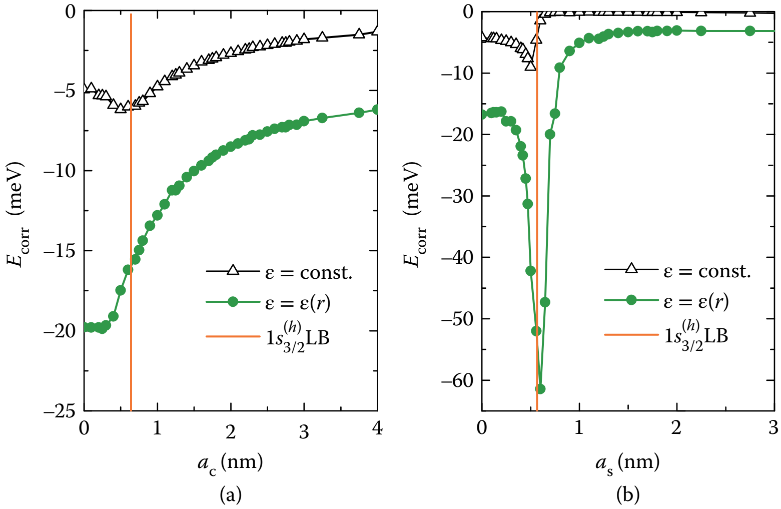


FIGURE 13.7 Correlation energy, E_{corr}, of $1S^{(e)}_{1/2}1S^{(h)}_{3/2}$ excitons in: (a) $a_s = 2.0$ nm CdTe/CdSe quantum dots (QDs), and (b) $a_c = 3.5$ nm CdSe/CdTe QDs. Cases of $\epsilon = \mathrm{const.}$ and $\epsilon = \epsilon(r)$ are represented by open and solid symbols, respectively. Vertical line points to the localization/delocalization boundary (LB) of $1s_{3/2}$ hole state.

in the vicinity of the $1s^{(h)}_{3/2}$ localization boundary (LB), i.e., once the SP hole starts to move significantly toward the shell region and becomes delocalized over the whole QD.

$1S^{(e)}_{1/2}1S^{(h)}_{3/2}$ *exciton:* In the case of a core-only CdSe QD correlation causes both carriers, correlated electron and correlated hole, to move closest toward the center of the QD compared with their strongest SP character. This is purely a result of the direct interparticle Coulomb interaction. Introduction of the self-polarization potential, i.e., $\epsilon = \epsilon(r)$, further exaggerates this move of the radial probability densities (RPD) of both carriers away from the QD surface in the core-only CdSe QD, Figure 13.8a. This effect increases localization of both carriers near the center of QD, increasing overlap between them and giving correlation energies $E_{\mathrm{corr}} = -18$ meV. For the $1S^{(e)}_{1/2}1S^{(h)}_{3/2}$ exciton in the CdSe QD, the shift in RPD is mainly due to an increase in $2s^{(e)}_{1/2}1s^{(h)}_{3/2}$ character.

To assess the effect of dielectric confinement on the correlated carriers in the CdSe QD, we consider the expectation value of the 1*s* electron (hole) radial coordinate, denoted $\langle r_{e(h)} \rangle$. When $\epsilon = \mathrm{const.}$ (no self-polarization) we find $\langle r_h \rangle = 1.55$ nm compared to $\langle r_h \rangle = 1.44$ nm when $\epsilon = \epsilon(r)$ for the $1s^{(h)}_{3/2}$ state. In contrast, the effect of dielectric confinement moves the $1s^{(e)}_{1/2}$ electron from $\langle r_e \rangle = 2.01$ nm to $\langle r_e \rangle = 1.89$. Although the SP $1s^{(e)}_{1/2}$ RPD has significantly greater overlap with the repulsive peak in self-polarization potential near the QD surface than the SP $1s^{(h)}_{3/2}$ hole RPD, the correlated electron is shifted by dielectric confinement by almost the same distance as the correlated hole. These results reflect the larger sensitivity of the correlated hole wave function to the dielectric environment compared to the electron in the CdSe core-only QD.

In Figure 13.8b, we see that the introduction of a thin CdTe shell allows the uncorrelated hole to start to localize nearer the QD surface (at $r = a_c + a_s$), dramatically reducing its overlap with the uncorrelated electron. However, the effect of correlation is strong enough to pull the hole back toward the center, mainly due to the addition of the $1s^{(e)}_{1/2}2s^{(h)}_{3/2}$ electron–hole pair (EHP) character to the exciton wave function. We see that the introduction of dielectric confinement exaggerates this move of the carriers further away from the QD surface compared to the $\epsilon = \mathrm{const.}$ case [58]. The effect of dielectric confinement is particularly strong in this case because $\langle r_h \rangle$ for the $1s^{(h)}_{3/2}$ state is close to the value of QD's outermost radius, $a_c + a_s$.

The close proximity of the hole to the QD surface reduces the distance $\xi = \langle r_h^{\text{QD}} \rangle - \langle r_h^{\text{ind.}} \rangle$ between the hole in the QD and its induced charge in the colloid, increasing the Coulombic repulsion between them which scales as $\propto 1/\xi$. Such repulsion, in addition to the strong effect of correlations as discussed above, causes the hole to be pushed back toward the center of the QD, thereby dramatically increasing overlap with the correlated electron wave function. The presence of dielectric confinement means the exciton wave function is an almost equal superposition of the $1s_{1/2}^{(e)} ns_{3/2}^{(h)}$ $(n = 1, 2)$ states, with characters $|c_1|^2 = 0.449$ and $|c_2|^2 = 0.458$. For comparison, when $\epsilon = \text{const.}$ the $1s_{1/2}^{(e)} 2s_{3/2}^{(h)}$ character amounts to only $|c_2|^2 = 0.019$. The much stronger configuration mixing in the dielectric confinement case allows E_{corr} to reach ~ -62 meV, compared to -9 meV without dielectric confinement.

Further increase of the CdSe/CdTe QD shell thickness to $a_s = 1$ nm, Figure 13.8c, allows the SP hole to fully localize in the shell, while the SP electron stays in the core, reaching the type-II localization limit. The carriers effectively enter the strong confinement regime (SCR) in which the Coulomb effects are overridden

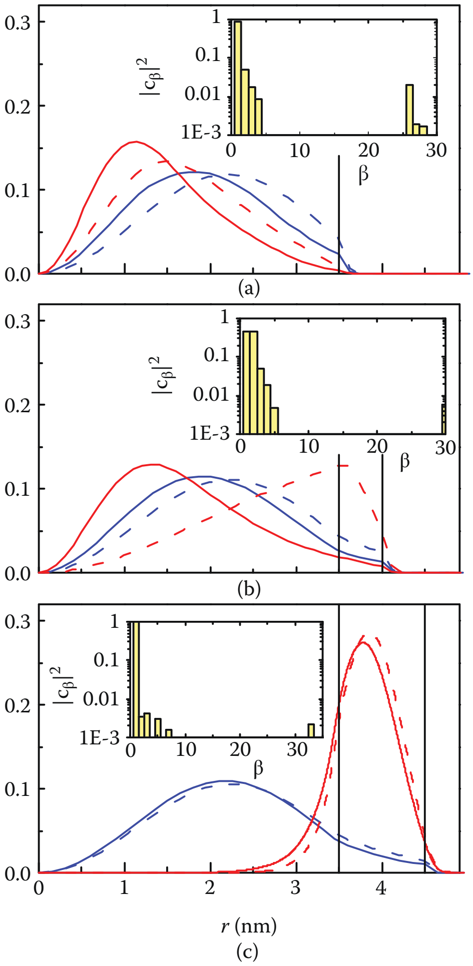


FIGURE 13.8 Solid (dashed) lines represent radial probability densities (RPDs) of the $1S_{1/2}^{(e)} 1S_{3/2}^{(h)}$ exciton (single particle $1s_{1/2}^{(e)} 1s_{3/2}^{(h)}$ electron–hole pair [EHP]) of CdSe/CdTe quantum dots (QDs) with $a_c = 3.5$ nm and (a) $a_s = 0$, (b) $a_s = 0.5$ nm, and (c) $a_s = 1$ nm. Vertical lines denote the boundaries between the core, shell, and external medium. Insets are bar charts of single EHP characters, $|c_\beta|^2$, in the exciton.

by the effects of the type-II spatial confinement. In the SCR the charge density of a correlated excitonic state is very similar to the charge density of the uncorrelated product of SP states, $\rho_{\mathrm{X}}^{e(h)} \simeq \rho_{\mathrm{SP}}^{e(h)}$, and the effect of correlations is lost. Again, E_{corr} is only nonzero when the hole is delocalized; once it localizes in the shell, the effect of VBM confinement overrides the interparticle Coulomb attraction.

Overall, we find that dielectric confinement affects the correlated hole density more than the correlated electron density for two reasons. First, the larger effective mass and deeper potential well experienced by SP hole states compared to electron states allows the former to localize more fully in the shell, closer to the peak in $V_{\mathrm{s}}(r)$ at $r \simeq a_{\mathrm{c}} + a_{\mathrm{s}}$. Second, the smaller energy spacing between the hole SP basis states (i.e., the larger density of hole states) compared to electron SP basis states means that the resulting correlated hole density has more "degrees of freedom" to adjust to the effects of dielectric confinement.

13.8.4 Absorption Spectra of Equivalent CdTe and CdTe/CdSe QD

In Figure 13.9, we compare the exciton dipole spectrum of (a) an $a_{\mathrm{c}} = 2$ nm, $a_{\mathrm{s}} = 1$ nm CdTe/CdSe QD and (b) an "equivalent" $a_{\mathrm{c}} = 3.8$ nm CdTe QD calculated using the CI Hamiltonian. The radius of the CdTe core-only QD is chosen such that the absorption wavelength of its ground-state exciton is the same as that of the CdTe/CdSe core/shell QD. The energy gap between the $1S_{1/2}^{(e)}1S_{3/2}^{(h)}$ and $1S_{1/2}^{(e)}2S_{3/2}^{(h)}$ excitons is increased from ∼ 0.1 eV in the core-only QD to ∼ 0.18 eV in the core/shell structure. The size of the ground-state exciton optical dipole matrix element is reduced by about 30%, from $0.15P_0^2$ (where P_0 is the bulk optical dipole matrix element) in the core-only QD to $0.11P_0^2$ in the core/shell QD due to reduced electron–hole overlap and electron delocalization (the $a_{\mathrm{c}} = 2$ nm, $a_{\mathrm{s}} = 1$ nm QD lies in the quasi-type-II regime). As expected, the CdTe QD shows slightly stronger absorption than its CdTe/CdSe QD counterpart (see Figure 13.9c), mainly due to the better overall overlap between electron and hole states in core-only QDs compared to type-II structures. However, we note that type-II QDs overall have superior absorption

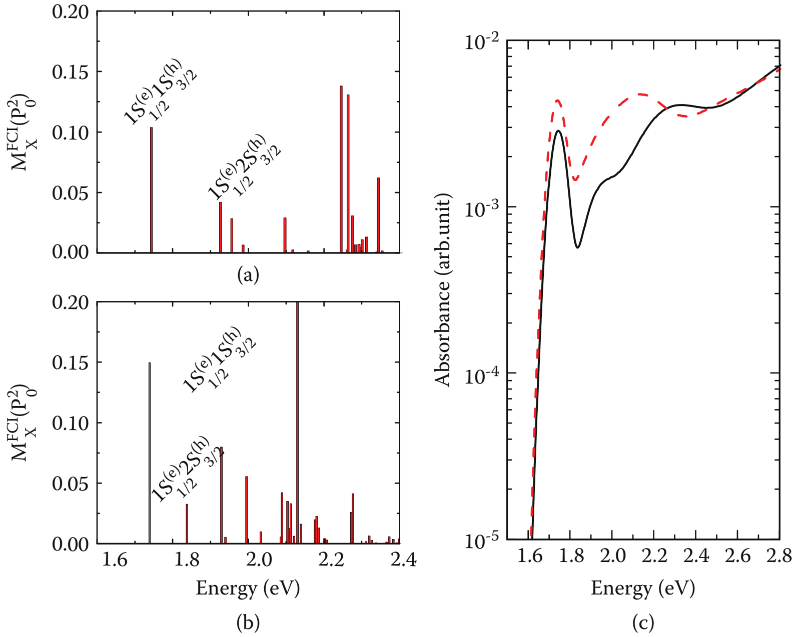


FIGURE 13.9 The exciton dipole spectrum of (a) an $a_{\mathrm{c}} = 2$ nm, $a_{\mathrm{s}} = 1$ nm CdTe/CdSe quantum dot (QD) and (b) an $a_{\mathrm{c}} = 3.8$ nm CdTe QD. (c) The absorption spectra corresponding to the dipole spectra shown in (a) and (b) are shown as solid and dashed lines, respectively.

properties compared with core-only QDs for the important application area of QD-sensitized solar cells. For example the type-II band alignment allows the band-edge absorption to be red-shifted compared to the core-only QD—this is often desirable because the optimum energy for exploitation of the solar spectrum is ~1.35 eV. Type-II QDs also allow greater absorption ranges to be achieved compared to core-only QDs since the absorption edges are not limited by the energy gap of the underlying bulk materials [94].

13.8.5 Charge Carrier Mobility in CdSe QD Arrays

In this section, we evaluate the charge carrier mobility in QD arrays. Good charge carrier mobility is essential for applications of QDs in solar cells since photogenerated carriers have to reach the contacts to form the current in external circuit.

We consider a three-dimensional lattice of CdSe QDs and calculate the temperature dependence of electron mobility in such a system. For hopping transport in a cubic lattice of QDs, the mobility is related to hopping rate between two neighboring dots via $\mu = \frac{ea^2}{k_B T} W$, where a is the distance between centers of two neighboring dots (lattice constant of a QD supercrystal) and W is the transition rate between the dots that can be evaluated using either Equation 13.67 or 13.70 from Section 13.7. If all dots are identical the mobility reads

$$\mu = \int_{-\infty}^{\infty} dt \frac{ea^2 t^2}{\hbar^2 k_B T} \prod_{s} \exp\left\{-2g_s^2 \left[2n_s + 1 - (n_s + 1)e^{-i\omega_s t} - n_s e^{i\omega_s t}\right]\right\} \tag{13.73}$$

and takes the form $\mu = \int_{-\infty}^{\infty} dt\, h(t)$. By exploiting the fact that $h(t) = h(-t)^*$, the last integral reduces to $\mu = \int_0^{\infty} dt\, f(t)$, where $f(t) = 2\,\mathrm{Re}\, h(t)$.

We consider two different dots. For dot 1, the diameter is $D = 3.6$ nm, relevant phonon energies and electron–phonon coupling constants are $\hbar\omega_s(1-6) = \{3.78, 8.10, 1.46, 2.87, 24.0, 24.0\}$ meV, $g_{1-6}\hbar\omega_s(1-6) = \{2.82, 0.14, 0.78, 0.52, 20.33, 3.52\}$ meV, the supercrystal lattice constant is $a = 5.0$ nm, and the electronic coupling parameter is $t = 4.0$ meV. For dot 2, the relevant parameters are $D = 7.0$ nm, $\hbar\omega_s(1-6) = \{1.95, 4.16, 0.75, 1.48, 24.0, 24.0\}$ meV, $g_{1-6}\hbar\omega_s(1-6) = \{1.06, 0.14, 0.40, 0.27, 14.58, 2.52\}$ meV, the supercrystal lattice constant is $a = 8.0$ nm, the electronic coupling parameter is $t = 2.5$ meV.

One might expect that it is difficult to numerically integrate the function $f(t)$ because of oscillatory terms of the form $\exp(i\omega_s t)$. However, in the strong electron–phonon coupling regime this function is smooth and decays with t very quickly. Its form for dot 1 at temperatures of $T = 200$ K and $T = 300$ K is presented in Figure 13.10.

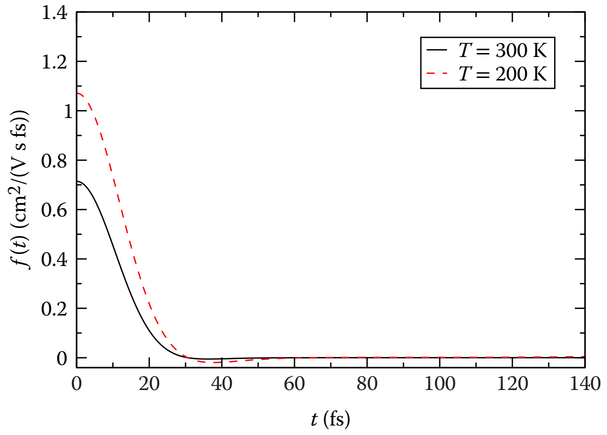


FIGURE 13.10 The function $f(t)$ for dot 1 at temperatures of $T = 200$ K and $T = 300$ K.

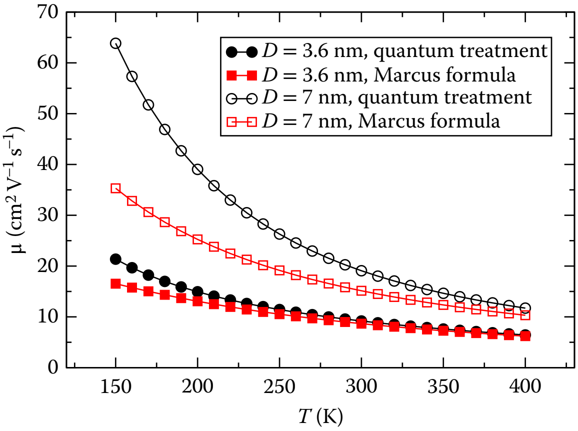


FIGURE 13.11 The dependence of mobility, μ, on temperature, T, for dots 1 and 2 with diameters D = 3.6 and 7 nm, respectively, calculated using Equations 13.70 and 13.67.

Temperature dependence of the mobility for dots 1 and 2 is presented in Figure 13.11. It has been checked that the assumption of strong electron–phonon coupling is valid in this temperature range [95]. The mobility was calculated using both the fully quantum treatment from Equation 13.67 and the Marcus formula given by Equation 13.70. We find that for this system in the relevant temperature range the mobility decreases with increasing temperature. This temperature dependence is often observed in the case of band transport. However, one should note that the charge transport in our case takes place by carrier hopping between neighboring dots. This suggests that the measurement of temperature dependence of mobility is not sufficient to identify the charge carrier transport regime. By comparing the results obtained using fully quantum treatment and Marcus formula, we find that, as expected, the agreement between these approaches is the best at high temperatures.

13.9 Conclusion

In conclusion, we have reviewed the methodology for calculation of SP, (multi)excitonic states, and charge transport in colloidal QDs. The methodology was then applied to analyze the optical absorption wavelength, radiative lifetimes, and correlation energy in CdSe/CdTe core/shell QDs, as well as charge carrier mobility in CdSe QD arrays.

Acknowledgments

Stanko Tomić acknowledges the EPSRC UK grant “Enhanced multiple exciton generation in colloidal quantum dots” (EP/K008587/1) for financial support. Nenad Vukmirović is supported by the Ministry of Education, Science, and Technological Development of the Republic of Serbia under project ON171017, and by the European Commission under H2020 project VI-SEEM, Grant No. 675121. We would also like to acknowledge the contribution of the COST Action MP1406 “MultiscaleSolar,” and we wish to thank David J. Binks, Jacek M. Miloszewski, Nikola Prodanović, and Edward J. Tyrrell for useful discussions.

References

1. B. N. Pal, Y. Ghosh, S. Brovelli, R. Laocharoensuk, V. I. Klimov, J. A. Hollingsworth, and H. Htoon. Giant CdSe/CdS core/shell nanocrystal quantum dots as efficient electroluminescent materials: Strong influence of shell thickness on light-emitting diode performance. *Nano Lett.* **12**, 331 (2012).

2. P. V. Kamat. Quantum dot solar cells. Semiconductor nanocrystals as light harvesters. *J. Phys. Chem. C* **112**, 18737 (2008).
3. V. Sukhovatkin, S. Hinds, L. Brzozowski, and E. H. Sargent. Colloidal quantum-dot photodetectors exploiting multiexciton generation. *Science* **324**, 1542 (2009).
4. G. Konstantatos, and E. H. Sargent. Nanostructured materials for photon detection *Nat. Nanotechnol.* **5**, 391 (2010).
5. J.-M. Caruge, Y. Chan, V. Sundar, H. J. Eisler, and M. G. Bawendi. Transient photoluminescence and simultaneous amplified spontaneous emission from multiexciton states in CdSe quantum dots. *Phys. Rev. B* **70**, 085316 (2004).
6. V. I. Klimov, S. A. Ivanov, J. Nanda, M. Achermannn, I. Bezel, J. A. McGuire, and A. Piryatinski. Single-exciton optical gain in semiconductor nanocrystals. *Nature* **447**, 441 (2007).
7. A. Shabaev, A. L. Efros, and A. J. Nozik. Multiexciton generation by a single photon in nanocrystals. *Nano Lett.* **6**, 2856 (2006).
8. O. E. Semonin, J. M. Luther, S. Choi, H.-Y. Chen, J. Gao, A. J. Nozik, and M. C. Beard. Peak external photocurrent quantum efficiency exceeding 100 in a quantum dot solar cell. *Science* **334**, 1530 (2011).
9. M. T. Trinh, L. Polak, J. M. Schins, A. J. Houtepen, R. Vaxenburg, G. I. Maikov, G. Grinbom, A. G. Midgett, J. M. Luther, M. C. Beard, A. J. Nozik, M. Bonn, E. Lifshitz, and L. D. A. Siebbeles. Anomalous independence of multiple exciton generation on different group IV–VI quantum dot architectures. *Nano Lett.* **11**, 1623 (2011).
10. S. Ithurria, M. D. Tessier, B. Mahler, R. P. S. M. Lobo, B. Dubertret, and A. L. Efros. Colloidal nanoplatelets with two-dimensional electronic structure. *Nat. Mater.* **10**, 936 (2011).
11. J. Jasieniak, M. Califano, and S. E. Watkins. Size-dependent valence and conduction band-edge energies of semiconductor nanocrystals. *ACS Nano* **5**, 5888 (2011).
12. S. Kim, B. Fisher, H.-J. Eisler, and M. Bawendi. Type-II quantum dots: CdTe/CdSe(core/shell) and CdSe/ZnTe(core/shell) heterostructures. *J. Am. Chem. Soc.* **125**, 11466 (2003).
13. J. J. Li, J. M. Tsay, X. Michalet, and S. Weiss. Wavefunction engineering: From quantum wells to near-infrared type-II colloidal quantum dots synthesized by layer-by-layer colloidal epitaxy. *Chem. Phys.* **318**, 82 (2005).
14. R. Xie, U. Kolb, J. Li, T. Basch, and A. Mews. Synthesis and characterization of highly luminescent CdSe-Core CdS/$Zn_{0.5}Cd_{0.5}S$/ZnS multishell nanocrystals. *J. Am. Chem. Soc.* **127**, 7480 (2005).
15. U. Aeberhard, R. Vaxenburg, E. Lifshitz, and S. Tomić. Fluorescence of colloidal PbSe/PbS QDs in NIR luminescent solar concentrators. *Phys. Chem. Chem. Phys.* **14**, 16223 (2012).
16. C. T. Smith, E. J. Tyrrell, M. A. Leontiadou, J. Miloszewski, T. Walsh, M. Cadirci, R. Page, P. O'Brien, D. Binks, and S. Tomić. Energy structure of CdSe/CdTe type {II}colloidal quantum dots: Do phonon bottle necks remain for thick shells? *Sol. Energ. Mater.* **158**, 160–167 (2016),
17. M. A. Leontiadou, E. J. Tyrrell, C. T. Smith, D. Espinobarro-Velazquez, R. Page, P. O'Brien, J. Miloszewski, T. Walsh, D. Binks, and S. Tomić. Influence of elevated radiative lifetime on efficiency of CdSe/CdTe type {II}colloidal quantum dot based solar cells. *Sol. Energ. Mater.* **159**, 657–663 (2017).
18. S. Kumar, M. Jones, S. Lo, and G. D. Scholes. Nanorod heterostructures showing photoinduced charge separation. *Small* **3**, 1633 (2007).
19. H. Zhong, Y. Zhou, Y. Yang, C. Yang, and Y. Li. Synthesis of Type II CdTe–CdSe nanocrystal heterostructured multiple-branched rods and their photovoltaic applications. *J. Phys. Chem. C* **111**, 6538 (2007).
20. S. Itzhakov, H. Shen, S. Buhbut, H. Lin, and D. Oron. Type-II quantum-dot-sensitized solar cell spanning the visible and near-infrared spectrum. *J. Phys. Chem. C* **117**, 22203 (2013).
21. N. McElroy, R. Page, D. Espinbarro-Valazquez, E. Lewis, S. Haigh, P. O'Brien, and D. Binks. Comparison of solar cells sensitised by CdTe/CdSe and CdSe/CdTe core/shell colloidal quantum dots with and without a CdS outer layer. *Thin Solid Films* **560**, 65 (2014).

22. S. Brovelli, R. D. Schaller, S. A. Crooker, F. Garcia-Santamaria, Y. Chen, R. Viswanatha, J. A. Hollingsworth, H. Htoon, and V. I. Klimov. Nano-engineered electron–hole exchange interaction controls exciton dynamics in core–shell semiconductor nanocrystals. *Nat. Commun.* **2**, 280 (2011).
23. D. Oron, M. Kazes, and U. Banin. Multiexcitons in type-II colloidal semiconductor quantum dots. *Phys. Rev. B* **75**, 035330 (2007).
24. P. G. McDonald, E. J. Tyrrell, J. Shumway, J. M. Smith, and I. Galbraith. Tuning biexciton binding and antibinding in core/shell quantum dots. *Phys. Rev. B* **86**, 125310 (2012).
25. S. A. Ivanov, J. Nanda, A. Piryatinski, M. Achermann, L. P. Balet, I. V. Bezel, P. O. Anikeeva, S. Tretiak, and V. I. Klimov. Light amplification using inverted core/shell nanocrystals: Towards lasing in the single-exciton regime. *J. Phys. Chem. B* **108**, 10625 (2004).
26. J. Nanda, S. A. Ivanov, M. Achermann, I. Bezel, A. Piryatinski, and V. I. Klimov. Light amplification in the single-exciton regime using exciton–exciton repulsion in type-II nanocrystal quantum dots. *J. Phys. Chem. C* **111**, 15382 (2007).
27. A. Szabo, and N. S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory* (New York, NY: Dover Publications, 1982).
28. V. A. Fonoberov, E. P. Pokatilov, and A. A. Balandin. Exciton states and optical transitions in colloidal CdS quantum dots: Shape and dielectric mismatch effects. *Phys. Rev. B* **66**, 085310 (2002).
29. E. Menéndez-Proupin, and C. Trallero-Giner. Electric-field and exciton structure in CdSe nanocrystals. *Phys. Rev. B* **69**, 125336 (2004).
30. N. Vukmirović, Ž. Gačević, Z. Ikonić, D. Indjin, P. Harrison, and V. Milanović. Electric-field and exciton structure in CdSe nanocrystals. *Semicond. Sci. Technol.* **21**, 1098 (2006).
31. N. Vukmirović, Z. Ikonić, I. Savić, D. Indjin, and P. Harrison. A microscopic model of electron transport in quantum dot infrared photodetectors. *J. Appl. Phys.* **100**, 074502 (2006).
32. N. Vukmirović, D. Indjin, Z. Ikonić, and P. Harrison. Origin of detection wavelength tuning in quantum dots-in-a-well infrared photodetectors. *Appl. Phys. Lett.* **88**, 251107 (2006).
33. N. Vukmirović, and S. Tomić. Plane wave methodology for single quantum dot electronic structure calculations. *J. Appl. Phys.* **103**, 103718 (2008).
34. N. Vukmirović, Z. Ikonić, V. D. Jovanović, D. Indjin, and P. Harrison. Optically pumped intersublevel mid-infrared lasers based on InAs/GaAs quantum dots. *IEEE J. Quant. Electron.* **41**, 1361 (2005).
35. M. Califano, A. Franceschetti, and A. Zunger. Lifetime and polarization of the radiative decay of excitons, biexcitons, and trions in CdSe nanocrystal quantum dots. *Phys. Rev. B* **75**, 115401 (2007).
36. S. Tomić, and N. Vukmirović. Excitonic and biexcitonic properties of single GaN quantum dots modeled by 8-band $\mathbf{k}\cdot\mathbf{p}$ theory and configuration-interaction method. *Phys. Rev. B* **79**, 245330 (2009).
37. M. Korkusinski, O. Voznyy, and P. Hawrylak. Fine structure and size dependence of exciton and biexciton optical spectra in CdSe nanocrystals. *Phys. Rev. B* **82**, 245304 (2010).
38. G. Allan, and C. Delerue. Tight-binding calculations of the optical properties of HgTe nanocrystals. *Phys. Rev. B* **86**, 165437 (2012).
39. A. L. Efros, and M. Rosen. The electronic structure of semiconductor nanocrystals. *Ann. Rev. Mater. Sci.* **30**, 475 (2000).
40. A. Franceschetti, A. Williamson, and A. Zunger. Addition spectra of quantum dots: The role of dielectric mismatch. *J. Phys. Chem. B* **104**, 3398 (2000).
41. J. I. Climente, M. Royo, J. L. Movilla, and J. Planelles. Strong configuration mixing due to dielectric confinement in semiconductor nanorods. *Phys. Rev. B* **79**, 161301 (2009).
42. L. E. Brus. A simple model for the ionization potential, electron affinity, and aqueous redox potentials of small semiconductor crystallites. *J. Chem. Phys.* **80**, 4403 (1984).
43. J. M. Luttinger, and W. Kohn. Motion of electrons and holes in perturbed periodic fields. *Phys. Rev.* **97**, 869 (1955).
44. E. O. Kane. Band structure of indium antimonide. *J. Phys. Chem. Solids* **1**, 249 (1957).
45. E. O. Kane. The $k \cdot p$ method. *Semiconduct. Semimet.* **1**, 75 (1966).

46. M. Cardona, and F. H. Pollak. Energy-band structure of germanium and silicon: The k · p method. *Phys. Rev.* **142**, 530 (1966).
47. D. Rideau, M. Feraille, L. Ciampolini, M. Minondo, C. Tavernier, H. Jaouen, and A. Ghetti. Strained Si, Ge, and $Si1{-}xGex$ alloys modeled with a first-principles-optimized full-zone k · p method. *Phys. Rev. B* **74**, 195208 (2006).
48. G. L. Bir, and G. E. Pikus. *Symmetry and Strain-Induced Effects in Semiconductors* (New York, NY: Wiley, 1974).
49. J. P. Loehr. *Physics of Strained Quantum Well Lasers* (New York, NY: Kluwer Academic Publishers, 1998).
50. C. R. Pidgeon, and R. N. Brown. Interband magneto-absorption and Faraday rotation in InSb. *Phys. Rev.* **146**, 575 (1966).
51. S. Tomić, and N. Vukmirović. Symmetry reduction in multiband Hamiltonians for semiconductor quantum dots: The role of interfaces and higher energy bands. *J. Appl. Phys.* **110**, 053710 (2011).
52. M. A. Cusack, P. R. Briddon, and M. Jaros. Electronic structure of InAs/GaAs self-assembled quantum dots. *Phys. Rev. B* **54**, R2300 (1996).
53. A. D. Andreev, and E. P. O'Reilly. Strain distributions in quantum dots of arbitrary shape. *Phys. Rev. B* **62**, 15851 (2000).
54. N. Vukmirović, D. Indjin, V. D. Jovanović, Z. Ikonić, and P. Harrison. Symmetry of $k \cdot p$ Hamiltonian in pyramidal InAs/GaAs quantum dots: Application to the calculation of electronic structure. *Phys. Rev. B* **72**, 075356 (2005).
55. S. Tomić, A. G. Sunderland, and I. J. Bush. Parallel multi-band $k \cdot p$ code for electronic structure of zinc blend semiconductor quantum dots. *J. Mater. Chem.* **16**, 1963 (2006).
56. N. Vukmirović, Z. Ikonić, D. Indjin, and P. Harrison. Symmetry-based calculation of single-particle states and intraband absorption in hexagonal GaN/AlN quantum dot superlattices. *J. Phys. Condens. Matter* **18**, 6249 (2006).
57. A. D. Andreev, J. R. Downes, D. A. Faux, and E. P. O'Reilly. Strain distributions in quantum dots of arbitrary shape. *J. Appl. Phys.* **86**, 297 (1999).
58. E. J. Tyrrell, and S. Tomić. Effect of correlation and dielectric confinement on $1s^{(e)}_{1/2}ns^{(h)}_{3/2}$ excitons in CdTe/CdSe and CdSe/CdTe type-II quantum dots. *J. Phys. Chem. C* **119**, 12720 (2015).
59. A. Franceschetti, H. Fu, L. W. Wang, and A. Zunger. Many-body pseudopotential theory of excitons in InP and CdSe quantum dots. *Phys. Rev. B* **60**, 1819 (1999).
60. Y. Z. Hu, M. Lindberg, and S. W. Koch. Theory of optically excited intrinsic semiconductor quantum dots. *Phys. Rev. B* **42**, 1713 (1990).
61. P. G. Bolcatto, and C. R. Proetto. Partially confined excitons in semiconductor nanocrystals with a finite size dielectric interface. *J. Phys. Condens. Matter* **13**, 319 (2001).
62. J. Movilla, and J. Planelles. Image charges in spherical quantum dots with an off-centered impurity: Algorithm and numerical results. *Comput. Phys. Comm.* **170**, 144 (2005).
63. C. G. Koay, J. E. Sarlls, and E. zarslan. Three-dimensional analytical magnetic resonance imaging phantom in the Fourier domain. *Magn. Reson. Med.* **58**, 430 (2007).
64. C. Xue, and S. Deng. Three-layer dielectric models for generalized coulomb potential calculation in ellipsoidal geometry. *Phys. Rev. E* **83**, 056709 (2011).
65. S. Deng. A robust numerical method for self-polarization energy of spherical quantum dots with finite confinement barriers. *Comput. Phys. Commun.* **181**, 787 (2010).
66. A. Sihvola, and I. Lindell. Polarizability and effective permittivity of layered and continuously inhomogeneous dielectric ellipsoids. *J. Electromagnet. Wave.* **4**, 1 (1990).
67. T. Takagahara. Effects of dielectric confinement and electron–hole exchange interaction on excitonic states in semiconductor quantum dots. *Phys. Rev. B* **47**, 4569 (1993).
68. E. Hanamura, and H. Haug. Condensation effects of excitons. *Phys. Rep.* **33**, 209 (1977).

69. G. Nan, X. Yang, L. Wang, Z. Shuai, and Y. Zhao. Nuclear tunneling effects of charge transport in rubrene, tetracene, and pentacene. *Phys. Rev. B* **79**, 115203 (2009).
70. S. H. Lin, C. H. Chang, K. K. Liang, R. Chang, Y. J. Shiu, J. M. Zhang, T.-S. Yang, M. Hayashi, and F. C. Hsu. *Ultrafast Dynamics and Spectroscopy of Bacterial Photosynthetic Reaction Centers* (New York, NY: John Wiley & Sons, Inc., 2002), pp. 1–88.
71. I.-H. Chu, M. Radulaski, N. Vukmirovic, H.-P. Cheng, and L.-W. Wang. Charge transport in a quantum dot supercrystal. *J. Phys. Chem. C* **115**, 21409 (2011).
72. C. de Mello Donegá, and R. Koole. Size dependence of the spontaneous emission rate and absorption cross section of CdSe and CdTe quantum dots. *J. Phys. Chem. C* **113**, 6511 (2009).
73. W. W. Yu, L. Qu, W. Guo, and X. Peng. Experimental determination of the extinction coefficient of CdTe, CdSe, and CdS nanocrystals. *Chem. Mater.* **15**, 2854 (2003), doi:10.1021/cm034081k.
74. T. Rajh, O. I. Micic, and A. J. Nozik. Synthesis and characterization of surface-modified colloidal cadmium telluride quantum dots. *J. Phys. Chem.* **97**, 11999 (1993), doi:10.1021/j100148a026.
75. C. B. Murray, D. J. Norris, and M. G. Bawendi. Synthesis and characterization of nearly monodisperse CdE (e = sulfur, selenium, tellurium) semiconductor nanocrystallites. *J. Am. Chem. Soc.* **115**, 8706 (1993), doi:10.1021/ja00072a025.
76. X. Peng, J. Wickham, and A. P. Alivisatos. Kinetics of II–VI and III–V colloidal semiconductor nanocrystal growth: "Focusing" of size distributions. *J. Am. Chem. Soc.* **120**, 5343 (1998), doi:10.1021/ja9805425.
77. A. L. Rogach, A. Kornowski, M. Gao, A. Eychmuller, and H. Weller. Synthesis and characterization of a size series of extremely small thiol-stabilized CdSe nanocrystals. *J. Phys. Chem. B* **103**, 3065 (1999), doi:10.1021/jp984833b.
78. V. N. Soloviev, A. Eichhofer, D. Fenske, and U. Banin. Molecular limit of a bulk semiconductor: Size dependence of the band gap in CdSe cluster molecules. *J. Am. Chem. Soc.* **122**, 2673 (2000), doi:10.1021/ja9940367.
79. F. V. Mikulec, M. Kuno, M. Bennati, D. A. Hall, R. G. Griffin, and M. G. Bawendi. Organometallic synthesis and spectroscopic characterization of manganese-doped CdSe nanocrystals. *J. Am. Chem. Soc.* **122**, 2532 (2000), doi:10.1021/ja991249n.
80. Y. Masumoto, and K. Sonobe. Size-dependent energy levels of CdTe quantum dots. *Phys. Rev. B* **56**, 9734 (1997).
81. D. V. Talapin, S. Haubold, A. L. Rogach, A. Kornowski, M. Haase, and H. Weller. A novel organometallic synthesis of highly luminescent CdTe nanocrystals. *J. Phys. Chem. B* **105**, 2260 (2001), doi:10.1021/jp003177o.
82. P. Dagtepe, V. Chikan, J. Jasinski, and V. J. Leppert. Quantized growth of CdTe quantum dots; observation of magic-sized CdTe quantum dots. *J. Phys. Chem. C* **111**, 14977 (2007).
83. E. J. Tyrrell, and J. M. Smith. Effective mass modeling of excitons in type-II quantum dot heterostructures. *Phys. Rev. B* **84**, 165328 (2011).
84. K. Gong, Y. Zeng, and D. F. Kelley. Extinction coefficients, oscillator strengths, and radiative lifetimes of CdSe, CdTe, and CdTe/CdSe nanocrystals. *J. Phys. Chem. C* **117**, 20268 (2013).
85. X. Ma, A. Mews, and T. Kipp. Determination of electronic energy levels in type-II CdTe-core/CdSe-shell and CdSe-core/CdTe-shell nanocrystals by cyclic voltammetry and optical spectroscopy. *J. Phys. Chem. C* **117**, 16698 (2013).
86. X. Cai, H. Mirafzal, K. Nguyen, V. Leppert, and D. F. Kelley. Spectroscopy of CdTe/CdSe type-II nanostructures: Morphology, lattice mismatch, and band-bowing effects. *J. Phys. Chem. C* **116**, 8118 (2012).
87. E. A. Lewis, R. C. Page, D. J. Binks, T. J. Pennycook, P. O'Brien, and S. J. Haigh. Probing the core–shell–shell structure of CdSe/CdTe/CdS type II quantum dots for solar cell applications. *J. Phys.* **522**, 012069 (2014).

88. C.-Y. Chen, C.-T. Cheng, C.-W. Lai, Y.-H. Hu, P.-T. Chou, Y.-H. Chou, and H.-T. Chiu. Type-II CdSe/CdTe/ZnTe (core/shell/shell) quantum dots with cascade band edges: The separation of electron (at CdSe) and hole (at ZnTe) by the CdTe layer. *Small* **1**, 1215 (2005).
89. B. D. Geyter, Y. Justo, I. Moreels, K. Lambert, P. F. Smet, D. V. Thourhout, A. J. Houtepen, D. Grodzinska, C. de Mello Donega, A. Meijerink, D. Vanmaekelbergh, and Z. Hens. The different nature of band edge absorption and emission in colloidal PbSe /CdSe core/shell quantum dots. *ACS Nano* **5**, 58 (2011), PMID: 21189031.
90. D. Espinobarro-Velazquez, M. A. Leontiadou, R. C. Page, M. Califano, P. O'Brien, and D. J. Binks. Effect of chloride passivation on recombination dynamics in CdTe colloidal quantum dots. *Chem. Phys. Chem.* **16**, 1239 (2015).
91. K. Nishikawa, Y. Takeda, T. Motohiro, D. Sato, J. Ota, N. Miyashita, and Y. Okada. Extremely long carrier lifetime over 200ns in GaAs wall-inserted type II InAs quantum dots. *Appl. Phys. Lett.* **100**, 113105 (2012).
92. S. Tomić. Effect of Sb induced type II alignment on dynamical processes in InAs/GaAs/GaAsSb quantum dots: Implication to solar cell design. *Appl. Phys. Lett.* **103**, 072112 (2013).
93. L. E. Brus. A simple model for the ionization potential, electron affinity, and aqueous redox potentials of small semiconductor crystallites. *J. Chem. Phys.* **79**, 5566 (1983).
94. S. Tomić, J. M. Miloszewski, E. J. Tyrrell, and D. J. Binks. Design of core/shell colloidal quantum dots for meg solar cells. *IEEE J. Photovolt.* **6**, 179 (2016).
95. N. Prodanović, N. Vukmirović, Z. Ikonić, P. Harrison, and D. Indjin. Synthesis and characterization of surface-modified colloidal cadmium telluride quantum dots. *J. Phys. Chem. Lett.* **5**, 1335 (2014).

IV

Light-Emitting Diodes (LEDs)

14 Light-Emitting Diode Fundamentals *Sergey Yu. Karpov* 451
Introduction • LED Family • Carrier Transport and Recombination in LED Heterostructures • Current Spreading, Heat Transfer, and Light Extraction in LED Dice • Phosphor-Converted White Light Emission • Concluding Remarks

15 Organic Light-Emitting Diodes *Pascal Kordt, Peter Bobbert, Reinder Coehoorn, Falk May, Christian Lennartz, and Denis Andrienko* 473
Introduction • Working Principles of an OLED • Drift–Diffusion Equations • Master Equation • Kinetic Monte Carlo • Rates • Density of States • Reorganization Energy • Electronic Coupling Elements • Morphology • Scale Bridging • Case Studies • Outlook

16 Tunnel-Junction Light-Emitting Diodes *Yen-Kuang Kuo, Jih-Yuan Chang, Ya-Hsuan Shih, Fang-Ming Chen, and Miao-Chan Tsai* 523
Introduction • Physical Models and Simulation Parameters • III-Nitride Tunnel Junction • Tunnel-Junction LED • Conclusion

17 Quantum Disk Nanowire Light-Emitting Diodes *Fabio Sacconi* 541
Introduction • Overview of Numerical Methods • Strain Maps • Transport Properties • Quantum Calculations: EFA Models • Quantum Calculations: Empirical Tight-Binding Approach • Self-Consistent Calculations • Modeling of Surface States • Sensitivity to Geometrical and Material Parameters • Conclusions

18 Influence of Random InGaN Alloy Fluctuations on GaN-Based Light-Emitting Diodes *Chen-Kuo Wu, Tsung-Jui Yang, and Yuh-Renn Wu* 559
Introduction to Random Alloy Distribution • Methodology in Random Alloy Modeling • Unipolar Transport for Random Alloy System • Modeling Result of MQW InGaN LED and the Comparison to the Traditional Model • Carrier Transport in Green LEDs: Influence of Possible Imperfection in QWs • Summary

19 Superluminescent Light-Emitting Diodes *Nicolai Matuschek and Marcus Duelk* 589
Introduction • Design and Modeling of SLED Devices • SLED Designs with Specified Targets • Summary

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

14

Light-Emitting Diode Fundamentals

14.1 Introduction 451
14.2 LED Family 452
14.3 Carrier Transport and Recombination in LED Heterostructures 453
Electromechanical Coupling • Transport Equations • Recombination Models • Efficiency of Light Emission
14.4 Current Spreading, Heat Transfer, and Light Extraction in LED Dice 464
Current Crowding • Surface Recombination
14.5 Phosphor-Converted White Light Emission 467
White Light Characteristics • Light Conversion by Phosphor
14.6 Concluding Remarks 470

Sergey Yu. Karpov

14.1 Introduction

Due to the breakthroughs in technological advances in the development of III-nitride semiconductors made in the early 1990s, visible light-emitting diodes (LEDs) now occupy the dominant sector of the optoelectronics market. The emerging LED production is also stimulated by ever-increasing demands from solid-state lighting, which is expected to not only reduce world energy consumption but also change human conceptions of light quality and ways to utilize it. However, because of the constantly increasing capacity of production equipment and, in particular, of growth reactors, the cost of research and development in the LED industry continues to increase at a very high rate, making comprehensive investigations virtually impossible for all but the biggest universities and industrial players. In this situation, modeling and simulation are especially important, saving time, money, and manpower at the research and development stage.

Simulation of LEDs is essentially a multiscale and multidisciplinary problem. The relevant specific sizes range from a few nanometers for the widths of LED active regions to a few centimeters for the LED lamp dimensions, that is, about seven orders of magnitude. This means that, electromechanical, electrical, thermal, and optical phenomena involved in LED operations are often considered to be self-consistent and, in the case of state-of-the-art LEDs, in three-dimensional (3D) formulation. In practice, this makes straightforward LED simulations unacceptably time- and computer resource-consuming. The commonly utilized approach to resolving this problem and making simulation a helpful research and engineering tool is to split the task into independent but interrelated subtasks. This approach correlates three stages of LED fabrication: heterostructure growth, chip fabrication, and lamp assembly. We choose the LED structure, chip, and lamp as the major objects of investigation within the above-mentioned subtasks.

This chapter focuses on fundamental mechanisms underlying the LED operation that should be included in the simulation models. LED heterostructures, chips, and lamps are considered separately, with a discussion on links between them. Here, inorganic LEDs will be considered only when the organic LED models require special modifications.

14.2 LED Family

There is no wonder that most LED applications are related to various kinds of lighting, and thus utilize the visible optical spectrum. The visible spectral range is shared by phosphide LEDs with $(\mathrm{Al}_x\mathrm{Ga}_{1-x})_{0.51}\mathrm{In}_{0.49}\mathrm{P}$ active regions (emission wavelength $\lambda = 555-625$ nm), usually lattice matched to the GaAs substrates and nitride LEDs with InGaN-based active regions ($\lambda = 400-600$ nm). Beyond the visible spectral range, the most important region is the ultraviolet (UV) region where UVA ($\lambda = 315-400$ nm), UVB ($\lambda = 280-315$ nm), and UVC ($\lambda = 100-280$ nm) diapasons are commonly distinguished. The active regions of UV LEDs are normally based on AlGaN compounds, with the emission wavelength decreasing with the Al content. These LEDs are used for resin curing, sterilization of water, air, and food, and in medical and biotechnology applications. Infrared (IR) LEDs are situated at the other siide of the visible spectral range, with the active region made up of AlGaInAs and GaInAsSb compounds. They are used in niche applications such as IR lighting, remote control, and gas sensing with mid-IR LEDs ($\lambda = 1.8-5.0$ μm).

LED efficiency is the most important characteristic of the devices (see Section 14.3.4). It is also simultaneously a measure for the maturity of their fabrication technology. Figure 14.1 summarizes the maximum external quantum efficiency (EQE) achieved to date in different spectral ranges by LEDs from various industrial companies and leading research centers. There are two distinct EQE maxima: one of ~77%–82% has been reported for InGaN-based LEDs ($\lambda = 410-440$ nm), whereas another of ~71% has been attained by red AlGaInP ($\lambda = 625$ nm) and IR AlGaAs ($\lambda = 850$ nm) devices. Between them, that is, in the spectral range of 500–600 nm, a so-called “green gap” is seen where EQE drops to ~10%–30% in both nitride and phosphide LEDs. Electron leakage into p-type layers has been proved to be responsible for the efficiency

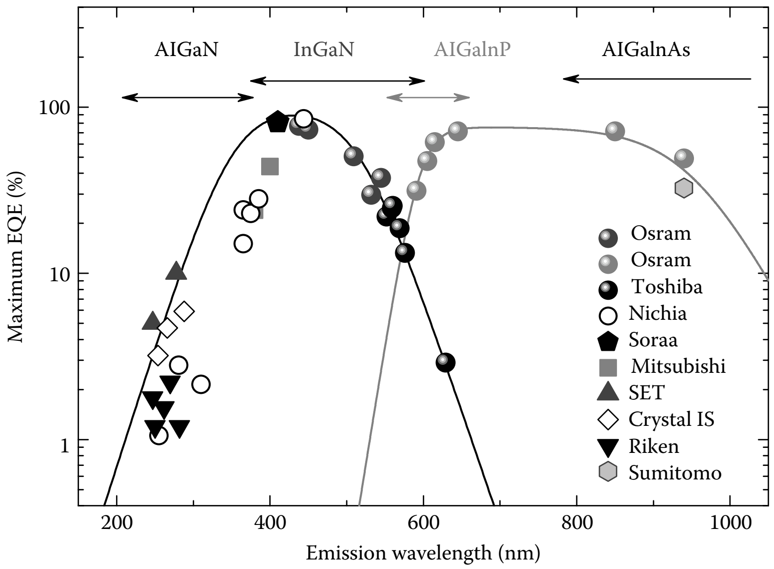


FIGURE 14.1 Maximum external quantum efficiencies (EQEs) of light-emitting diodes (LEDs) as a function of their emission wavelength (symbols are data compiled from numerous sources, lines are drawn for eye). Arrows indicate the ranges of utilizing various III–V compounds for the active region fabrication.

reduction in phosphide LED (Altieri et al., 2005), whereas the nature of the "green gap" in nitride LEDs is not yet completely understood.

Figure 14.1 shows that EQE of AlGaN-based ultraviolet (UV) LEDs decrease dramatically with the Al content in the alloy (at shorter λ), which is attributed to (1) insufficiently high light extraction efficiency partly related to strong transverse-magnetic polarization of the emitted light and absence of highly reflective metallic contacts in this spectral range, (2) problems of p-doping of AlGaN alloys with high Al content, and (3) higher defect density in the grown materials, as compared to Al-free nitride compounds.

Efficiency of IR LEDs suffers primarily from nonradiative Auger recombination inherent in narrow-bandgap semiconductors. In particular, EQE of mid-IR GaInAsSb LEDs emitting in the spectral range of 2.0–4.5 μm (not shown in Figure 14.1) does not exceed ~0.5%–3.0% at room temperature.

To date, as shown earlier, nitride and phosphide LEDs have attained the best performance in terms of emission efficiency. Therefore, simulation examples given in this chapter focus primarily on these two material systems.

14.3 Carrier Transport and Recombination in LED Heterostructures

Any LED structure consists of a stack of epitaxial layers grown in a certain sequence on a substrate of choice. At an arbitrary orientation of heterostructure growth surface, the epilayer interfaces do not always correspond to the main low-index facets of the grown crystals. Therefore, it is convenient to distinguish between the epitaxial coordinate system (ECS) $x'y'z'$, with the x' and y' axes lying in the interface plane and the z'-axis being normal to this plane, and the crystal coordinate system (CCS) xyz, with the axes corresponding to the symmetry axes of the crystal (see Figure 14.2). While ECS is suitable for analysis of carrier transport across the LED structure, CCS enables easy description of anisotropic properties of the crystalline materials (Nye, 1964).

Any crystal of wurtzite symmetry has elastic properties isotropic in the plane normal to its hexagonal C-axes. If the CCS z-axis corresponds to the C-axis of the crystal, the other x- and y-axes can be chosen rather arbitrarily. It is convenient to choose the x-axis of CCS and the x'-axis of ECS to coincide with each other and to be normal to both z- and z'-axes (Romanov et al., 2006). Then the y'- or y-axis orthogonal to both x'- and z'- or x- and z-axes can be unambiguously determined. In ECS defined in such a way, the in-plane lattice constant mismatch can be characterized by two parameters corresponding to the lattice

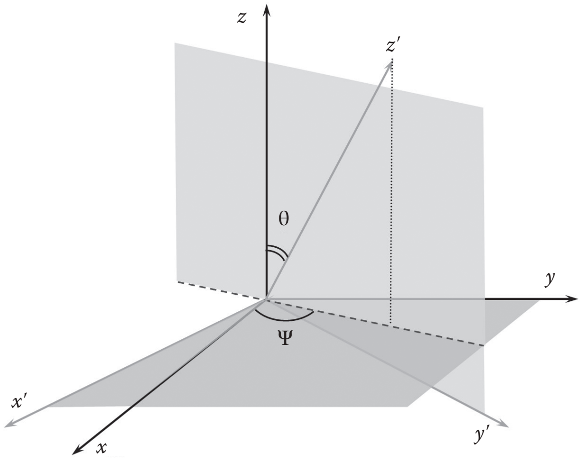


FIGURE 14.2 Crystal coordinate system (CCS) xyz, epitaxial coordinate system (ECS) $x'y'z'$, and specific angles between their axes.

constant variation along the x'- and y'-axes (Romanov et al., 2006):

$$\eta_x = 1 - a_S/a_E \quad \text{and} \quad \eta_y = 1 - [(a_S/a_E)^2 \cos^2\theta + (c_S/c_E)^2 \sin^2\theta]^{1/2}, \tag{14.1}$$

where a_E and c_E are the lattice constant of the epitaxial layer considered, a_S and c_S are the lattice constants of the substrate or a thick template layer on which the LED structure is coherently grown, and θ is the polar angle between the z'- and z-axes (see Figure 14.2). This angle can be expressed via four-digit Miller indices $[klmh]$ $(m = -k-l)$ characterizing the orientation of the LED structure interfaces (Gil, 2009):

$$\cos\theta = \frac{h}{[h^2 + \frac{4}{3}(c_E/a_E)^2(k^2 + l^2 + kl)]^{1/2}} \tag{14.2}$$

In the case of zinc-blende materials having no in-plane isotropy of elastic properties, the x'-axis normal to both z'- and z-axes does not coincide with the x-axis of CCS but lies in the xy-plane (Figure 14.2). Then the angles θ and ψ (see Figure 14.2) can be expressed via three-digit Miller indexes $[klm]$ as follows:

$$\cos\theta = \frac{m}{(k^2 + l^2 + m^2)^{1/2}}, \quad \cos\psi = \frac{l}{(k^2 + l^2)^{1/2}} \tag{14.3}$$

The lattice mismatch in zinc-blende crystals is described by the only parameter η_x (see Equation 14.1), where a_E and a_S denote the lattice constants of the epilayer and substrate/template, respectively.

Generally, epitaxial layers may have the lattice constants highly mismatched with those of substrate/template layer. The mismatch produces elastic strain affecting the band structure of the material, its electric polarization, and the band offsets at the structure interfaces. Therefore, determination of strain distribution in an LED structure is the primary task in analysis of its operation.

14.3.1 Electromechanical Coupling

Under isothermal conditions, the equations for elastic strain, stress, and electric polarization in every epitaxial layer of the LED structure can be written in matrix form (Nye, 1964):

$$\boldsymbol{\sigma} = \hat{C}\boldsymbol{\varepsilon} - \hat{e}^T\mathbf{E} \text{ (a)}, \quad \boldsymbol{\varepsilon} = \hat{S}\boldsymbol{\sigma} + \hat{d}^T\mathbf{E} \text{ (b)}, \quad \mathbf{D} = \mathbf{P}_{sp} + \hat{d}\boldsymbol{\sigma} + \kappa_0\hat{\kappa}\mathbf{E} \text{ (c)}. \tag{14.4}$$

Here, $\boldsymbol{\sigma}$ is the vector built up of six components of the stress tensor (σ_{xx}, σ_{yy}, σ_{zz}, σ_{yz}, σ_{xz}, and σ_{xy}), $\boldsymbol{\varepsilon}$ is the vector consisting of the diagonal and doubled nondiagonal components of the strain tensor (u_{xx}, u_{yy}, u_{zz}, $2u_{yz}$, $2u_{xz}$, and $2u_{xy}$), $\mathbf{E}$ and $\mathbf{D}$ are the electric field and induction vectors, respectively, and $\mathbf{P}_{sp}$ is the spontaneous electric polarization that does not vanish in wurtzite crystals. The first and second of Equations 14.4 correspond to the direct and reverse Hook laws, where $\hat{C}$ and $\hat{S}$ are the 6×6 matrices of stiffness and compliance constants having a standard form in CCS (Nye, 1964); $\hat{C} \cdot \hat{S} = \hat{S} \cdot \hat{C} = \hat{I}$, where $\hat{I}$ is the 6×6 unity matrix. Here, $\hat{d}$ and $\hat{e}$ are the 3×6 matrices of piezoelectric coefficients ($\hat{e} = \hat{d} \cdot \hat{C}$ and $\hat{d} = \hat{e} \cdot \hat{S}$) having a standard form in CCS too (Nye, 1964), and the superscript "T" denotes transposition of a matrix. The third of Equations Equation 14.4 links the electric field, induction, and stress vectors; κ_0 is the electric constant; and $\hat{\kappa}$ is the 3×3 matrix of static dielectric constants, being diagonal in CCS for both zinc-blende and hexagonal crystals.

The strain components in every epitaxial layer can be determined as follows (Romanov et al., 2006). First, the stress components are assumed to be uniform within the layer, which is valid, if the lateral dimensions of an LED structure are much larger than the vertical ones and bending of the structure is negligible. Second, the components of the stress vector in ECS $\sigma_{z'z'}$, $\sigma_{y'z'}$, and $\sigma_{x'z'}$ are regarded as vanishing according to the absence of external mechanical loading of the top heterostructure surface. Third, the shear component $\sigma_{x'y'}$ is also equal to zero due to uniformity of the strain distribution in the x'- and y'-directions. Then the strain vector contains only two components, $\sigma_{x'x'}$ and $\sigma_{y'y'}$. To determine them, the respective components of the strain vector $\varepsilon_{x'x'}$ and $\varepsilon_{y'y'}$ should be expressed through $\sigma_{x'x'}$ and $\sigma_{y'y'}$ using Equation 14.4 in ECS. These components are known from geometrical consideration, that is, $\varepsilon_{x'x'} = -\eta_x$ and $\varepsilon_{y'y'} = -\eta_y$, which enables the determination of two nonzero components $\sigma_{x'x'}$ and $\sigma_{y'y'}$. As soon as they are found other components of the strain vector $\boldsymbol{\varepsilon}$ in ECS can be calculated using Equation 14.4 and then transformed into CCS in a standard way.

In the simplest case of wurtzite crystals, the above procedure provides the analytical solution, which is formulated as follows, neglecting the electric field contribution to the stress components:

$$\sigma_{x'x'} = \sigma_1 = -\frac{S'_{22}\eta_x - S'_{12}\eta_y}{S'_{11}S'_{22} - (S'_{12})^2}, \quad \sigma_{y'y'} = \sigma_2 = -\frac{S'_{11}\eta_y - S'_{12}\eta_x}{S'_{11}S'_{22} - (S'_{12})^2}, \tag{14.5}$$

where $S'_{11} = S_{11}$, $S'_{12} = S_{12}\cos^2\theta + S_{13}\sin^2\theta$ and $S'_{22} = S_{11}\cos^2\theta + S_{33}\sin^2\theta + \frac{1}{4}(2S_{13} + S_{44} - S_{11} - S_{33})\sin^2 2\theta$.

Then the strain components in CCS can be obtained from Equation 14.5:

$$\begin{aligned} &\varepsilon_{xx} = -\eta_a; \quad \varepsilon_{xz} = \varepsilon_{xy} = 0 \\ &\varepsilon_{yy} = S_{12}\sigma_1 + (S_{11}\cos^2\theta + S_{13}\sin^2\theta)\,\sigma_2 \\ &\varepsilon_{zz} = S_{13}\sigma_1 + (S_{13}\cos^2\theta + S_{33}\sin^2\theta)\,\sigma_2 \\ &\varepsilon_{yz} = -\tfrac{1}{2}S_{44}\sigma_2\sin 2\theta. \end{aligned} \tag{14.6}$$

Nonzero stress components in zinc-blende crystals depend generally on both angles θ and ψ. Using the above approach, one can obtain the strain components in CCS for three practically important orientations of the LED structure interfaces, that is,

- for (001) orientation:

$$\varepsilon_{xx} = \varepsilon_{yy} = -\eta_x, \quad \varepsilon_{zz} = -\frac{2S_{12}}{S_{11} + S_{12}}\eta_x, \quad \varepsilon_{xz} = \varepsilon_{xy} = \varepsilon_{yz} = 0; \tag{14.7}$$

- for (111) orientation:

$$\varepsilon_{xx} = \varepsilon_{yy} = \varepsilon_{zz} = -\eta_x\frac{S_{11} + 2S_{12}}{S_{11} + 2S_{12} + \frac{1}{4}S_{44}}, \quad \varepsilon_{xz} = \varepsilon_{xy} = \varepsilon_{yz} = \eta_x\frac{\frac{1}{2}S_{44}}{S_{11} + 2S_{12} + \frac{1}{4}S_{44}}; \tag{14.8}$$

- for (0*lm*) orientations: strain components can be found from Equation 14.6), assuming $S_{33} = S_{11}$, $S_{13} = S_{12}$, and finding the angle θ from Equation 14.3); in particular, θ = 45° for the (011) orientation.

The obtained strain components in CCS can then be used for calculating strain-dependent band structure and electric polarization of every epitaxial layer. In wurtzite crystals, the components of the total electric polarization vector **P** in ECS are

$$\begin{aligned} P_{x'} &= 0 \\ P_{y'} &= e_{15}\varepsilon_{yz}\cdot\cos\theta - [P_{\text{sp}} + e_{31}(\varepsilon_{xx} + \varepsilon_{yy}) + e_{33}\varepsilon_{zz}]\cdot\sin\theta \\ P_{z'} &= [P_{sp} + e_{31}(\varepsilon_{xx} + \varepsilon_{yy}) + e_{33}\varepsilon_{zz}]\cdot\cos\theta + e_{15}\varepsilon_{yz}\cdot\sin\theta. \end{aligned} \tag{14.9}$$

In zinc-blende crystals,

$$
\begin{aligned}
&P_{x'} = P_{y'} = P_{z'} = 0, && \text{for (001) orientation} \\
&P_{x'} = P_{y'} = 0 \text{ and } P_{z'} = \sqrt{3}e_{14}\varepsilon_{yz}, && \text{for (111) orientation} \\
&P_{x'} = e_{14}\varepsilon_{yz} \text{ and } P_{y'} = P_{z'} = 0, && \text{for } (0lm) \text{ orientation.}
\end{aligned}
\tag{14.10}
$$

The difference in the z'-component of the electric polarization vector $\Delta P_{z'}$ of neighboring layers determines the polarization charge accumulated at their interface. Figure 14.3a shows $\Delta P_{z'}$ at the interfaces between the $In_{0.2}Ga_{0.8}N$, $Al_{0.2}Ga_{0.8}N$, $In_{0.17}Al_{0.83}N$, and GaN layer as a function of the C-axis inclination angle θ (the sign of $\Delta P_{z'}$ corresponds to the sign of the interface charge). To calculate $\Delta P_{z'}$ for the nitride alloys, Vegard's rule was used for estimation of the compliance constants S_{ij}. The application of Vegard's rule to the stiffness constants C_{ij} had proved to provide the results graphically indistinguishable from those shown in Figure 14.3a. The $\Delta P_{z'}$ values plotted in Figure 14.3a refine the similar results reported by Romanov et al. 2006, as the nondiagonal component of the strain tensor u_{yz} was erroneously used in this study instead of ε_{yz}.

An alternative approach to analysis of electromechanical coupling in nitride compounds is based on minimization of the elastic energy in every epitaxial layer (Park and Chuang, 1999). This approach provides incorrect results as the predicted stress does not correspond to unloaded free surface of the LED structure and its interfaces.

Equation 14.10 relevant to zinc-blende semiconductors shows, in particular, that only the (111)-orientation possesses polarization charges at the LED structure interfaces.

14.3.2 Transport Equations

In the hierarchy of transport models, the drift-diffusion one (DDM) was first successfully applied to electron and hole transport in semiconductor devices. Introduced in the mid-twentieth century (Van Roosbroeck, 1950), DDM is commonly recognized as the best compromise between the predictability of the theory and the cost of its numerical implementation.

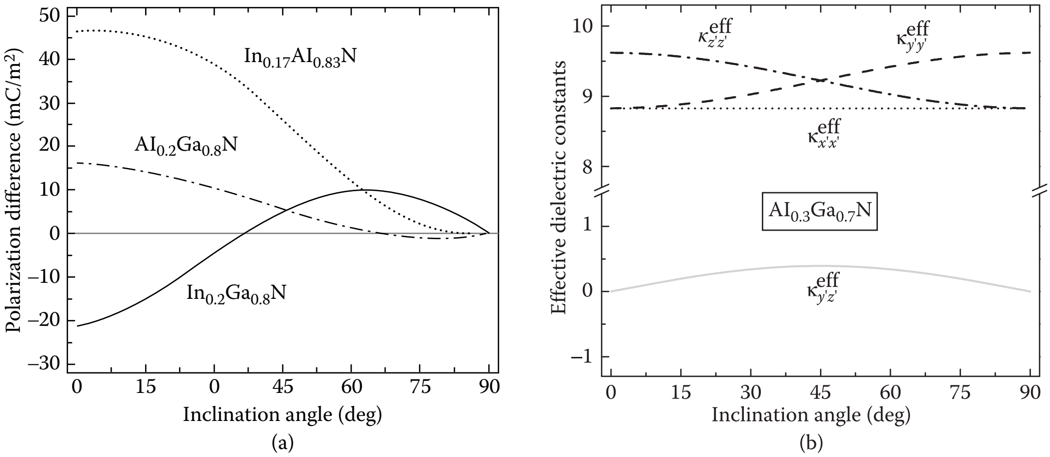


FIGURE 14.3 Difference between the electric polarization of III-nitride alloys and GaN (a) and components of the dielectric constant matrix of $Al_{0.3}Ga_{0.7}N$ (b) in the epitaxial coordinate system (ECS) as a function of inclination angle θ.

Within DDM, the electron n and hole p concentrations can be found from the continuity equations solved in ECS:

$$\frac{\partial n}{\partial t} + \nabla \cdot \mathbf{J_n} = G - R, \quad \frac{\partial p}{\partial t} + \nabla \cdot \mathbf{J_p} = G - R, \tag{14.11}$$

where G is the carrier generation rate, for example, due to light absorption, R is the electron/hole recombination rate, and $\mathbf{J_n}$ and $\mathbf{J_p}$ are the electron and hole fluxes, respectively, related to their partial current densities $\mathbf{j_n}$ and $\mathbf{j_p}$ as follows: $\mathbf{j_n} = -q\mathbf{J_n}, \mathbf{j_p} = q\mathbf{J_p}$; q is the elementary charge.

The most compact definition of the carrier fluxes $\mathbf{J_n}$ and $\mathbf{J_p}$ is based on linear nonequilibrium thermodynamics, involving electron (F_n) and hole (F_p) quasi-Fermi levels measured in electron volts (eV):

$$\mathbf{J_n} = -\mu_n n \nabla F_n, \quad \mathbf{J_p} = -\mu_p p \nabla F_p. \tag{14.12}$$

Here, scalar electron (μ_n) and hole (μ_p) mobilities are used as there is no distinct evidence in the literature for noticeable anisotropy of carrier electric conductivities in various LED structures. The expression 14.12 for the carrier fluxes account for all the forces driving the carrier movement if the electron and hole concentrations are defined as

$$n = N_C \cdot \bar{\ }_n \left(\frac{F_n - E_C + q\phi}{kT} \right)^{\mathbb{1}}, \quad p = \sum_s N_s \cdot \bar{\ }_s \left(\frac{E_s - F_p - q\phi}{kT} \right)^{\mathfrak{I}}. \tag{14.13}$$

Here, E_C is the conduction band energy determined from the known electron affinity of every epitaxial layer in the LED structure, E_s is the energy of the sth valence sub-band obtained from E_C and the energy gap of the sub-band, ϕ is the electric potential, k is the Boltzmann constant, T is temperature, N_C is the effective density of states in the conduction band, and N_s is effective density of states in the s-th valence sub-band (heavy-hole, light-hole, or split-off hole one), that is,

$$N_C = 2m_{xx}^n (m_{zz}^n)^{1/2} \left(\frac{kT}{2\pi\hbar^2} \right)^{\mathbb{1}\,3/2}, \quad N_s = 2m_{xx}^s (m_{zz}^s)^{1/2} \left(\frac{kT}{2\pi\hbar^2} \right)^{\mathbb{1}\,3/2}, \tag{14.14}$$

where $\hbar$ is Planck's constant $(m_{xx}^n)^{-1}$, $(m_{zz}^n)^{-1}$, and $(m_{xx}^s)^{-1}$, $(m_{zz}^s)^{-1}$ are the components of the inverse effective mass tensor of electrons and holes in the s-th valence sub-band ($s = hh, lh$, or so) in CCS, respectively ($m_{xx}^n = m_{zz}^n = m_n$ and $m_{xx}^s = m_{zz}^s = m_s$ in cubic crystals within isotropic approximation). Integrals $\bar{\ }_n$ and $\bar{\ }_s$ in Equation 14.13 are

$$\bar{\ }_\nu(\zeta) = \int_0^\infty \frac{g_\nu(x)}{1 + \exp(x - \zeta)}, \quad (\nu = n, s), \tag{14.15}$$

where $g_\nu(x)$ is the density of states in the ν-th band, which is expressed as a function of carrier kinetic energy normalized by kT. For a simple parabolic band, $g_\nu(x) = x^{1/2}$ and the integral 14.15 is reduced to the standard Fermi integral of the 1/2-order. In a general case, Equation 14.15 enables accounting for the band nonparabolicity or a more complex energy spectrum typical for holes.

Electric potential distribution in the LED structure can be found from the Poisson equation solved in ECS. In wurtzite heterostructures of arbitrary orientation, the Poisson equation should account for anisotropic properties of the materials:

$$\kappa_0 \nabla \cdot (\hat{\kappa}^{\text{eff}} \nabla \phi) = \nabla \cdot \mathbf{P} - \rho, \quad \hat{\kappa}^{\text{eff}} = \hat{\kappa} - \frac{(\hat{e}\hat{S}\hat{e}^T)}{\kappa_0}, \tag{14.16}$$

where $\rho = q(N_D^+ + p - N_A^- - n)$ is the charge density, and N_D^+ and N_A^- are the concentrations of ionized donors and acceptors, respectively. The effective dielectric constant $\hat{\kappa}^{\text{eff}}$ is a diagonal matrix in CCS. To obtain its components in the ECS, the standard transformation should be applied. The second term in the expression for $\hat{\kappa}^{\text{eff}}$ originates from the elastic stress produced by the electric field (see Equation 14.4) and the back contribution of the stress to the electric polarization.

The dielectric constant of cubic crystals κ is a scalar as well as the effective dielectric constant $\kappa^{\text{eff}} = \kappa - S_{44}e_{14}^2/\kappa_0$. Therefore, the Poisson equation for the cubic heterostructures has the same form in both CCS and ECS. As a rule, $\kappa >> S_{44}e_{14}^2/\kappa_0$, which allows neglecting the stress contribution to the dielectric constants. In the case of wurtzite semiconductors, the stress contribution to κ_{zz}^{eff} of ~7%–8% is not negligible, changing remarkably the anisotropy degree of the dielectric constant. Variation of $\hat{\kappa}^{\text{eff}}$ with the *C*-axis inclination angle (Figure 14.3b) may exceed ~10% and should be accounted for in the analysis of nonpolar and semipolar LEDs; in the latter case, the nondiagonal $\kappa_{y'z'}^{\text{eff}}$ component of the effective dielectric constant should also be considered.

Concentrations of ionized donors and acceptors in Equation 14.16 depend locally on the quasi-Fermi level positions relative to the band edges, that is,

$$N_{\text{D}}^{+} = \frac{N_{\text{D}}}{1 + g_{\text{D}} \exp\left(\frac{F_n - E_{\text{C}} + E_{\text{D}} + q\phi}{kT}\right)}, \quad N_{\text{A}}^{-} = \frac{N_{\text{A}}}{1 + g_{\text{A}} \exp\left(\frac{E_{\text{V}} + E_{\text{A}} - F_p - q\phi}{kT}\right)}, \tag{14.17}$$

Here N_{D} and N_{A} are total donor and acceptor concentrations, respectively, E_{D} and E_{A} are the activation energies of the impurities, and g_{D} and g_{A} are their degeneracy factors. These equations assume free carriers and those trapped by donors and acceptors to be in thermodynamic equilibrium with each other.

Equations 14.11 through 14.17 form the simplest DDM for the carrier transport in LED structures. More elaborate models (Lundstrom, 2000; Jüngel, 2009; Jacoboni, 2010; Querlioz and Dollfus, 2010), both semiclassical and quantum mechanical, are not discussed here as most of the practically important results have been obtained with DDM allowing clear and straightforward interpretation.

Figure 14.4 presents band diagrams of single-quantum well (SQW) LED structures consisting of a thick n-GaN contact layer, an undoped 15 nm i-GaN interlayer, an undoped 3-nm $\text{In}_{0.17}\text{Ga}_{0.83}\text{N}$ SQW, a an 8-nm i-GaN spacer where Mg was assumed to diffuse to from the neighboring 20-nm p-$\text{Al}_{0.1}\text{Ga}_{0.9}\text{N}$ electron-blocking layer (EBL), and a 150 nm p-GaN contact layer. The above structures differ in crystal orientation only, demonstrating orientation-dependent band diagrams primarily controlled by the interface polarization charges. The Ga- and N-polar structures (Figure 14.4a and b) form QWs with strong built-in polarization fields and asymmetric barriers, whereas nonpolar (Figure 14.4c) and semipolar (Figure 14.4d) ones possess a rather weak electric field inside the well, and nearly symmetric barriers. The QW asymmetry varies with the forward bias applied and produces a strong, by tens of nanometers, bias-dependent blue shift of the emission wavelength of the Ga- and N-polar structures. In contrast, the emission wavelengths of the nonpolar and semipolar structures are predicted to be stable within ~3–5 nm under the bias/current variation. It is also interesting that GaN claddings of the InGaN SQW in N-polar structure form rather high natural barriers, preventing electrons and holes from leakage to the p- and n-layers of the structure, respectively (Figure 14.4b).

14.3.3 Recombination Models

Various recombination channels should be accounted for in LED simulations. The first one is the nonradiative Shockley–Read–Hall (SRH) recombination mediated by defects. The respective recombination rate is

$$\begin{gathered} R^{\text{SRH}} = \frac{np}{\tau_p n(1 + \xi_n) + \tau_n p(1 + \xi_p)} \cdot \left[1 - \exp\left(-\frac{F_n - F_p}{kT}\right.^{\mathbb{1}}\right]; \\ \xi_n = \exp\left(\frac{E_{\text{t}} - F_n}{kT}\right.^{\mathbb{1}}, \quad \xi_p = \exp\left(\frac{F_p - E_{\text{t}}}{kT}\right.^{\mathbb{1}}. \end{gathered} \tag{14.18}$$

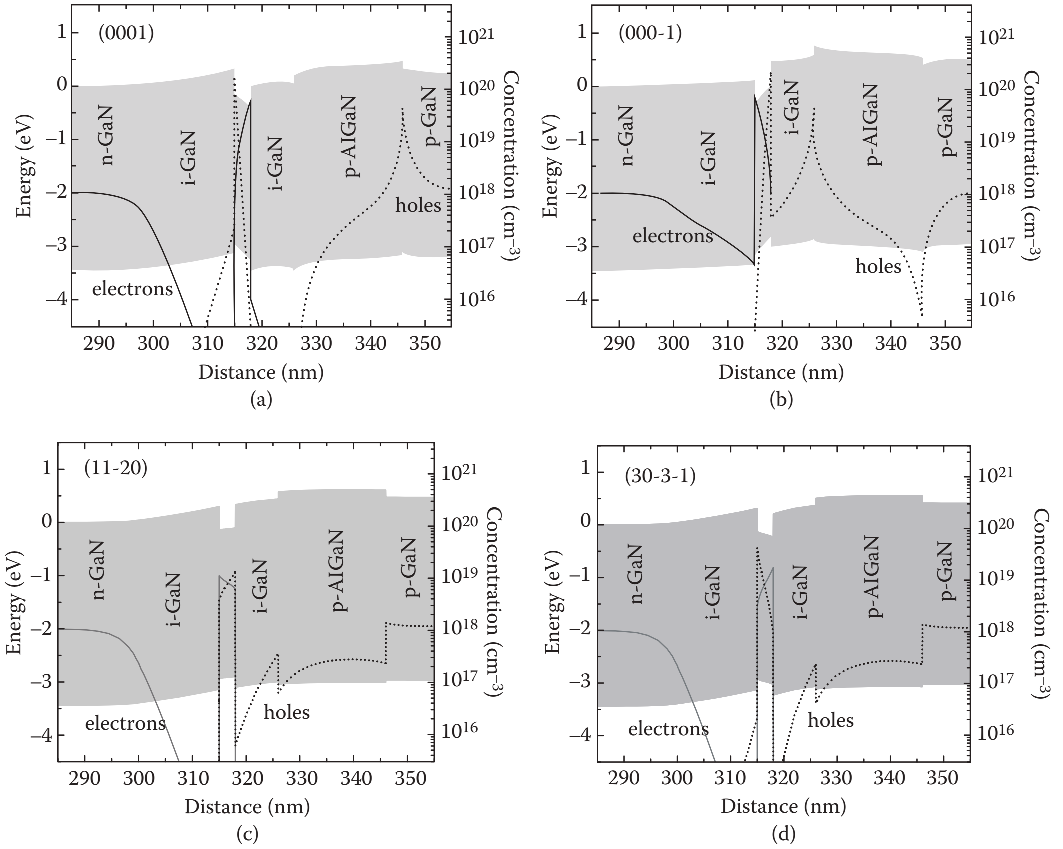


FIGURE 14.4 Band diagrams (gray shading indicates the bandgap) and distributions of electron (solid lines) and hole (dotted lines) concentrations in Ga-polar (a), N-polar (b), nonpolar (c), and semipolar (d) InGaN-based SQW light-emitting diode (LED) structures at the current density of 30 A/cm^2.

Here, E_t is the dominant energy level of a defect responsible for SRH recombination, τ_n and τ_p are the electron and hole life times. One can see from Equation 14.18 that ξ_n , ξ_p << 1, if the energy E_t is located deep in the materials bandgap.

In the case of recombination via point defects, $\tau_{n,p} = (c_{n,p}V_{n,p}N_\text{t})^{-1}$, where $c_{n,p}$ is the electron/hole trapping cross-section by the defects, $V_{n,p}$ is the carrier thermal velocity, and N_t is the defect concentration in the material. Dislocations threading through the LED active region are inherent defects in III-nitride epitaxial structures because of lack of native substrates. Here, the electron/hole life times can be estimated by the expression (Karpov and Makarov, 2002)

$$\tau_{n,p} = \frac{1}{4\pi D_{n,p}N_\text{d}}\left\{\ln\left(\frac{1}{\pi r_\text{c}^2 N_\text{d}}\right)^{\mathbb{1}} + \frac{2D_{n,p}}{r_\text{c}V_{n,p}} - \frac{3}{2}\right\}, \tag{14.19}$$

where $D_{n,p}$ is the carrier diffusivity, N_d is the dislocation density, and r_c is the dislocation core radius equal to the lattice constant a in the order of magnitude.

Another recombination channel is the radiative one. The rate of radiative recombination between free carriers in the conduction and valence bands can be calculated as

$$R^\text{rad} = Bnp \cdot \left[1 - \exp\left(-\frac{F_n - F_p}{kT}\right)^{\mathbb{1}}\right], \tag{14.20}$$

where the radiative recombination constant B depends essentially on the particular mechanism of light emission. In the case of bulk unstrained zinc-blende semiconductors, the B constant corresponding to optical transitions occurring with the **k**-selection rule at a rather low excitation level can be estimated as

$$B_0 = (N_{\mathrm{cv}}\tau_{\mathrm{B}})^{-1} \cdot \left(1 + \frac{3kT}{2E_{\mathrm{g}}}\right); \quad N_{\mathrm{cv}} = 2\left[\frac{(m_n + m_{\mathrm{hh}})kT}{2\pi\hbar^2}\right]^{3/2}, \quad \tau_{\mathrm{B}} = \frac{3\hbar m_0 c^2}{2\alpha \bar{n}_{\mathrm{r}} E_{\mathrm{g}} E_{\mathrm{P}}} \tag{14.21}$$

Here, m_n and m_{hh} are the effective masses of electrons and heavy holes, respectively, m_0 is the free electron mass, c is the light velocity in vacuum, $\alpha = q^2/(4\pi\varepsilon_0\hbar c) \approx 1/137$ is the fine structure constant, $\bar{n}_{\mathrm{r}}$ is the group refractive index at the wavelength corresponding to the optical transition, E_{g} is the bandgap of the semiconductor, and E_{P} is the energy related to the Kane's matrix element P (Kane, 1957): $E_{\mathrm{P}} = 2m_0P^2/\hbar^2$. At high excitation levels, the B constant starts to depend on the carrier concentrations. Numerical calculations of the radiative recombination rate under these conditions, accounting for the fact that $m_n << m_{\mathrm{hh}}$ have found the B constant to be well fitted by the expression: $B = B_0N_{\mathrm{cv}}/(N_{\mathrm{cv}} + p)$.

Many other processes may also contribute to radiative recombination, including excitonic optical transitions, band-to-impurity (donor, acceptor, or a deep center), donor-to-acceptor, phonon-assisted band-to-band recombination, and so on (Pankove, 1971), each providing its own specific radiative recombination constant. In addition, each of the above processes occurs differently in bulk materials, QWs, and quantum dots (QDs), multiplying a variety of B constants. Therefore, the choice of dominant mechanisms should be based on experimental data in every particular case and corresponding models should be applied in the framework of DDM.

The third important mechanism is nonradiative Auger recombination with the rate

$$R^{\mathrm{A}} = (C_nn + C_pp)\,np \cdot \left[1 - \exp\left(-\frac{F_n - F_p}{kT}\right)\right], \tag{14.22}$$

where C_n and C_p are the Auger recombination coefficients related to the microscopic processes involving two electrons and a hole and two holes and an electron, respectively. In bulk unstrained zinc-blende semiconductors, the C_n coefficient corresponding to the process involving two electrons and a heavy hole, that is, the CHCC[†] process, is given by the equation (Gelmont and Sokolova, 1982; Abakumov et al., 1991)

$$\begin{aligned} C_n &= \frac{16}{\sqrt{\pi}}\chi\frac{\nu_{\mathrm{D}}}{N_{\mathrm{C}}N_{\mathrm{hh}}}\left(\frac{kT}{E_{\mathrm{g}}}\right)^{3/2}\frac{\langle\varepsilon_{\mathrm{c}}\rangle}{E_{\mathrm{g}}}\exp(-E_{\mathrm{th}}/kT); \\ E_{\mathrm{th}} &= \frac{m_n}{m_{\mathrm{hh}}}E_{\mathrm{g}}\frac{(2E_{\mathrm{g}} + \Delta)(3E_{\mathrm{g}} + 2\Delta)}{(E_{\mathrm{g}} + \Delta)(3E_{\mathrm{g}} + \Delta)}, \quad \nu_{\mathrm{D}} = \frac{E_{\mathrm{D}}}{\hbar}, \\ \chi &= \left(\frac{E_{\mathrm{g}} + \Delta}{3E_{\mathrm{g}} + 2\Delta}\right)^{3/2}\left(\frac{3E_{\mathrm{g}} + \Delta}{2E_{\mathrm{g}} + \Delta}\right)^{1/2}, \quad E_{\mathrm{D}} = \frac{m_{\mathrm{n}}q^4}{32\pi^2(\kappa_0\kappa)^2\hbar^2}. \end{aligned} \tag{14.23}$$

[†] To distinguish between various microscopic Auger processes, the following four-letter notation is commonly accepted. First two letters indicate the initial and final electronic states of the recombining particle, e.g., an electron in the conduction band (C) and a hole in the heavy hole sub-band (H). Last two letters indicate the initial and final states of the third particle to which the excess energy released in course of recombination is transferred. Therefore, the CHCC notation means that an electron in the conduction band and a heavy hole recombine with each other, transferring the excess energy to one more electron (C), which is scattered within the same conduction band (C). In turn, the CHHS notation means that an electron in the conduction band and a heavy hole recombine, transferring the excess energy to another heavy hole (H), which is scattered into the split-off hole sub-band (S).

Here, N_C and N_{hh} are the effective densities of states in the conduction band and heavy-hole sub-band, respectively (see Equation 14.14), E_D is the ionization energy of a shallow donor, $\langle \varepsilon_c \rangle$ is the mean electron energy in the conduction band, and Δ is the spin–orbital splitting of the valence band. The energy $\langle \varepsilon_c \rangle = \frac{3}{2}kT$ in the case of nondegenerated electrons and $\langle \varepsilon_c \rangle = \frac{3}{5}\varepsilon_F \propto n^{2/3}$ for degenerated electrons, and ε_F is the Fermi energy. In the latter case, Auger recombination depends even more on the electron concentration than in the case of nondegenerated carriers.

The CHCC Auger process has a threshold accounted for by the exponential factor in Equation 14.23, with E_{th} being the threshold energy. The existence of the threshold originates from simultaneous conservation of the momentum and energy during recombination. As follows from Equation 14.23, the rate of the CHCC Auger recombination decreases dramatically with the material bandgap; therefore, this nonradiative process is especially important in narrow-bandgap semiconductors.

Among the processes involving two holes and an electron in bulk cubic semiconductors, the most important is the CHHS one, where the high-energy hole transfers from the heavy-hole to the split-off hole sub-band. At $E_g - \Delta >> kT$, the recombination coefficient associated with this process has an analytical representation (Abakumov et al., 1991):

$$C_p = \frac{216}{\sqrt{\pi}}\chi\frac{\nu_D}{N_{hh}N_{so}}\left(\frac{kT}{E_{th}}\right)^{1/2}\left(\frac{kT}{E_g}\right)^3\exp\left(\frac{-E_{th}}{kT}\right);$$
$$E_{th} = \frac{m_{so}}{m_{hh}}E_g\frac{E_g(E_g - \Delta)}{(E_g + \Delta)(3E_g - 2\Delta)},\quad \chi = \left(\frac{m_{hh}}{m_{so}}\right)^{1/2}\left(\frac{E_g + \Delta}{E_g}\right)^2. \tag{14.24}$$

Equation 14.24 shows that the energy threshold of the CHHS process disappears under resonance conditions $\Delta \approx E_g$; near the resonance, the threshold energy is relatively small, resulting in a considerable increase in the recombination rate.

Similarly to the case of radiative recombination, there is a variety of microscopic Auger recombination processes, involving, in addition to free carriers, those trapped by shallow donors, acceptors, and deep centers (Abakumov et al., 1991), or injected into QWs and QDs (Zegrya and Kharchenko, 1992; Dyakonov and Kachorovskii, 1994; Zegrya and Samosvat, 2007) as well as phonons (Kioupakis et al., 2011). In all these cases, the momentum conservation is violated in one or more directions, leading to the disappearance of the Auger recombination threshold and, as a result, to an increase in the nonradiative recombination rate. Unfortunately, not all of the mechanisms allow even simplified analytical treatment to derive the recombination coefficient; moreover, the theory for some important mechanisms is not yet completely developed (Abakumov et al., 1991). Therefore, only experimental observations may provide guidelines for the choice of dominant Auger mechanisms, in particular materials and device structures, and fitting of the C_n and C_p coefficient to the available data instead of using a rigorous theory may be applied in some cases.

The final nonradiative mechanism of carrier losses to be considered is that their surface recombination is occurs mainly at free surfaces of the LED active region. The surface recombination is accounted for via the flux J_s of carriers arriving at the free surface to recombine nonradiatively. The flux may be calculated as

$$J_s = \frac{np}{n/S_p + p/S_n}\cdot\left[1 - \exp\left(-\frac{F_n - F_p}{kT}\right)\right], \tag{14.25}$$

where S_n and S_p are the surface recombination velocities associated with n-type and p-type materials, respectively. Frequently the carrier losses localized at the active region surface are accounted for by the effective recombination rate distributed over the whole active region with the area and perimeter $\mathcal{P}$, which can be calculated by Equation 14.18 with $\xi_n = \xi_p = 0$ and $\tau_{n,p} = \ /(\mathcal{P}S_{n,p})$ (Royo et al., 2002). Such a simplified model is valid if the carrier concentration variation in the active region is much lower than their mean concentrations. In other cases, a more elaborate approach should be applied (see Section 14.4.2).

14.3.4 Efficiency of Light Emission

A number of parameters characterize the efficiency of LED structures. One is the injection efficiency η_{inj}, which accounts for the losses of minority carriers arriving at the contact electrodes without recombination inside the LED structure. The injection efficiency is defined as the ratio of the recombination current $I_{\text{rec}} = q \int_V R(\mathbf{r})\, d^3\mathbf{r}$ to the total current I flowing through the LED (here, the integration is performed over the whole LED volume V). Another important parameter is the internal quantum efficiency (IQE) η_i, defined as the ratio of the radiative recombination current $I_{\text{rad}} = q \int_V R^{\text{rad}}(\mathbf{r})\, d^3\mathbf{r}$ to I_{rec}. EQE η_{e} characterizes the number of photons outgoing from the LED per one electron–hole pair passing through the LED contact electrodes. Generally, $\eta_e = \eta_{\text{ext}}\eta_{\text{inj}}\eta_i$, where η_{ext} is the efficiency of light extraction (LEE) from the LED die. Finally, the wall–plug efficiency (WPE) η_{w} is the ratio of the output optical power to the electrical power consumed by LED during its operation. The ratio $\eta_{\text{w}}/\eta_{\text{e}}$ is sometimes regarded as the so-called electric efficiency. However, this ratio may be greater than unity at low currents and, therefore, use of this parameter appears to be somewhat misleading.

Both blue/green InGaN-based and red/amber AlGaInP-based LEDs exhibit rather strong EQE reduction at high operating currents but the dominant mechanisms responsible for the reduction are different in nitride and phosphide devices. In AlGaInP LEDs, the EQE reduction with current has been primarily attributed to electron leakage into the p-side of the LED structure because of insufficiently high barriers inherent in the AlGaInP materials system (Royo et al., 2002; Altieri et al., 2005). The leakage is found to be strongly increased by elevated temperature, with the activation energy nearly equal to the difference between the electron energy levels involved in optical transitions and the edge of the barrier conduction band (Altieri et al., 2005). This experimental fact was the major argument for attributing the EQE reduction with current to the electron leakage.

Blue/green nitride LEDs operate at the nonequilibrium carrier density in the active region, at least an order of magnitude higher than that in phosphide LEDs (compare Figures 14.4 and 14.5a). Therefore, the EQE reduction (droop) with current is associated here with Auger recombination (Shen et al., 2007; Bulashevich and Karpov, 2008; Laubsch et al., 2009); see also various reviews (Piprek, 2010; Avrutin et al., 2013; Verzellesi et al., 2013; Cho et al., 2013) for more detailed discussion on this and alternative mechanisms and experimental studies (Iveland et al., 2013; Galler et al., 2013; Binder et al., 2013) directly demonstrating the importance of Auger recombination. Although particular microscopic mechanisms of Auger recombination in nitride LEDs are not reliably identified, the concept of enlarged recombination volume used in the LED structure design (see, e.g., the review by Weisbuch et al. 2015) and aimed at suppression of the nonradiative carrier losses due to lowering the carrier concentration in the active region works well in practice.

Figure 14.5a shows the band diagrams and nonequilibrium carrier concentrations in a red ($\lambda = 622$ nm) LED structure grown on an n-GaAs substrate and consisting of an undoped 5 × (3.5 nm $\text{Ga}_{0.51}\text{In}_{0.49}\text{P}$/7 nm $(\text{Al}_{0.8}\text{Ga}_{0.2})_{0.51}\text{In}_{0.49}\text{P}$) multiple-quantum well (MQW) active region sandwiched between n- and p-$\text{Al}_{0.51}\text{In}_{0.49}\text{P}$ confinement layers, both 100 nm thick; a thick p-GaP contact layer completes the structure. As one can see from a comparison of Figures 14.5a and 14.4, the electron/hole concentration in the active region of the red LED is ~1–2 orders of magnitude lower than in InGaN-based LEDs at the same operating current density of 30 A/cm^2. This originates from different properties of the phosphide and nitride semiconductors: conduction and valence band offsets, carrier mobilities, and recombination constants. The band alignment in the MQW active region is rather flat, providing a uniform distribution of the recombination rate among all the QWs.

The injection efficiency of the red LED structure plotted in Figure 14.5b is strongly dependent on temperature. It differs from unity rather weakly at 300 K, and is considerably reduced at 340 and 370 K, indicating increasing carrier losses via leakage. The electron leakage is predicted to dominate over the hole one at all current densities considered. IQE of the red LED structure varies weakly with temperature (Figure 14.5c) due to temperature-dependent recombination constants. The IQE dependence on current

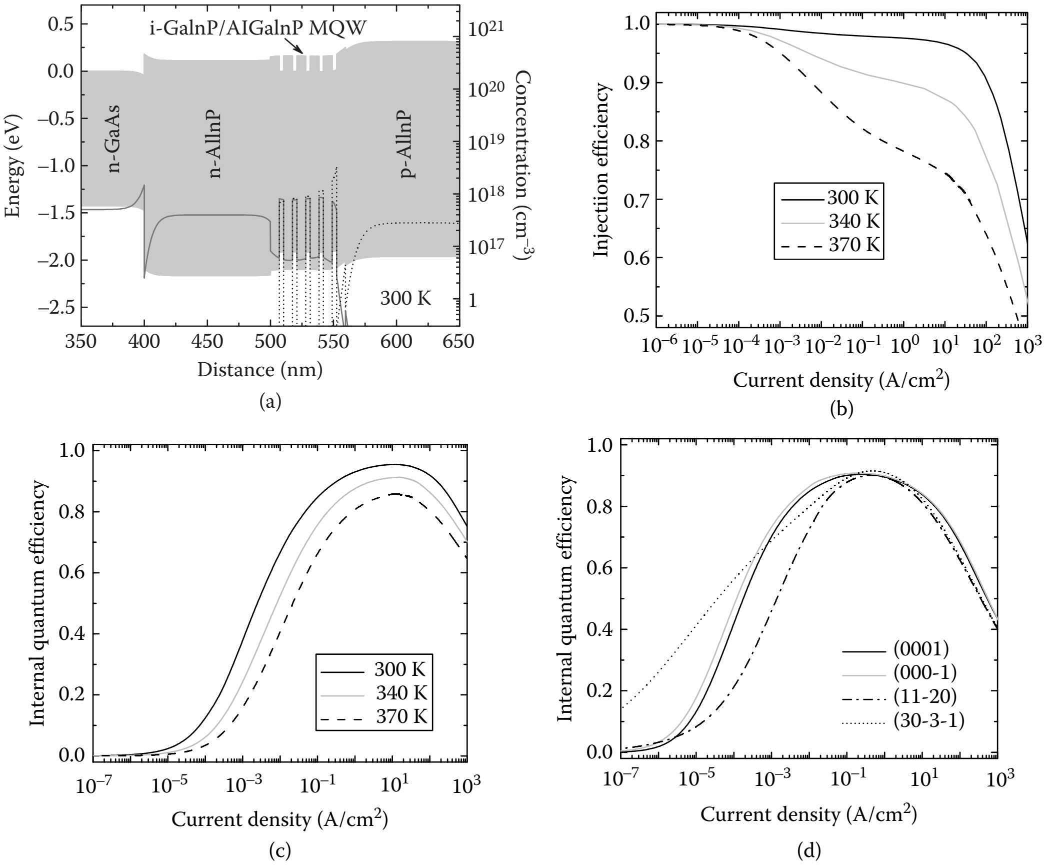


FIGURE 14.5 Band diagram (gray shading indicates the bandgap) and distributions of electron (solid lines) and hole (dotted lines) concentrations in a red MQW light-emitting diodes (LEDs) structure at the current density of 30 A/cm^2 (a). Injection efficiency (b) and IQE (c) of the structure as a function of current density at three different temperatures. IQEs of the InGaN-based SQW LED structures shown in Figure 14.4 are also given versus current density for comparison (d).

has a dome-like shape, reflecting competition between the radiative and nonradiative recombination channels. In particular, the IQE droop at high currents is related to nonradiative carrier losses caused by Auger recombination. Because IQE starts to decline at much higher current density than IE does, degradation of the overall red LED efficiency with temperature should be primarily attributed to the carrier leakage.

Figure 14.5d displays current-dependent IQEs of the nitride SQW LED structures discussed in Section 14.3.2. Simulation of the structures was intentionally carried out with the same recombination constants in order to emphasize the role of the crystal polarity. The computed injection efficiency is equal to unity for all the structures in the whole range of the current density variation. Therefore, the IQE droop clearly seen in Figure 14.5d should be attributed entirely to Auger recombination. As the carrier concentrations in the active region of InGaN-based LEDs considerably exceed those in the red LED structure, the IQE droop in nitride emitters starts at lower current densities (compare Figure 14.5c and d).

The efficiency droop is also observed in UV LEDs with AlGaN active regions (Mickevičius et al., 2015). The nature of the droop here, however, remains unclear and requires further investigation.

14.4 Current Spreading, Heat Transfer, and Light Extraction in LED Dice

State-of-the-art LEDs utilize rather complex essentially 3D chip designs aimed at achieving various goals: providing a uniform carrier injection into the active regions at minimum series resistance, efficient light extraction from the LED dice, and efficient heat removal from the active region. Most of the physical processes involved in the chip operation are coupled with one another so that their straightforward modeling requires a large amount of computational time and computer resources. Therefore, implementation of approximation approaches to create engineering simulations of the processes occurring in the LED chips is highly desirable.

A rather effective approach can be developed on the basis of the equation for the total current density $\mathbf{j} = \mathbf{j_n} + \mathbf{j_p}$; under steady-state conditions, this equation follows from Equation 14.11: $\nabla \cdot \mathbf{j} = 0$. It does not contain any recombination terms and enables much easier and faster numerical solution than full DDM equations, if the relationship between **j** and electric potential ϕ is specified. In the contact layers, this relationship can be approximated by a unipolar conductivity, that is, $\mathbf{j} = -\sigma_{n,p}\nabla\phi$, where σ_n and σ_p are the electric conductivities of the n- and p-contact layers, respectively. This approximation does not work, however, in the LED active region. On the other hand, electrons and holes are transferred here mainly in the vertical direction due to dominant vertical components of the electric field and gradients of carrier concentrations. Therefore, the carrier transport in the active region may be well approximated by a 1D model. The 1D solution to the DDM equations provides a nonlinear dependence of the vertical current density j_v on the p-n junction bias $U_b = F_n - F_p$, where the values of the quasi-Fermi levels F_n and F_p are taken at the n- and p-sides of the active region, respectively. Thus, the $j_v(U_b)$ dependence obtained may be used as a nonlinear boundary condition in 3D chip simulations, connecting the vertical component of the local current density **j** with the electric potential difference on the n-side (ϕ_n) and p-side (ϕ_p) borders of the active region: $\mathbf{n} \cdot \mathbf{j} = j_v(\phi_p - \phi_n)$, where **n** is the normal to the active region directed from the p- to n-contact layer. Solution to the above problem provides 3D distribution of the current density in an LED die.

This approximate approach has proved to be quite efficient for coupled modeling of LEDs with rather complex chip designs (Bogdanov et al., 2008); it can also be easily refined by using additional experimental input (López and Margalith, 2008). Comparison of theoretical results obtained by the above approach with available observations has demonstrated its good predictability (Chernyakov et al., 2013).

As it was already mentioned, the carrier transport in the LED active region occurs primarily in the vertical direction. Therefore, considering the lateral electron and hole transfer as a perturbation, one can also account for the lateral redistribution of carriers due to ambipolar diffusion and their surface recombination at the free surface of the active region.

To illustrate, Figure 14.6 shows simulated 2D current density distribution over the active region and some characteristics of a vertical red LED die shaped in such a way as to increase its light extraction efficiency (LEE) using inclined side walls of the die (see Figure 14.7a). The LED structure presented in Figure 14.5a has been chosen for the simulations, whereas modeling of current spreading, heat transfer, and light extraction was carried out self-consistently in 3D approximation. Due to the vertical chip geometry, current crowding (see Section 14.4.1 for a more detailed discussion) is rather weak in the chip (Figure 14.6a), resulting in a nearly linear relationship between the mean current density and current (Figure 14.6b). Simulated current–voltage and light–current characteristics of the red LED are shown in Figure 14.6c, whereas EQE and WPE variation with current is given in Figure 14.6d. At relatively low currents, WPE exceeds EQE (see Figure 14.6d), which is typical for LEDs operating at the forward voltage $V_f < E_{ph}/q$, where E_{ph} is the mean energy of emitted photons. The decrease of EQE and WPE at high currents is caused by the thermally enhanced electron leakage, as discussed in Section 14.3.4. One can see from Figure 14.6b that the mean temperature in the LED die increases superlinearly with the operating current.

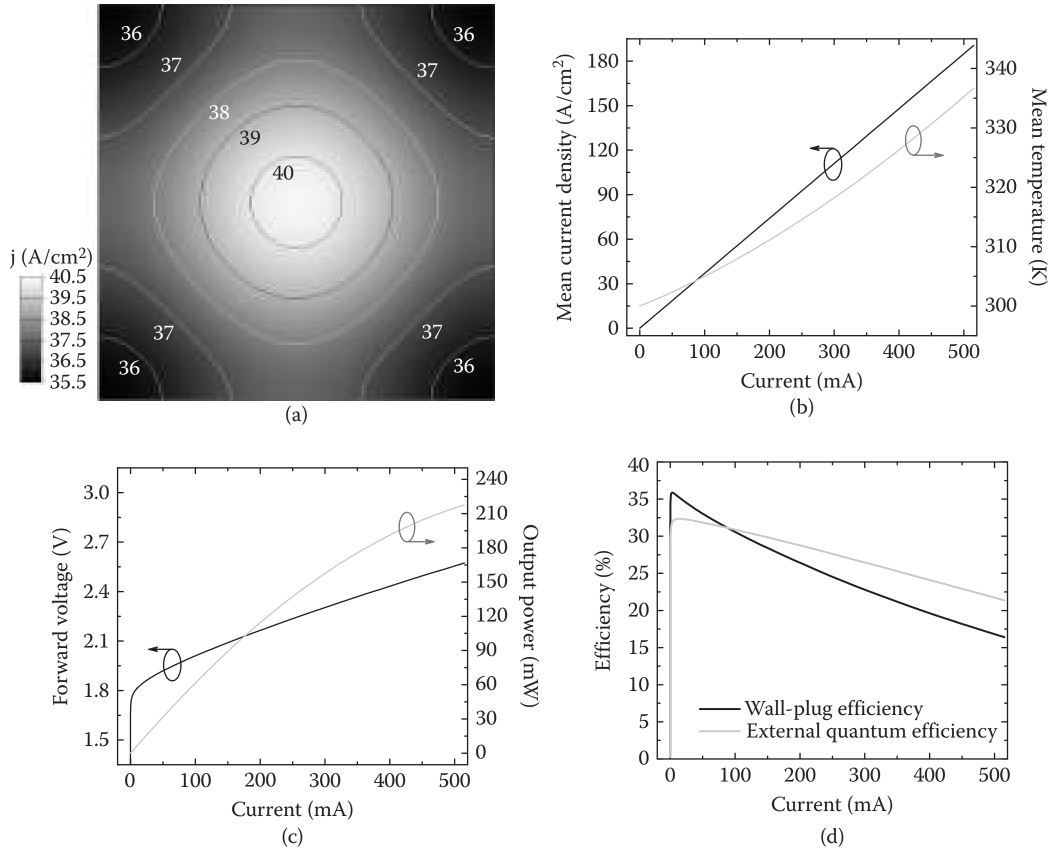


FIGURE 14.6 Two-dimensional (2D) vertical current density distribution in the active region at the current of 100 mA (a) and mean current density and temperature (b), forward voltage and output optical power (c), and external quantum efficiency (EQE) and wall–plug efficiency (WPE) (d) as a function of operating current of the red light-emitting diode (LED) chip shown in Figure 14.7a.

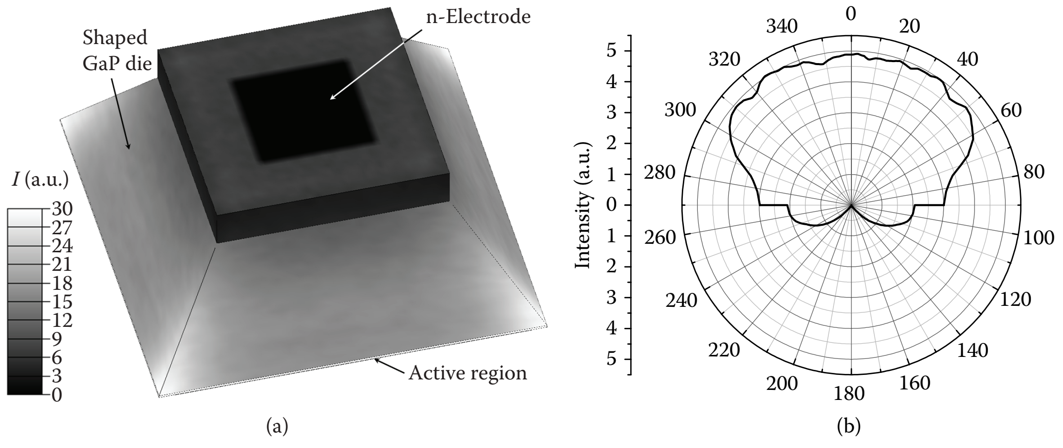


FIGURE 14.7 Near-field emission intensity distribution over the outer surfaces of red light-emitting diode (LED) die with shaped side walls (a) and far-field emission pattern of the LED (b).

As soon as the distribution of electron and hole concentrations in the active region is determined, the radiative recombination rate in each point can be obtained (Bogdanov et al., 2008). Using the radiation pattern specific for the active region band structure, one can use the ray tracing to simulate photon propagation and absorption in the LED die, as well as their extraction efficiency.

Figure 14.7 shows results of the ray-tracing modeling of the shaped red LED die considered previously. LEE has been found to be independent of operating current due to weak current crowding and equal to 35.2%. Near-field emission intensity distribution over the outer surfaces of the LED die (Figure 14.7a) demonstrates that most of the emitted photons are extracted through the inclined side walls of the die and primarily through the regions adjacent to the die edges. As a result, the far-field emission pattern differs substantially from the Lambertian one, providing nearly constant intensity in a wide observation angle of ~100° (Figure 14.7b).

Ray tracing is quite suitable for accounting properties of the photon ensemble in LED dice. However, its applicability is rather limited as the approach ignores the wave nature of photons, that is, the effects of their diffraction and interference. Ray tracing is warranted if specific LED chip dimensions are much larger than the photon wavelength. However, any elements of the chip design based on the photon interference or diffraction, such as photonic crystals or textured surface commonly used for better extraction of photons from LEDs, cannot be considered to be accurate in the ray tracing procedure.

An alternative approach frequently applied for estimation of LEE is the direct finite-difference time domain (FDTD) simulations, where the Maxwell's equations are solved for the electromagnetic field induced by a dipole imitating the act of radiative recombination of an electron and a hole. The FDTD method is quite suitable, in particular, for analysis of light diffraction at a textured surface of the LED dice. Despite the accounting for photon wave nature, the FDTD approach has also some limitations related to (1) ignoring the ensemble properties of electrons/holes and, hence, of emitted photons and (2) using some artificial boundary conditions for the Maxwell's equations, the impact of which on the accuracy of theoretical predictions is still uncertain (Zhmakin, 2011).

To summarize, ray tracing and FDTD simulation are complementary approaches to LEE calculation rather than competing ones. The choice between them depends on the LED chip geometry and parameters desirable to predict LEE value, emission pattern in the near- and far-field zones, light absorption in LED die, emission polarization, and so on.

14.4.1 Current Crowding

Current crowding is the effect of the current density localization inside the LED die around the streamlines with minimum path resistance. Generally, every LED chip can be represented via two distributed linear resistors corresponding to n- and p-contact layers and a nonlinear resistor corresponding to the LED active region, including p-n junction, all connected in series. The nonlinear p-n junction resistor has a nearly exponential resistance dependence on the internal p-n junction bias. At a low operating current, the p-n junction resistance is much higher than both linear ones and the lateral current flow in the contact layers dominates, resulting in a rather uniform distribution of the vertical current density over the LED active region. At higher currents, the increase in the p-n junction bias leads to a dramatic reduction of the p-n junction resistance, which becomes comparable to or lower than both contact layer resistances. As a result, the vertical current density redistributes significantly in such a way as to minimize the total path resistance along the current streamlines. Therefore, current crowding is essentially a nonlinear effect, which develops with the LED operating current I and depends on temperature primarily through the nonlinear p-n junction resistance.

Because of current crowding, the main LED characteristics should be obtained by averaging over the active region area with the local current density j_V serving as the weight function. So, the mean current density in the active region $\bar{j} = I^{-1} \int j_V^2(\mathbf{r})\, d^2\mathbf{r} > I/$ is no longer proportional to $I = \int j_V(\mathbf{r})\, d^2\mathbf{r}$, if the current crowding is pronounced. Also the mean IQE of an LED becomes dependent on current crowding (see simulation results obtained by Bogdanov et al. (2010) and experimental data by Malyutenko et al.

(2010)): $\overline{\eta_i} = I^{-1} \int \eta_i(j_v) \cdot j_v(\mathbf{r})\, d^2\mathbf{r}$, where $\eta_i(j_v)$ is the IQE of LED structure dependent on the local current density j_v.

Another important effect of current crowding found by simulations (Bogdanov et al., 2010) and reported experimentally by Laubsch et al. (2010) is the current localization under metallic electrodes deposited on the emitting surface of LEDs. Depending largely on the operating current, the localization results in LEE decreasing with current, enhancing the overall efficiency droop. The effect was found to be much stronger than the IQE reduction caused by current crowding (Bogdanov et al., 2010). Avoiding metallic electrodes on the emitting surface by the use of advance chip designs (see, for instance, Laubsch et al. 2010) improves LEE substantially.

Generally, current crowding is especially pronounced in lateral LED chips utilizing nonconductive sapphire substrate and having one-side electrode access to both n- and p-contact layers. This is because of a relatively thin n-contact layer providing dominant lateral current flow inside the chip. In a vertical LED die similar to that shown in Figure 14.7a, the current crowding is weaker but, nevertheless, still tangible (see Figure 14.6a) despite a relatively large distance, ~200 μm, between the n- and p-electrodes, which looks at first glance to be quite sufficient for the complete homogenization of the current density distribution inside the die.

14.4.2 Surface Recombination

There are three main parameters affecting the surface recombination impact on the LED efficiency: the recombination velocity S, the ambipolar carrier diffusion coefficient D_a in the active region, and the carrier differential life time τ_d generally dependent on the carrier concentration in the active region. The latter two parameters determine the carrier diffusion length $L_d = (D_a \tau_d)^{1/2}$, which characterizes the width of the region adjacent to the free surface of the active region where contribution of surface recombination to the carrier losses is significant.

Red/amber AlGaInP LEDs operate normally at a reduced, ~10^{17}–10^{18} cm^{-3}, carrier concentration in the active region (Figure 14.5a), providing a relatively long τ_d. The carrier diffusivity in the active region enhanced in undoped QWs may be rather large. As a result, L_d may approach ~10–25 μm in such devices. Therefore, surface recombination results in a remarkable rise to the carrier losses already at the LED chip dimensions as small as ~150–250 μm (Royo et al., 2002).

The carrier concentration in the active regions of blue and green InGaN-based LEDs is ~1–2 orders of magnitude higher (Figure 14.4), shortening significantly the lifetime τ_d; the carrier diffusivity here is also about an order of magnitude smaller. Therefore, L_d does not normally exceed a few micrometers in nitride LEDs. On the other hand, the free active region surface is normally located close to the regions of strong current crowding in the III-nitride LED chips. Estimates show that surface recombination may be valuable in blue/green LEDs at low currents corresponding to the maximum of their efficiency. Surface recombination is expected to be especially pronounced in deep-UV LEDs because of (1) a high Al content in the active region normally increasing the surface recombination velocity, (2) a weaker carrier localization resulting in their higher diffusivity in the active region, and (3) smaller LED die dimensions usually utilized in practice.

14.5 Phosphor-Converted White Light Emission

An important sector of LED industry is the production of white light sources for solid-state lighting. Such sources are based on emulation of sunlight by a number of LEDs emitting in different spectral ranges or by using a partial light conversion by one or more phosphors. As it has been shown by Bulashevich et al. (2015), the phosphor-converted LEDs (PC-LEDs) provide a higher efficacy at a better color rendition, as compared to white light sources utilizing LEDs only. In addition, fabrication of PC-LEDs represents the main industrial approach to the white light production. Therefore, this section will focus on simulation models necessary for the analysis of light conversion by phosphors pumped with LEDs.

14.5.1 White Light Characteristics

The fundamentals of color perception and white light characterization are described in detail by Judd and Wyszecki (1975). Here, the main white light characteristics will be briefly discussed in view of discussion on their further modeling.

Starting with J.C. Maxwell's studies in the mid-nineteenth century, it became customary to decompose the light color into three components. According to CIE (1986), three color coordinates, X, Y, and Z, can be brought in by the convolution of a light source spectral power density emission spectrum $S_\lambda(\lambda)$ with three tabulated color matching functions, $\bar{x}(\lambda)$, $\bar{y}(\lambda)$, and $\bar{z}(\lambda)$. Then the relative chromatic coordinates $x = X/(X + Y + Z)$ and $y = Y/(X + Y + Z)$ project the emission spectrum into the CIE1931 color chart, which is represented by points in the (x, y) space, all the possible colors that can be encountered in practice.

A natural source producing white light is a black body heated up to a certain temperature. Under the temperature variation, chromatic coordinates of the black-body radiation (BBR) form in the CIE1931 diagram, a locus nearly corresponding to white light. In practice, some deviations of the white-light chromatic coordinates x_w and y_w from the BBR locus are admissible. In this case, a certain projection of the (x_w, y_w) point to the BBR locus produces the so-called correlated color temperature (CCT), which is one of the most important characteristics of white light.

Another important characteristic is the luminous efficacy of radiation (LER) obtained by convolution of the spectral power density spectrum $S_\lambda(\lambda)$ with the human eye sensitivity function $\Phi(\lambda)$ (Judd and Wyszecki, 1975), accounting statistically for the perception of human eye: $\mathrm{LER} = \int S_\lambda(\lambda)\,\Phi(\lambda)\,d\lambda$. The efficacy of a white light source is the product of LER and total WPE of the source, which depends on WPEs of individual LEDs and the power fractions of the emission spectra supplied by these LEDs into the total spectrum $S_\lambda(\lambda)$ (Bulashevich et al., 2015). If a phosphor emission is used for white light generation, its effective WPE depends on the WPE of the pumping LED, the Stokes shift caused by down conversion of the photon energy, and quantum yield of the phosphor emission (Bulashevich et al., 2015).

Ability of a light source to reproduce the colors of illuminated objects accurately is characterized by the color rendering index (CRI). The CRI accounts for the differences between the chromatic coordinates of light reflected from a number of standard color samples at their illuminating by the light source studied and by one or another standard illuminant. Detailed procedure of the CRI evaluation is recommended by CIE (1986). The CRI is reduced from the maximum value of 100 by every chromatic coordinate difference obtained from the selected color samples.

14.5.2 Light Conversion by Phosphor

Conversion medium in an LED lamp typically consists of a silicone layer with embedded particles of one or more phosphors deposited on the surface of pumping LED. Therefore, the light absorption and scattering by the phosphor particle, as well as emission of the down-converted light and its scattering by the particles, occurs in the medium. Here, spectral ray tracing is the approach most suitable for simulations of light conversion by phosphors. Assuming the phosphor particles to be randomly and uniformly distributed in silicone, one can derive the light conversion to depend on scattering properties of an individual particle and total volume fraction f_{ph} of the particles in the conversion medium.

Rigorous Mie theory (Van de Hulst, 1957) is currently the best model for analysis of light scattering by individual phosphor particles. Assuming spherical particle shape, the theory predicts the wavelength-dependent scattering and absorption cross-sections necessary for ray tracing simulations, as well as the scattering pattern in the far-field zone. For particles of realistic size and monochromatic light, the scattering patterns exhibit strong interference oscillations (black lines in Figure 14.8), which should be absent in

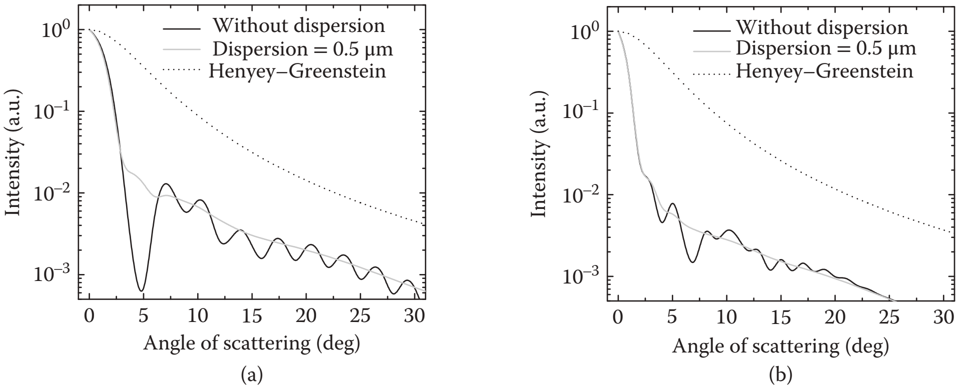


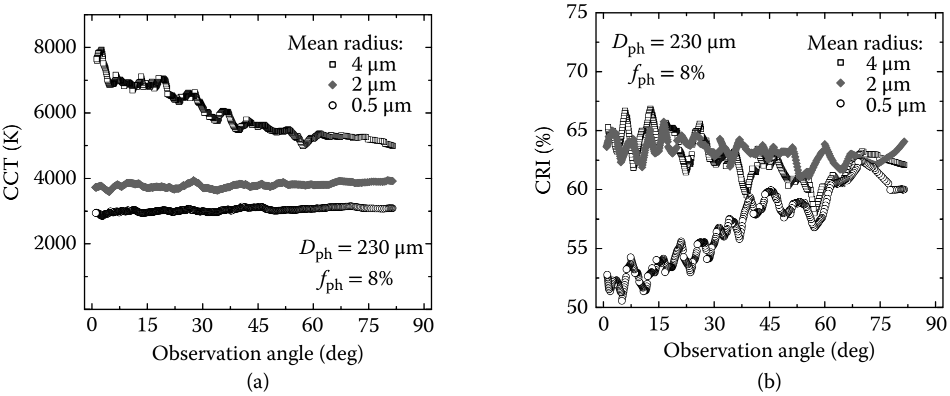
FIGURE 14.8 Scattering patterns of YAG:Ce^{3+} particle with the radius of 4 μm obtained by the Mie theory with and without account of particle size dispersion and by Henyey–Greenstein approximation for the wavelength of 600 nm (a) and 400 nm (b).

practice in view of (1) dispersion of the particle sizes, (2) deviation of the particle shapes from the spherical one, and (3) a nonmonochromatic spectra of the pumping LED and phosphor. Accounting for the size dispersion of phosphor particles is already sufficient to avoid the oscillations in the scattering pattern (gray lines in Figure 14.8). It is interesting that the Henyey–Greenstein scattering functions (Henyey and Greenstein, 1941) commonly used for empirical fitting of the scattering pattern do not work well in the case of realistic phosphor particles, if the asymmetry parameter *g* of the function is correctly estimated as the first moment of the scattering pattern predicted by the Mie theory (dotted lines in Figure 14.8). Being fitted to the computed pattern, the Henyey–Greenstein scattering function at g ~0.98–0.99 is capable of predicting the enhanced forward scattering seen in Figure 14.8 but fails to reproduce the wings of the scattering pattern predicted by the Mie theory. So, there is an unlikely alternative to the use of Mie theory for modeling the light interaction with the phosphor particles.

The absorption/scattering cross-sections and scattering pattern of individual phosphor particles involved in spectral ray tracing allow simulating the total emission spectrum in the far-field zone, estimating the color characteristics of white light, that is, chromatic coordinates *x* and *y*, CCT, and CRI, and their dependence on the observation angle. The latter dependence is quite important as the overall white light LED performance requires uniform color characteristics in a wide range of the observation of angle variation. As soon as the angle dependence of the total emission spectrum is obtained, integration over the angle enables calculation of LER and then of efficacy of the white light source.

One of the approaches aimed at improving angle uniformity of color characteristics is based on the fact that the scattering cross-section of phosphor particles depends very critically on the particle size. Figure 14.9 shows the simulated angle dependence of CCT and CRI of the white light source comprising of a blue (450 nm) LED and a yellow YAG:Ce^{3+} phosphor with the emission spectrum peaked at 560 nm and its full width at half maximum of 120 nm. Simulations were carried out for a square-shaped LED chip uniformly coated by a 230-nm silicone layer with 8% of its volume filled by the phosphor particles. One can see from Figure 14.9 that the angle uniformity of CCT and CRI is strongly dependent on the particle size. Among the chosen values of the mean particle radius, only that of 2 μm provides simultaneously the angle-uniform CCT and CRI.

The above results have been obtained by ray tracing performed in 50 spectral intervals from 380 to 730 nm, using 10^8 rays. Even at that, the simulated color characteristics in the far-field zone are rather noisy because of insufficient ray statistics. Therefore, parallel computing is normally required for the spectral ray tracing in order to accelerate simulations.



(a)

(b)

FIGURE 14.9 Angle-dependent correlated color temperature (CCT) (a) and color rendering index (CRI) (b) of a white light source consisting of a blue light-emitting diode (LED) and a YAG:Ce^{3+} yellow phosphor with the particles of various sizes. The conversion media thickness D_{ph} = 230 μm is assumed in the ray-tracing simulations.

14.6 Concluding Remarks

This chapter presents a general frame and a strategy for multiscale and multidisciplinary simulations of state-of-the-art LEDs from heterostructure to lamp. The basic models and the most important mechanisms underlying the simulations are reviewed with the focus on those that were not discussed in detail previously. It should be noted that many aspects of simulations are just outlined in the chapter and should be adopted to specific subjects of the study. In particular, the radiative and nonradiative recombination mechanisms require more detailed description that accounts for both direct and phonon-assisted processes, as well as the variety of active region designs (quasi-bulk, QWs, or QDs). Another issue beyond the scope of the chapter is the proper choice of materials properties, which are known to often be highly inaccurate, and its impact on the simulation results. Finally, the role of a number of material factors important for LED operation, such as high-density threading dislocations and compositional fluctuations in InGaN and other alloys, are still in the early stages of study. A better understanding of these factors may result in a reconsideration of the basic carrier transport and recombination models in the future.

References

Abakumov VN, Perel VI, Yassievich IN (1991) Nonradiative recombination in semiconductors (Ch. 11). In: *Modern Problems in Condensed Matter Science*, V. M. Agranovich and A. A. Maradudin (Eds.), vol. 33. Amsterdam: North-Holland.

Altieri P, Jaeger A, Windisch R, Linder N, Stauss P, Oberschmid R, Streubel K (2005) Internal quantum efficiency of high-brightness AlGaInP light-emitting devices. *J. Appl. Phys.* 98: 086101.

Avrutin V, Hafiz S-D-A, Zhang F, Özgür Ü, Morkoç H (2013) InGaN light-emitting diodes: Efficiency-limiting processes at high injection. *J. Vac. Sci. Technol. A* 31: 050809.

Binder M, Nirschl A, Zeisel R, Hager T, Lugauer H-J, Sabathil M, Bougeard D, Wagner J, Galler B (2013) Identification of nnp and npp Auger recombination as significant contributor to the efficiency droop in (GaIn)N quantum wells by visualization of hot carriers in photoluminescence. *Appl. Phys. Lett.* 103: 071108.

Bogdanov MV, Bulashevich KA, Evstratov IY, Zhmakin AI, Karpov SY (2008) Coupled modeling of current spreading, thermal effects, and light extraction in III-nitride light-emitting diodes. *Semicond. Sci. Technol.* 23: 125023.

Bogdanov MV, Bulashevich KA, Khokhlev OV, Evstratov IY, Ramm MS, Karpov SY (2010) Current crowding effect on light extraction efficiency of thin-film LEDs. *Phys. Status Solidi C* 7: 2124–2126.

Bulashevich KA, Karpov SY (2008) Is Auger recombination responsible for the efficiency rollover in III-nitride light-emitting diodes? *Phys. Status Solidi C* 5: 2066–2069.

Bulashevich KA, Kulik AV, Karpov SY (2015) Optimal ways of colour mixing for high-quality white-light LED sources. *Phys. Status Solidi A* 212: 914–919.

Chernyakov AE, Bulashevich KA, Karpov SY, Zakgeim AL (2013) Experimental and theoretical study of electrical, thermal, and optical characteristics of InGaN/GaN high-power flip-chip LEDs. *Phys. Status Solidi A* 210: 466–469.

Cho J, Schubert EF, Kim JK (2013) Efficiency droop in light-emitting diodes: Challenges and countermeasures. *Laser Photonics Rev.* 7: 408–421.

CIE Publ. No. 15.2 (1986) Colorimetry. *Official Recommendations of the International Commission on Illumination*. Vienna: Central Bureau of the CIE.

Dyakonov MI, Kachorovskii VY (1994) Nonthreshold Auger recombination in quantum wells. *Phys. Rev. B*, 49: 17130–17138.

Galler B, Lugauer H-J, Binder M, Hollweck R, Folwill Y, Nirschl A, Gomez-Iglesias A, Hahn B, Wagner J, Sabathil M (2013) Experimental determination of the dominant type of Auger recombination in InGaN quantum wells. *Appl. Phys. Express* 6: 112101.

Gelmont B, Sokolova Z (1982) Auger recombination in direct-gap N-type semiconductors. *Sov. Phys. Semicond.* 16: 1067–1068.

Gil B (2009) Wurtzitic semiconductors heterostructures grown on (hk.l) oriented substrates: The interplay between spontaneous and piezoelectric polarization fields, elastic energy and the modification of quantum confined Stark effect. *Proc. SPIE* 7216: 7216OE.

Henyey LG, Greenstein JL (1941) Diffuse radiation in the galaxy. *Astrophys. J.* 93: 70–83.

Iveland J, Martinelli L, Peretti J, Speck JS, Weisbuch C (2013) Direct measurement of Auger electrons emitted from a semiconductor light-emitting diode under electrical injection: Identification of the dominant mechanism for efficiency droop. *Phys. Rev. Lett.* 110: 177406.

Jacoboni C (2010) *Theory of Electron Transport in Semiconductors*. Berlin: Springer.

Judd DB, Wyszecki G (1975) *Color in Business, Science and Industry*. New York, NY: Wiley.

Jüngel A (2009) *Transport Equations for Semiconductors*. Berlin: Springer.

Kane EO (1957) Band structure of indium antimonide. *J. Phys. Chem Solids* 1: 249–261.

Karpov SY, Makarov YN (2002) Dislocation effect on light emission efficiency in gallium nitride. *Appl. Phys. Lett.* 81: 4721–4723.

Kioupakis E, Rinke P, Delaney KT, Van de Walle CG (2011) Indirect Auger recombination as a cause of efficiency droop in nitride light-emitting diodes. *Appl. Phys. Lett.* 98: 161107.

Laubsch A, Sabathil M, Bergbauer W, Strassburg M, Lugauer H, Peter M, Lutgen S, Linder N, Streubel K, Hader J, Moloney JV, Pasenow B, Koch SW (2009) On the origin of IQE-“droop” in InGaN LEDs. *Phys. Status Solidi C* 6: S913–S916.

Laubsch A, Sabathil M, Baur J, Peter M, Hahn B (2010) High-power and high-efficiency InGaN-based light emitters. *IEEE Trans. Electron Devices* 57: 79–87.

López T, Margalith T (2008) Electro-thermal modelling of high power light emitting diodes based on experimental device characterisation. www.comsol.de/paper/5004.

Lundstrom M (2000) *Fundamentals of Carrier Transport*. Cambridge: Cambridge University Press.

Malyutenko VK, Bolgov SS, Podoltsev AD (2010) Current crowding effect on the ideality factor and efficiency droop in blue lateral InGaN/GaN light emitting diodes. *Appl. Phys. Lett.* 97: 251110.

Mickevičius J, Tamulaitis G, Jurkevičius J, Shur MS, Shatalov M, Yang J, Gaska R (2015) Efficiency droop and carrier transport in AlGaN epilayers and heterostructures. *Phys. Status Solidi B* 252: 961–964.

Nye JF (1964) *Physical Properties of Crystals: Their Representation by Tensors and Matrices* (Ch. 10). Oxford: Clarendon Press.

Pankove JI (1971) *Optical Processes in Semiconductors*. New York, NY: Dover Publications.

Park S-H, Chuang S-L (1999) Crystal-orientation effects on the piezoelectric field and electronic properties of strained wurtzite semiconductors. *Phys. Rev. B* 59: 4725–4737.

Piprek J (2010) Efficiency droop in nitride-based light-emitting diodes. *Phys. Status Solidi A* 207: 2217–2225.

Querlioz D, Dollfus P (2010) *The Wigner Monte Carlo Method for Nanoelectronic Devices*. London: ISTE Ltd. & John Wiley & Sons.

Romanov AE, Baker TJ, Nakamura S, Speck JS (2006) Strain-induced polarization in wurtzite III-nitride semipolar layers. *J. Appl. Phys.* 100: 023522.

Royo P, Stanley RP, Ilegems M, Streubel K, Gulden KH (2002) Experimental determination of the internal quantum efficiency of AlGaInP microcavity light-emitting diodes *J. Appl. Phys.* 91: 2563–2568.

Shen YC, Müller GO, Watanabe S, Gardner NF, Munkholm A, Krames MR (2007) Auger recombination in InGaN measured by photoluminescence. *Appl. Phys. Lett.* 91: 141101.

Van de Hulst HC (1957) *Light Scattering by Small Particles*. New York, NY: Dover Publications.

Van Roosbroeck W (1950) Theory of flow of electrons and holes in germanium and other semiconductors. *Bell System Tech. J.* 29: 560–607.

Verzellesi G, Saguatti D, Meneghini M, Bertazzi F, Goano M, Meneghesso G, Zanoni E (2013) Efficiency droop in InGaN/GaN blue light-emitting diodes: Physical mechanisms and remedies. *J. Appl. Phys.* 114: 071101.

Weisbuch C, Piccardo M, Martinelli L, Iveland J, Peretti J, Speck JS (2015) The efficiency challenge of nitride light-emitting diodes for lighting. *Phys. Status Solidi A* 212: 899–913.

Zegrya GG, Kharchenko VA (1992) New mechanism of Auger recombination of nonequilibrium current carriers in semiconductor heterostructures. *Sov. Phys. JETP* 74: 173–181.

Zegrya GG, Samosvat DM (2007) Mechanisms of Auger recombination in semiconducting quantum dots. *J. Exp. Theor. Phys.* 104: 951–965.

Zhmakin AI (2011) Enhancement of light extraction from light emitting diodes. *Phys. Rep.* 498: 189–241.

15

Organic Light-Emitting Diodes

15.1 Introduction 473
15.2 Working Principles of an OLED 474
Electrodes • Charge Transport Layers • Blocking Layers • Emission Layer
15.3 Drift–Diffusion Equations 478
Charges • Recombination • Excitons • Boundary Conditions and Numerical Solution • Limitations
15.4 Master Equation 480
Single-Carrier Charge Transport • Finite Charge-Carrier Densities • Mobility and Diffusion Constant
15.5 Kinetic Monte Carlo 482
Variable Step Size Method • Forbidden Events • Efficiency and Parallelization
15.6 Rates 483
Charge Transfer • Electron–Hole Recombination • Energy Transfer
15.7 Density of States 485
Gaussian Disorder • Perturbative Approach • Hybrid Approaches
15.8 Reorganization Energy 487
15.9 Electronic Coupling Elements 488
15.10 Morphology 488
Classical Force Fields • Morphology Simulations
15.11 Scale Bridging 490
Stochastic Models • Parametrization of Gaussian Disorder Models • Tabulated Mobilities
15.12 Case Studies 494
Current–Voltage Characteristics • Impedance Spectroscopy • Efficiency • Electroluminescence of a White OLED • Degradation
15.13 Outlook 511

Pascal Kordt
Peter Bobbert
Reinder Coehoorn
Falk May
Christian Lennartz
and
Denis Andrienko

15.1 Introduction

The 2014 Nobel Prize in Physics was awarded “for the invention of efficient blue light-emitting diodes (LEDs) which has enabled bright and energy-saving white light sources” [1–3], setting a clear target to mankind: energy-efficient and environmental-friendly light sources [4]. The first *organic* LED (OLED) was reported in 1987 by a team at Kodak [5]. This publication, cited to date more than 10,000 times, stipulated the entire field of organic electronics. Shortly afterward, a polymer LED (PLED) was demonstrated [6], paving the way for flexible lighting applications [7]. Nowadays, OLEDs are successfully used in displays of mobile phones and televisions: In 2008, Samsung announced a flexible display that was only 50 μm thick [8], about half the thickness of a sheet of paper. A prototype of an OLED display for the automotive market was presented recently by Continental [9]: In OLED displays, black pixels are completely switched off, allowing the driver's eye to adapt better to the darkness. Contrary to liquid crystal displays (LCDs), OLED screens do not require backlight illumination, yielding exceptionally good contrast ratios and reduced power consumption. OLED displays also provide viewing angles and response times superior to LCDs and are, in general, thinner and lighter. Last but not least, many organic materials can be printed from solution, enabling cost-effective large-scale manufacturing on mechanically flexible films.

The prime challenge in OLED development is an improvement of device operation lifetimes. In many cases, the steady decrease of luminescence efficiency of OLEDs under continuous operation is compensated by a gradual rise in bias voltages, or by using larger pixel sizes, which can then be operated at lower voltages and luminances. In particular, blue phosphorescent devices, with their high energy of emitted photons and long-lived excited states, are prone to rapid material degradation. The underlying mechanisms of degradation, discussed in Section 15.12.5, are unfortunately not fully understood.

Designing new organic materials is crucial for tuning OLED properties, performance, and stability. One, therefore, needs to provide rigorous links between device characteristics and the chemical composition of the layers, or structure–property relationships. This can be done with the help of cheminformatics tools [10], i.e., by correlating available experimental data to underlying chemical structures. However, the number of experimental samples is normally limited to a few hundred since synthesis, device optimization, and characterization are often costly and time consuming. A substantial extension of the training set is therefore experimentally not feasible to achieve, again motivating the development of computer simulations techniques capable of predicting device characteristics *in silico* [11]. In this chapter, we aim to provide an overview of theoretical models and simulation approaches developed to study charge and exciton transport in organic semiconductors, as well as to predict and eventually optimize current–voltage characteristics, electroluminescence efficiency, and lifetimes of OLEDs.

15.2 Working Principles of an OLED

We start by reviewing the elementary processes taking place in an OLED. The simplest two-layer OLED consists of hole and (luminescent) electron transporting layers (ETLs), which are sandwiched between two electrodes, as shown in Figure 15.1a. When a voltage is applied to the electrodes, holes are injected from the anode into the hole transporting layer (HTL) and electrons from the cathode into the ETL. The external field forces electrons and holes to drift toward the interface between these layers, where they recombine and emit light. Energetically, such a heterojunction is designed to facilitate the hole injection from HTL to ETL as well as to block electron diffusion into the opposite direction.

To fine-tune device properties such as luminescence, driving voltage, light outcoupling, and lifetime, more layers are added to a two-layer OLED: A typical phosphorescent OLED is shown in Figure 15.1b. Here, when a voltage is applied to the electrodes, electrons are injected from the reflective metal cathode (Al) and holes from the semitransparent anode (indium tin oxide or ITO). The cathode has a low work function, while the anode has a high one. This energy difference is compensated by the externally applied voltage, which forces charge carriers to drift into the emission layer. In the emission layer, holes and electrons form excitons, predominantly on the emitter (guest) molecules. Finally, radiative decay of an exciton leads to light emission. In the next sections, we briefly review the functionality of all layers.

15.2.1 Electrodes

The outermost layers, metallic or metal-oxide electrodes, inject electrons and holes into organic layers. A common material for the semitransparent anode is ITO [12]. Its low ionization potential (IP) ensures

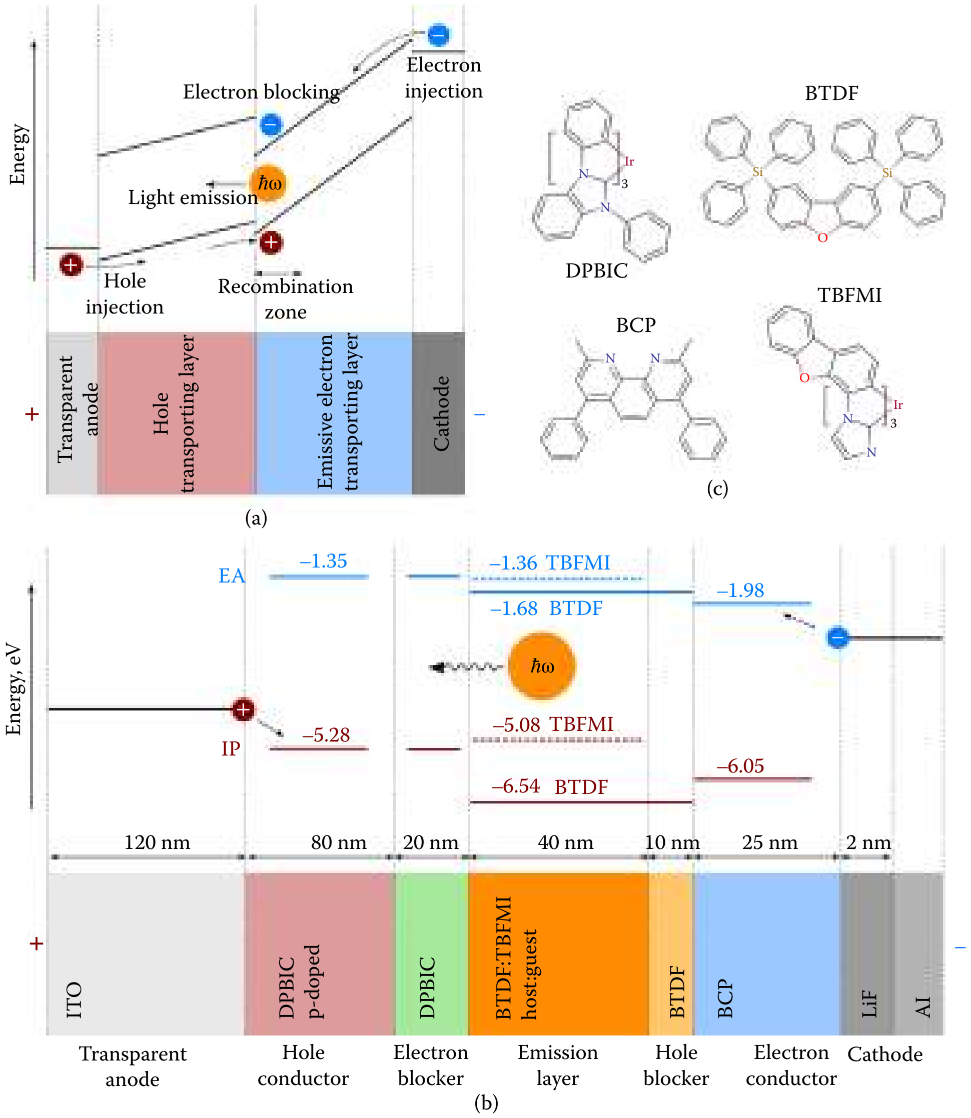

Energy
Electron blocking
Electron injection
Light emission
ħω
Hole injection
Recombination zone
+
Transparent anode
Hole transporting layer
Emissive electron transporting layer
Cathode
−
(a)
BTDF
DPBIC
BCP
TBFMI
Ir
3
N
Si
O
(c)
Energy, eV
EA
IP
−1.35
−1.36 TBFMI
−1.68 BTDF
−1.98
−5.28
−5.08 TBFMI
−6.54 BTDF
−6.05
120 nm
80 nm
20 nm
40 nm
10 nm
25 nm
2 nm
ITO
DPBIC p-doped
DPBIC
BTDF:TBFMI host:guest
BTDF
BCP
LiF
Al
Transparent anode
Hole conductor
Electron blocker
Emission layer
Hole blocker
Electron conductor
Cathode
(b)

FIGURE 15.1 (a) Schematic of a two-layer electroluminescent organic light-emitting diode (OLED) and its energy level diagram under applied voltage. (b) Structure of a blue phosphorescent OLED with a transparent anode (indium tin oxide or ITO), a reflective cathode (Al), emission layer, and hole/electron conduction/blocking layers. Indicated energy levels are without the applied bias. (c) Compounds used in the OLED (b). (Adapted from P. Kordt et al., *Adv. Funct. Mater.*, 25, 1955–1971, 2015.)

efficient hole injection into the organic layer. For efficient electron injection a low work function cathode is required. Since such materials are often air-unstable, normally bilayered structures are used [13], e.g., Al/LiF [14,15] or Ag/MgAg [16,17].

15.2.2 Charge Transport Layers

The role of electron and hole transport layers (HTLs) is to provide an Ohmic contact to the electrodes (barrierless injection) and to help control light outcoupling. These layers are often doped. By inserting electron donating or accepting impurities into a material, their intrinsic charge carrier density is increased, leading to higher conductivities and a reduction of the injection barrier. While it is a standard technique in inorganic semiconductors, it can be challenging in organic materials: Dopant molecules create energetic traps, hindering the formation of mobile charges [18].

Realizing *p*-type doping, which increases the hole density, requires a dopant material with a lowest unoccupied molecular orbital (LUMO) level close to the highest occupied molecular orbital (HOMO) of the host. This is feasible in most cases and, consequently, it is used in many OLEDs. For example, the OLED stack in Figure 15.1b has a *p*-doped Tris[(3-phenyl-1H-benzimidazol-1-yl-2(3H)-ylidene)-1,2-phenylene]Ir (DPBIC) HTL. On the contrary, *n*-type doping requires a dopant material with an HOMO level near or above the LUMO of the host. The necessity to find air-stable dopant materials with such a high HOMO level makes *n*-type doping challenging. Moreover, when using small molecules such as O_2, Br_2, and I_2 as dopant materials, these can diffuse into an organic host, leading to undesired doping profiles. This effect, however, can also be used to actually dope the adjacent layer with an interlayer of dopant molecules. In the stack shown in Figure 15.1b this technique is applied: Here an LiF interlayer is placed between the aluminum cathode and the electron transport layer (ETL) doping the latter [19]. Alq_3 (aluminum-tris(8-hydroxychinolin)) is a material that can be used for both the hole- and electron-conducting layers. In the stack in Figure 15.1b DPBIC is used for hole conduction and bathocuproine-4,7-diphenyl-2,9-dimethyl-1,10-phenanthroline (BCP) serves as an electron-conducting layer [14,16,17,20,21].

15.2.3 Blocking Layers

Hole/electron blocking layers suppress charge flow to the opposite electrodes, enhancing their recombination probability in the emission layer. The hole blocking layer has an IP lower than the IP of the emission layer. Similarly, an electron blocking layer has an electron affinity (EA) higher than the EA of the emission layer. In Figure 15.1, DPBIC is used as an electron blocker and BCP as a hole blocker [14,16,17,20,21].

15.2.4 Emission Layer

In the emission layer, electrons and holes recombine, leading to exciton formation. Upon radiative decay of the excitons, photons are emitted. To avoid exciton quenching, emitters (host molecules) are embedded into a matrix (guest molecules). In the stack shown in Figure 15.1b, Tris[(1,2-dibenzofurane-4-ylene)(3-methyl-1/1-imidazole-1-yl-2(3/1)-ylidene)]Ir(III) (TBFMI) is a blue phosphorescent emitter, while 8-bis(triphenylsilyl)-dibenzofuran (BTDF) is the host material.

Exciton generation and the emission of photons proceeds in several steps, which are illustrated in Figure 15.2. First, a carrier (in this case a hole) is trapped on an emitter molecule and an oxidized complex is formed. Driven by Coulomb forces, a carrier of opposite charge (electron) moves on host molecules toward the trapped carrier. When the electron reaches a host molecule' neighboring oxidized emitter, a charge transfer (CT) state is formed, as shown in Figure 15.2b. In a CT state, the short-range exchange interaction leads to an energy splitting of singlet (S) and triplet (T) states. The triplet consists of three triplet substates, which differ from one another by their relative spin orientations. Statistically, one obtains a population ratio of 1:3 between singlet and triplet substates. In a final step, the electron moves in a very rapid process directly to the emitter molecule, forming an excited emitter, see Figure 15.2c. This process may occur via a singlet or a triplet path, depending on the initial spin orientation of the electron–hole pair. The two spins of an electron and a hole are then coupled to four new combined states: one singlet state, with total spin momentum 0, and one triplet state, with the total spin momentum 1. In a statistical limit, all four substates are populated with an equal probability.

An exciton formed on an emitter molecule can decay radiatively by fluorescence or phosphorescence. In *fluorescence*, which is the most common radiative decay pathway for organic molecules, the radiative decay occurs from the excited singlet state, S_1, to the ground state. If triplet states are not harvested, the internal quantum efficiency of an OLED is limited to 25%. *Phosphorescence*, conversely, allows harvesting of

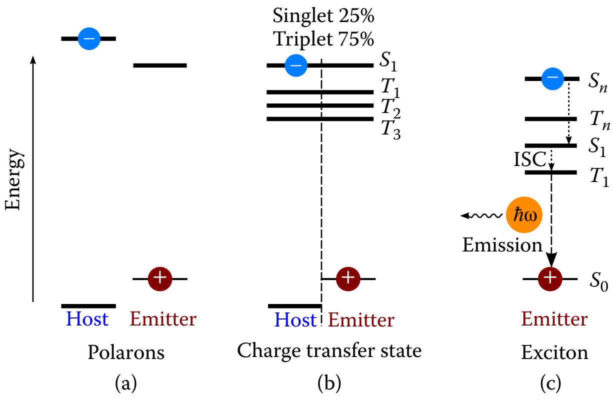


FIGURE 15.2 Exciton formation in the emission layer of a phosphorescent organic light-emitting diode (OLED). (a) Electron moves toward a hole which is trapped on the emitter molecule. (b) Charge transfer state is formed with the electron on a host and hole on a neighboring emitter molecule. Spins of hole and electron are already correlated to one singlet and three triplet states. (c) Exciton is formed on the emitter molecule with a statistical probability to be in a 25% singlet and 75% triplet state. Singlet state undergoes fast intersystem crossing (ISC) into a triplet state which then emits light.

triplet states. One can use compounds with large spin–orbit couplings, such as organometallic complexes, to achieve this. The large spin–orbit coupling capacitates a spin flip, (see Figure 15.2c), leading to ISC, i.e., a transition between singlet and triplet states. Thus, triplets can also decay radiatively ($T_1 \rightarrow S_0$), yielding an internal quantum efficiency of 100%.

Another way of harvesting triplet states is via a *thermally activated delayed fluorescence* (TADF) [22]. TADF can be realized in organic molecules even with small spin–orbit coupling. In TADF triplets are converted to singlets thermally (reverse ISC) and a radiative $S_1 \rightarrow S_0$ decay (fluorescence) takes place. TADF requires small energy differences between T_1 and S_1 levels $\Delta E_{ST} \sim k_B T$. To achieve this, a large spatial HOMO–LUMO separation is required. This, however, competes with the requirement of large oscillator strengths of fluorescence, i.e., a large transition dipole moment [23].

Several processes in an OLED can lead to efficiency losses. In a phosphorescent OLED, these are *triplet–triplet quenching*, where two triplets are annihilated as a result of the interaction of neighboring excited emitter molecules, and *triplet–polaron quenching*, i.e., an interaction between electronic and vibrational modes, energy transfer to polarons, and dissociation into free carriers [12,24]. These processes are discussed in more detail in Sections 15.12.3 and 15.12.5.

A comprehensive OLED model should incorporate all the aforementioned elementary processes: charge injection from electrodes, charge trapping, and transport to the emission layer, long-range electron–hole interactions, formation of CT and excited states, exciton–exciton and exciton–polaron interactions, and radiative decay of excitons. The ultimate goal of such a model is to predict the current–voltage–luminescence characteristics of the multilayered OLED structure. Since the typical thickness of an organic layer of an OLED stack ranges from 10 to 100 nm, it is computationally prohibitive to use only first-principles methods for OLED modeling. State-of-the-art OLED simulations employ either continuous models, such as drift–diffusion equations (see Section 15.3), or solve the master equation for charge/exciton occupation probabilities (see Section 15.4). Both approaches use models of different complexity. For example, if the target is to optimize the composition of the stack made of well-characterized organic layers, drift–diffusion equations and lattice models with phenomenologically fitted parameters are used. If one also needs to retain the link to the underlying chemical composition of layers, off-lattice models based on first-principles parametrizations are required. In all cases, we are dealing with a typical multiscale problem, which requires the development of scale-bridging techniques (see Section 15.11). In the next sections, we review these approaches, paying special attention to limitations, parametrizations, and scale-bridging issues.

15.3 Drift–Diffusion Equations

We start with the most coarse model: On a macroscopic level, the drift and diffusion of electrons, holes, and excitons in an OLED stack can be described by the corresponding densities, denoted here as n, p, and s, respectively. The drift–diffusion model assumes that local charge densities, charge mobilities, μ_n, μ_p, diffusion constants, D_e, D_n, and field strength, F, all vary continuously in space. The set of equations describing the time-dependent spatial distribution of charge and exciton densities is then based on the respective conservation laws. In what follows, we describe a model where charges drift-diffuse in a given density of states (DOS) $g(E)$ (see also Section 15.7), can be trapped, have density- and field-dependent mobilities, and recombine either radiatively or nonradiatively.

15.3.1 Charges

Charge conservation leads to two continuity equations, for electrons and holes

$$e\frac{\partial n}{\partial t} = \vec{\nabla}\cdot\vec{J}_n - eR(n,p) - e\frac{\partial n_{\mathrm{t}}}{\partial t}, \qquad e\frac{\partial p}{\partial t} = -\vec{\nabla}\cdot\vec{J}_p - eR(n,p) - e\frac{\partial p_{\mathrm{t}}}{\partial t}, \tag{15.1}$$

where $R(n,p)$ is the recombination rate (see Section 15.3.2), n_{t} and p_{t} are densities of trapped charges, and e is the electron charge. The current equations for electron and holes drift–diffusion in an electrostatic potential ψ read [25,26]

$$\vec{J}_n = -en\mu_n\vec{\nabla}\psi + eD_n\vec{\nabla}n, \qquad \vec{J}_p = -ep\mu_p\vec{\nabla}\psi - eD_p\vec{\nabla}p. \tag{15.2}$$

Note that only mobile charges contribute to the current. Summing up electron and hole currents yields the total current in the device, $\vec{J} = \vec{J}_n + \vec{J}_p$. The electrostatic potential ψ is related to the electron and hole densities via the Poisson equation

$$\varepsilon_0\varepsilon_r\Delta\psi = e\left(n - p + n_{\mathrm{t}} - p_{\mathrm{t}}\right), \tag{15.3}$$

where ε_0 is the vacuum and ε_r is the relative permittivity. The densities of trapped charges obey the phenomenological rate equations

$$\frac{\partial n_{\mathrm{t}}}{\partial t} = r_{\mathrm{c},n}n(N_t - n_{\mathrm{t}}) - r_{\mathrm{e},n}n_t, \qquad \frac{\partial p_{\mathrm{t}}}{\partial t} = r_{\mathrm{c},p}p(N_t - p_{\mathrm{t}}) - r_{\mathrm{e},p}p_t, \tag{15.4}$$

where it is assumed that traps occupy two levels (one for holes and one for electrons). Here N_{t} is the trap density, r_{e} is the escape rate, and r_{c} is the capture rates of electrons (n) and holes (p), respectively.

Since two electrons or holes cannot occupy the same energy level at the same time, the occupation probabilities of energy levels follow the Fermi–Dirac statistics [27]

$$f(E,E_{\mathrm{F}}) = \left[1 + \exp\left(\frac{E - E_{\mathrm{F}}}{k_{\mathrm{B}}T}\right)\right]^{-1}, \tag{15.5}$$

where $f(E,E_{\mathrm{F}})$ is the occupation probability of a level with energy E, k_{B} is the Boltzmann constant, and T is the temperature. The carrier density is then related to the quasi-Fermi level, E_{F}, as

$$p\left(E_{\mathrm{F}}\right) = \frac{N}{V}\int_{-\infty}^{\infty} g(E)f(E,E_{\mathrm{F}})\mathrm{d}E, \tag{15.6}$$

where $g(E)$ is the DOS (see Section 15.7), N is the number of holes, and V is the box volume. A similar relation holds also for electrons. Fermi–Dirac statistics implies that the diffusion coefficient and mobility in Equation 15.1 are related via the generalized Einstein relation [28]

$$D = \frac{p\mu}{e}\left(\frac{\partial p}{\partial E_{\mathrm{F}}}\right)^{-1}. \tag{15.7}$$

In order to solve the drift–diffusion equations one needs to know the dependence of mobility on charge density, electric field, and temperature, $\mu(p, n, \vec{F}, T)$. These can be obtained either from experiments or Monte Carlo simulations, as discussed in Section 15.11.

15.3.2 Recombination

The recombination rate $R(n, p)$ includes two loss mechanisms: Shockley–Read–Hall (SRH), or trap-assisted recombination, and bimolecular recombination. In SRH recombination, electrons are trapped in low-energy states and recombine with free holes. The macroscopic rate for this process reads [29]

$$R_{\mathrm{SRH}} = \frac{C_n C_p N_{\mathrm{t}}(np - n_i p_i)}{C_n(n + n_i) + C_p(p + p_i)}. \tag{15.8}$$

Here C_n and C_p are the capture coefficients for electrons and holes, respectively, N_{t} is the trap density for electrons, n and p are the electron and hole densities, and n_i and p_i are the intrinsic electron and hole densities.

Bimolecular recombination is often modeled using the Langevin rate

$$R_{\mathrm{L}} = \gamma(np - n_i p_i), \qquad \gamma = \frac{q}{\varepsilon_0 \varepsilon_{\mathrm{r}}}\left(\mu_n + \mu_p\right). \tag{15.9}$$

Here γ is the recombination constant and μ_n and μ_p are electron and hole mobilities. The intrinsic electron and hole carrier densities are often neglected [29,30], simplifying the expression to

$$R_{\mathrm{L}} \approx \gamma np. \tag{15.10}$$

There are several assumptions on which the Langevin rate is based and which might not hold in the case of organic semiconductors. Its validity can be verified by performing kinetic Monte Carlo (KMC) simulations on lattices, as discussed in Section 15.6.2.

The overall recombination rate, $R_{\mathrm{SRH}} + R_{\mathrm{L}}$, enters the continuity Equation 15.1. Experimentally, the prevailing recombination mechanism can be probed by means of the classical Shockley diode equation with an ideality factor η that differs depending on the dominant mechanism [31,32].

15.3.3 Excitons

The generation, transport, and decay of excitons can be described with a phenomenological equation for the population of excitons S_i, where i denotes the exciton type (triplet, singlet) [26]

$$\frac{\partial S_i}{\partial t} = G_i R(n, p) + \vec{\nabla} \cdot \vec{J}_i - \left(k_i^{(r)} + k_i^{(n)}\right) S_i - k_i^{(a)} f S_i^2 + \sum_{j=S,T} (k_{ji} S_j - k_{ij} S_i) - k_{\mathrm{TPQ}}(n + p). \tag{15.11}$$

Here G_i is the exciton generation efficiency: For singlet excitons $G_{\mathrm{S}} = 0.25$, while for triplet excitons $G_{\mathrm{T}} = 0.75$. $k^{(r)}$ and $k^{(n)}$ are the radiative and nonradiative (position-dependent) decay rates, and $k^{(a)}$ is the annihilation rate. The factor f is 0.5 if only one triplet is lost or 1 if both are lost. The radiative decay rate can be calculated using the dipole emission model [26,33]. The exciton energy transfer rate, k_{ij}, describes the conversion of triplet excitons to singlets and vice versa. k_{TPQ} is the triplet–polaron quenching rate [24] (assumed here to be the same for electrons and holes). Since excitons are charge-neutral, their transport is purely diffusive, i.e., $\vec{J}_i = D_i \vec{\nabla} S_i$. Note that the exciton dissociation into an electron and a hole has been neglected in Equation 15.11.

15.3.4 Boundary Conditions and Numerical Solution

Equations 15.1 through 15.11 are complemented by the boundary conditions for the electrostatic potential, ψ, by setting the potential difference at the boundaries to $\psi_{\mathrm{eff}} = V_{\mathrm{app}} - V_{\mathrm{int}}$, where V_{app} is the applied potential and V_{int} is the built-in potential, defined as the difference of the materials' work functions.

The set of Equations 15.1 through 15.11 is normally solved using iterative schemes, until self-consistency for the electrostatic potential, density, and current is reached. The Gummel iteration method [34] with a discretization according to a scheme proposed by Scharfetter and Gummel [35] can be used to solve linearized equations. This method is less sensitive to the initial guess than a Newton algorithm and thus is the method of choice despite its slower convergence in terms of iteration steps [36].

15.3.5 Limitations

Field, temperature, and charge density dependencies of the mobility entering drift–diffusion equations are normally parametrized in equilibrium or under stationary conditions. These dependencies cannot in general be used to describe nonequilibrium, time-dependent processes. For example, charge-carrier relaxation makes the mobility effectively time dependent, as evident from, e.g., impedance and dark-injection studies (see Section 15.12.2). Moreover, it has been shown that the current density in the device can be spatially inhomogeneous in all three dimensions (filamentary structures) [37–39], which then questions the applicability of mean-field descriptions. Last, some processes, e.g., exciton–electron interactions or energetic barriers between organic layers, are difficult to incorporate into drift–diffusion equations. The master equation, which we review in the next section, allows us to consider the individual rates between CT and other events. In the subsequent section, we describe the Monte Carlo technique that also allows us to include excitonic processes.

15.4 Master Equation

In inorganic, crystalline semiconductors charges are delocalized and energy eigenvalues of the electronic Hamiltonian form smooth bands. Consequently, one deals with band-like charge transport, where mobility decreases with increasing temperature. In contrast, in amorphous organic semiconductors, molecules are weakly bound by van der Waals forces and, consequently, intermolecular electronic couplings are small. As a result, excess electrons are localized on single molecules or their fragments. Moreover, orientational and positional *energetic* disorder helps further localize charged excitations. In this situation, charge transport can be modeled as thermally activated charge *hopping* between neighboring molecules. Mathematically, this is described by a Poisson process: The probability of an event to happen depends only on the time interval (no memory) and events do not occur simultaneously. The corresponding master equation, which describes the time evolution of a system with a discrete set of states i, then reads

$$\frac{\mathrm{d}P_i(t)}{\mathrm{d}t} = \sum_{j \neq i} \left[w_{ij} P_i(t) - w_{ji} P_j(t) \right], \tag{15.12}$$

where $P_i(t)$ is the probability of finding the system in state i at time t and w_{ij} is the transition rate from state i to j. In most situations, the dimension of the state space is too large to use direct numerical differential equation solvers to solve for P_i. In special cases, e.g., for single charge carrier transport or transport of many carriers in a mean-field approximation, this equation can be rewritten in terms of site occupation probabilities, p_i.

15.4.1 Single-Carrier Charge Transport

If only one charge carrier is present in the system, the state of the system is fully determined by the position of the charge or the index of the molecule which this charge occupies. In this situation, the occupation probability of the system state, P_i, is equivalent to the site (molecule) occupation probability, p_i. The rates for transitions between states are then given by CT rates, ω_{ij} (see Section 15.6), and the master equation can be rewritten in terms of site occupation probabilities

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t} = \sum_{j\neq i} \left[\omega_{ij} p_i(t) - \omega_{ji} p_j(t)\right]. \tag{15.13}$$

Equation 15.13 is a set of linear equations of size N, where N is the number of molecules in the system. It can be solved using standard numerical differential equation solvers [40]. In special cases, e.g., for one-dimensional charge transport, it is even possible to obtain an analytic solution [41–43]. Alternatively, one can use the KMC algorithm, which is discussed in Section 15.5.

15.4.2 Finite Charge-Carrier Densities

In experimentally relevant conditions, charge densities have finite values. The system state is now given by a vector of indices of occupied molecules and the number of states increases dramatically. It is, however, still possible to reformulate the master Equation 15.12 in terms of site occupation probabilities by using a mean field approximation [44,45]. The master equation then reads

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = \sum_{j\neq i} \left[\omega_{ij} p_i \left(1 - p_j\right) - \omega_{ji} p_j \left(1 - p_i\right)\right]. \tag{15.14}$$

The resulting equation is no longer linear in p_i and thus requires more involved numerical solvers. In most cases, where one deals with many carriers the KMC method (see Section 15.5) becomes more practical: It does not rely on the mean field approximation but solves the original master equation, and it can easily be extended to other processes, such as exciton transport, triplet–triplet annihilation, etc.

15.4.3 Mobility and Diffusion Constant

The stationary solution of the master equation for a system in an external field $\vec{F}$ allows to evaluate both charge carrier mobility and diffusion constant. For single-carrier transport, the mobility tensor $\hat{\mu} = \vec{v} \otimes \vec{F}/F^2$ reads

$$\mu_{\alpha\beta} = F^{-2} \sum_{ij} \omega_{ij} p_i r_{ij,\alpha} F_\beta, \tag{15.15}$$

where $r_{ij,\alpha} = r_{i,\alpha} - r_{j,\alpha}$ and $\vec{v}$ is the average velocity of a charge carrier [45]. For finite charge carrier density $\rho = N/V$, where N is the number of carriers and V is the volume of the box, Equation 15.15 takes the form

$$\mu_{\alpha\beta} = \frac{1}{\rho F^2 V} \sum_{ij} \omega_{ij} p_i \left(1 - p_j\right) r_{ij,\alpha} F_\beta. \tag{15.16}$$

In a similar fashion, occupation probabilities can be used to calculate charge and current distributions in the system [45]. Alternatively, one can directly analyze charge trajectories of KMC simulations and evaluate mobility tensor as

$$\hat{\mu} = \frac{\langle \vec{v} \rangle \otimes \vec{F}}{F^2}, \qquad \text{with} \qquad \langle \vec{v} \rangle = \frac{\Delta \vec{r}}{\Delta t}. \tag{15.17}$$

Here $\vec{F}$ is the external electric field, $\vec{r}$ is the charge position, and Δt is the simulation time. The diffusion tensor can also be calculated directly from the charge trajectory generated without the applied external field [46]

$$D_{\alpha\beta} = \frac{\left\langle \Delta r_\alpha \Delta r_\beta \right\rangle}{2\tau}, \tag{15.18}$$

where $\Delta r_\alpha(t) = r_\alpha(t+\tau) - r_\alpha(\tau)$, or can be obtained from charge mobility with the help of the generalized Einstein relation, Equation 15.7.

15.5 Kinetic Monte Carlo

As we saw in the previous section, rewriting the master Equation 15.12 to a form tractable for large systems (i.e., in terms of site occupation probabilities) requires certain approximations. Moreover, this step becomes less and less straightforward if events other than CT are included, e.g., electron–hole recombination, exciton splitting, exciton decay, or transfer. For these reasons, a different way of solving the master equation becomes more practical. Here, system dynamics, or trajectories in the phase space, or Markov chains, are generated explicitly by using the so-called KMC method.

KMC, also known as dynamic Monte Carlo, Gillespie algorithm, residence-time algorithm, *n*-fold way, or the Botz–Kalos–Lebowitz algorithm, was initially developed by Doob [47,48]. The time update as it is used here was first introduced by Young and Elcock [49]. Bortz, Kalos, and Lebowitz developed the same algorithm independently and applied it to the Ising model [50]. Gillespie provided a physics-based derivation of the algorithm [51,52], which was then improved in terms of computational efficiency by Fichthorn [53] and Jansen [54]. The version we describe here is known as a variable step size method (VSSM) [54].

15.5.1 Variable Step Size Method

The VSSM allows us to group certain events and treat the groups and events in the group hierarchically. In the case of charge transport with multiple charge carriers, e.g., a two-level approach can be used. First, a charge is selected with the probability proportional to its escape rate ω_i, that is the sum of the rates of all possible moves of this charge away from the occupied site

$$\omega_i = \sum_{j=1}^{m} \omega_{ij}, \tag{15.19}$$

where m is the coordination number, i.e., the number of neighbors to which the site is connected, and ω_{ij} is the rate for all possible moves to connected sites. Afterward, the destination site is chosen with a probability ω_{ij}/ω_i. Since we are dealing with a Poisson process, waiting times are exponentially distributed with a parameter λ that is the inverse of the sum of all escape rates

$$\Delta t \sim \exp(\lambda), \ \lambda = \left(\sum_i \omega_i \right)^{-1}. \tag{15.20}$$

In order to reproduce this distribution, the time in the VSSM algorithm is updated after each move with Δt drawn from an exponential distribution. In practice, this is achieved by drawing a random number u from a uniform distribution on the interval $(0, 1]$ and setting Δt to

$$\Delta t = -\lambda \ln(u), \qquad u \sim \mathcal{U}((0,1]). \tag{15.21}$$

15.5.2 Forbidden Events

The VSSM algorithm can be adapted to efficiently treat forbidden events [54]. In the case of charge transport, e.g., each site (molecule) can be occupied by one charge carrier at a time (the Pauli exclusion principle). As a result, charge carriers obey Fermi–Dirac occupation statistics (see Section 15.3). Hence, if site j is already occupied, all incoming rates, $\omega_{i\to j}$ should be set to zero. This is computationally inefficient, since all rates (as well as escape rates) must be updated after every event. A much more efficient way is to keep the forbidden event in the event list. Once this event is attempted, i.e., a carrier attempts to hop to an occupied site, this destination site is marked as forbidden and another destination is selected from the remaining possibilities. If all surrounding sites are occupied, the algorithm switches to the level above, i.e., a different charge is selected and the previous one is added to a temporary list of forbidden events. The time is updated regardless of whether or not the event is forbidden (the charge moved or not). One can show that this strategy results in the same statistics as *a priori* removing all forbidden events from the event list [54].

15.5.3 Efficiency and Parallelization

Developing an efficient KMC code for charge/exciton transport is a rather challenging task. Indeed, if only the Pauli exclusion principle is taken into account, a single charge transfer event does not affect the rates, and only the (local) list of forbidden events should be updated. Hence, all rates can be precomputed before the KMC simulation is performed. If (long-range) Coulomb interactions are included, every CT event changes all rates. In other words, energy differences between two states of the system depend now on the relative positions of all charges in the system. To avoid updating rates at every KMC step, one can use scale-separation techniques: Since mesoscopic charge densities evolve on time scales much longer than typical CT times, electrostatic fields created by far-off charges can be treated in a mean-field way, e.g., by solving the Poisson equation. In this case, one can update local rates at every step, while all rates are updated only every 100–1000 steps [55].

Even the local rate updates can become computationally costly for systems with many charge carriers. In this situation, the binary tree search algorithm helps improve computational efficiency by providing a faster way of identifying which events and escape rates should be updated. By using this algorithm one can gain a factor of 10 speed-up for systems with 10% occupied sites [55]. Another alternative is to use the next reaction method [56], an improved version of the first reaction method, where a dependency graph is used to determine which rates to update.

Apart from trivial parallelization, where multiple copies of the system with different initial conditions are run in parallel, Monte Carlo schemes can be parallelized by dividing the simulation volume into subvolumes treated as individual simulations [57–60]. An efficient parallel implementation on graphics processing units (GPUs) has also been demonstrated [61].

15.6 Rates

To parametrize the master Equation 15.12, we need to evaluate the rates of all elementary processes of interest. For example, for charge transport, CT rates must be evaluated for all neighboring molecular pairs. To make this computationally feasible, we often rely on simplified theoretical treatments of transfer reactions. In this section, we review several of such theories as well as introduce the corresponding rate expressions.

15.6.1 Charge Transfer

The simplest expression for the CT rate, proposed by Miller and Abrahams [62], is the rate of thermally activated, barrierless tunneling between two localized electronic states. The corresponding rate is proportional to the Boltzmann prefactor of the free energy difference between the initial and final states, ΔE_{ij}, and the electronic overlap between the states, which decays exponentially with the molecular separation

$$\omega_{ij} = \begin{cases} \omega_0 \exp\left(-2\gamma r_{ij}\right) \exp\left(-\frac{\Delta E_{ij}}{k_B T}\right) & \Delta E_{ij} > 0, \\ 1, & \Delta E_{ij} \leq 0, \end{cases} \tag{15.22}$$

where $1/\gamma$ is the localization length of the charge, r_{ij} is the distance between the two molecules, and ω_0 is the attempt frequency. The simplicity and intuitiveness of this rate justified its use in the early stages of understanding charge transport in organic semiconductors [63].

Environmental effects, as well as molecular reorganization upon charging/discharging (i.e., the energetic barrier between the initial and final states), have been accounted for in the so-called Marcus CT rate [64,65]

$$\omega_{ij} = \frac{2\pi}{\hbar} \frac{J_{ij}^2}{\sqrt{4\pi\lambda_{ij} k_B T}} \exp\left[-\frac{\left(\Delta E_{ij} - \lambda_{ij}\right)^2}{4\lambda_{ij} k_B T}\right]. \tag{15.23}$$

Again, $\Delta E_{ij} = E_i - E_j$ is the driving force or the free energy difference of the final and initial states of the CT reaction [66]. An accurate evaluation of these energies using polarizable force-fields is discussed in Section 15.7. J_{ij} is the electronic coupling element. They are intimately related to the overlap of the diabatic electronic states, as discussed in Section 15.9. Finally, λ_{ij} is the sum of the external and internal reorganization energies, discussed in more detail in Section 15.8.

The main issue with the classical Marcus rate is that the intramolecular vibrational modes promoting the CT reaction are energetically comparable to the C–C bond stretching mode at room temperature, $\hbar\omega_{CC} \sim 0.2\,\text{eV} \gg k_B T \sim 0.025\,\text{eV}$. Therefore, these modes should be treated quantum mechanically. For a common set of intramolecular high-frequency (quantum-mechanical) and an outer sphere low-frequency (classical) vibrational coordinates, a mixed quantum-classical multichannel generalization of the Marcus formula is readily available [66,67]. A generalization for the bimolecular electron transfer rate with independent sets of coordinates for donor and acceptor has also been proposed [45].

Another rate expression has been proposed by Weiss and Dorsey [68–70]. In the low temperature limit, $\hbar\nu_c/k_B T \gg 1$, it reads

$$\omega_{ij}(\epsilon) = \frac{J_{ij}^2}{\hbar^2 \nu_c} \left(\frac{\hbar\nu_c}{2\pi k_B T}\right)^{1-2\alpha} \frac{\left|\Gamma\left(\alpha + i\frac{\Delta E_{ij}}{2\pi k_B T}\right)\right|^2}{\Gamma(2\alpha)} \exp\left(\frac{\Delta E_{ij}}{2k_B T} - \frac{\left|\Delta E_{ij}\right|}{\hbar\nu_c}\right). \tag{15.24}$$

Here $\Gamma(z)$ is the gamma function and ν_c is the characteristic frequency, or the largest frequency in the Ohmic bath, which is related to the reorganization energy, λ, by $\lambda = 2\alpha\hbar\nu_c$. The Kondo parameter, α, describes the coupling strength between the charge and the heat bath [71]. In the high temperature limit, the Weiss–Dorsey rate simplifies to the Marcus rate. Indications that the Weiss–Dorsey rate is better suited for describing charge transport, especially at low temperatures and high fields, have recently been reported [72].

15.6.2 Electron–Hole Recombination

The microscopic process of electron–hole recombination is straightforward: Whenever a hole and an electron occupy the same site (molecule), they recombine with a certain rate, w_r. Radiative recombination leads to a photon emission which is the basis of OLED functionality.

On a macroscopic level, electron–hole recombination is traditionally described by the Langevin equation (see Section 15.3.2), with a prefactor proportional to the sum of electron and hole mobilities. Extensive KMC simulations were performed to verify this dependency [73]. An excellent agreement was found if the electron and hole mobilities are extracted from simulations with both carrier types, which was attributed to the change in mobilities upon inclusion of two carrier types due to their Coulomb attraction. Deviations at high charge concentrations ($>10^{-3}$ carriers per molecule) were also observed: Here, the average electron–hole distance becomes smaller than the thermal capture radius, $r_c = e^2/4\pi\epsilon\epsilon_0 k_B T$, violating assumptions used in the derivation of the Langevin formula.

15.6.3 Energy Transfer

We now briefly review rate expressions used to describe electronic excitation (energy) transfer (EET). EET is also known as a resonant energy transfer (RET).

A theory explaining the mechanism of EET was first proposed by Förster [74]. The electronic interaction promoting EET relies on a coupling of the donor and acceptor molecules via Coulombic interaction. Similar to CT, Förster theory relies on Fermi's Golden Rule with the electronic coupling between donor and acceptor treated perturbatively. Additional assumptions are that the system equilibrates after the electronic excitation of the donor on a time scale much faster than that of EET and that coupling to the bath (given by the absorption line shape) is much greater that the electronic coupling between donor and acceptor. Energy conservation in the weak coupling limit results in a coupling element, which is proportional to the overlap of the donor fluorescence spectrum with the acceptor absorption spectrum. The spectral overlap includes nuclear overlap (in the form of Franck–Condon factors), which depends on the spectral line shapes and thus provides the temperature dependence of the EET rate. Under these assumptions, the EET rate between emitters i (donor) and j (acceptor) takes a simple form

$$\omega_{ij}^{\mathrm{F}} = \frac{1}{\tau_{r,i}} \left(\frac{R_{\mathrm{F},ij}}{r_{ij}} \right)^6, \tag{15.25}$$

where $\tau_{r,i}$ is the radiative life time of the emitter (donor), r_{ij} is the separation between donor and acceptor, and $R_{\mathrm{F},ij}$ is the Förster radius for transfer from donor to acceptor.

An EET process can occur even when Coulomb-term-mediated electronic transitions are forbidden. Dexter provided a derivation for the case when the Coulombic interaction is negligible and EET is due to the exchange part [75]. This rate decays exponentially with intermolecular separation

$$\omega_{ij}^{\mathrm{D}} = k_{\mathrm{D}} \exp\left(-2\gamma r_{ij}\right), \tag{15.26}$$

where $1/\gamma$ is the wavefunction decay length and k_{D} is the Dexter prefactor proportional to the exchange integral.

Note that since the exchange integral is a quantum mechanical correction to the Coulombic repulsion, the total EET rate is always a sum of the two rates. However, due to the exponential decay of the Dexter coupling, the electronic interaction that mediates EET at separations greater than 5 Å is invariably Coulombic.

The importance of exchange and other short-range interactions, as well as higher multipole contributions to the Coulombic interaction, has been examined fairly extensively, helping improve the accuracy of rate expressions 15.25 and 15.26. An overview of these extensions can be found in Refs. [76–80].

15.7 Density of States

In the previous section, we saw that CT rates depend on the energies of the initial and final states of the CT complex. In an amorphous organic semiconductor, every molecule is embedded in a unique environment and therefore every molecule has its unique set of energy levels, EA, and IP. The set of energy levels available for an excess charge is termed the DOS. In this section, we discuss several methods of evaluation of the DOS of an amorphous organic semiconductor.

15.7.1 Gaussian Disorder

The simplest approach to model the DOS of organic semiconductor is to assume a Gaussian distribution of energy levels

$$g(E) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{\left(E-\bar{E}\right)^2}{2\sigma^2}\right]. \tag{15.27}$$

This phenomenological expression, first proposed by Bässler [63] and termed the Gaussian disorder model (GDM), is motivated as follows: Randomly oriented dipole moments in an amorphous material interact with one another, thus influencing the energy levels of neighboring molecules. As a result of the central limit theorem, this leads to approximately Gaussian distributed energy levels. The width of the distribution, σ, is called the *energetic disorder* and can be extracted from temperature-dependent mobility measurements [81, 82] or from simulations, as discussed in Section 15.11.1.

The original model was extended to the correlated disorder model (CDM) by accounting for spatial correlations that arise from the dipole interaction of neighboring molecules [83]. For randomly oriented dipoles, $\vec{p}_j$, the the electrostatic energy of a site i is given by [84]

$$E_i = -\frac{q}{4\varepsilon} \sum_{j\neq i} \frac{q\vec{p}_j \cdot (\vec{r}_j - \vec{r}_i)}{|\vec{r}_j - \vec{r}_i|^3}, \tag{15.28}$$

where ε is the dielectric permittivity and q is the charge. In the case of equal absolute values, $|\vec{p}_j| = p$, and dipoles fixed on a cubic lattice with lattice constant a, the sum can be evaluated using Ewald summation [85,86], yielding a Gaussian DOS with energetic disorder $\sigma = 2.35\, qp/\varepsilon\, a^2$ [83,87]. The spatial energy correlation function

$$\kappa(r) = \frac{\mathbb{E}\left[\left(E(\vec{r}_i) - \bar{E}\right)\left(E(\vec{r}_j) - \bar{E}\right)\right]}{\sigma^2} \tag{15.29}$$

is then given by $\kappa(r) \approx 0.74\frac{a}{r}$ [88]. Here $r = \left|\vec{r}_i - \vec{r}_j\right|$ is the distance between two molecules, $\bar{E}$ is the mean of the energy distribution, and $\mathbb{E}\left[\cdot\right]$ is the expectation value. Note that the spatial correlation function of this simple model system depends only on the lattice spacing a, limiting its applicability to realistic morphologies [89], as discussed in Section 15.11.1.

The Gaussian DOS is at the heart of the family of GDM, where Miller–Abrahams rates, Equation 15.22, with the Gaussian DOS are assumed for a charge hopping on a cubic lattice. Extensive KMC simulations of these models helped parametrize the mobility as a function of temperature, field, and field carrier density [63]. It was extended by including the important influence of charge carrier density in the extended Gaussian disorder model (EGDM) [90] without spatial correlation of site energies. The mobility expression in the EGDM at finite can be written as a product of three functions,

$$\mu(T, F, \rho) = \mu_0(T) g(T, \rho) f(T, F), \tag{15.30}$$

where

$$\mu_0(T) = 1.8 \times 10^{-9} \mu_0 \exp\left[-C\hat{\sigma}^2\right], \qquad g(T, \rho) = \exp\left[\frac{1}{2}\left(\hat{\sigma}^2 - \hat{\sigma}\right)\left(2\rho a^3\right)^{\delta}\right], \tag{15.31}$$

$$f(T, F) = \exp\left[0.44\left(\hat{\sigma}^{3/2} - 2.2\right)\left(\sqrt{1 + 0.8\hat{F}^2} - 1\right)\right], \qquad \delta = 2\frac{\ln\left(\hat{\sigma}^2 - \hat{\sigma}\right) - \ln(\ln 4)}{\hat{\sigma}^2},$$

with $C = 0.42$, $\hat{F} = eaF/\sigma$ and $\hat{\sigma} = \sigma/k_{\mathrm{B}}T$. A similar parametrization, termed the extended correlated disorder model (ECDM), also exists for the spatially correlated DOS [91].

15.7.2 Perturbative Approach

While Gaussian disorder models have been successful in describing many properties of organic semiconductors, they do not provide a direct link to the material morphology or chemical composition. A perturbative approach allows us to evaluate site energies in atomistically resolved (see Section 15.10), large-scale morphologies. Furthermore, it can also be used to parametrize Gaussian disorder models and perform large-scale simulations (see Section 15.11.2).

For every molecular pair, our quantity of interest is the energy difference $\Delta E_{ij} = E_i - E_j$, which is the energy separation between the minima of the diabatic potential energy surfaces (PES). In systems with weak intermolecular couplings, one can treat interactions with the environment perturbatively: The intermolecular electrostatic and induction contributions are then given by the first- and second-order terms in the expansion of the interaction energy [92].

The total energy energy of an ion embedded in a molecular environment thus includes an internal contribution E_i^{int}, i.e., the electron affinities for electrons and IPs for holes of *isolated* molecules. These can vary from one molecular pair to another because of different energy levels for different types of molecules, or different conformers of the same molecule. Correspondingly, the *external* contribution is due to the electrostatic and induction interactions, E_i^{el} and E_i^{ind}, of a charged molecule with the environment. These interaction energies are determined by the electrical charge distribution and the polarizability distribution, respectively, in the environment of the charged molecule. Overall, the total energy of molecule i is then given by

$$E_i = E_i^{\text{int}} + E_i^{\text{el}} + E_i^{\text{ind}}. \tag{15.32}$$

Most difficult to evaluate is the interaction with the environment. This is for two reasons: First, the underlying interactions are long-ranged and thus large system sizes are needed to converge the values of site energies. Second, special summation techniques are needed in order to evaluate interactions of an ion with a neutral *periodic* environment [93–96]. Also note that the perturbative evaluation of site energies relies on accurate molecular representations in terms of distributed multipoles and polarizabilities and is computationally demanding, in spite of being classical [97].

15.7.3 Hybrid Approaches

Using Equation 15.32, the site energies can be evaluated by solving a microscopic analogue of the Poisson equation. The self-consistent solution is normally achieved using iterative schemes for induced multipoles and Ewald summation techniques for static multipoles, which is computationally demanding. The role of hybrid schemes is to reduce computational cost. Here, one relies on the fact that the fields created by the far-off charges can be treated in a mean-field way, i.e., obtained by solving the Poisson equation. Nearest-neighbor interactions (i.e., interactions between molecular pairs within a certain cutoff distance) are still evaluated explicitly. This scheme also allows us to add metallic electrodes as image charges and has been used extensively to simulate multilayered OLED devices [98–100]. To improve computational efficiency, in this approach all induction interactions are taken into account effectively, by rescaling interaction energies by the (effective) dielectric constant of the medium.

15.8 Reorganization Energy

The reorganization energy, responsible for the energy barrier between diabatic states as well as for the broadening of energy levels of the electron detachment/attachment spectra, has two contributions. The *internal* reorganization energy is a measure for how much the geometry of the CT complex adapts while the charge is transferred. It can be estimated based on four points on the diabatic PES [45,101]

$$\lambda_{ij}^{\text{int}} = E_i^{nC} - E_i^{nN} + E_j^{cN} - E_j^{cC}. \tag{15.33}$$

Here, small letters denote the state and capital the geometry of a molecule, e.g., E_i^{nC} is the internal energy of the molecule i in the geometry of its charged state. Treatments that do not approximate the PES in terms of a single shared normal mode are also available [45,66,102].

An additional contribution to the overall reorganization energy results from the rearrangement of the environment in which the CT takes place. In a classical case, this outer-sphere reorganization energy, λ^{out}, contributes to the exponent in the rate expression in the same way as its internal counterpart. Assuming that CT is significantly slower than electronic polarization but much faster than the nuclear rearrangement of the environment, λ^{out} can be evaluated from the electric displacement fields created by the CT complex [66], provided that the Pekar factor is known. Alternatively, one can use polarizable force-fields [103] or quantum mechanics/molecular mechanics (QM/MM) methods [104]. It also turns out that the classical Marcus expression for the outersphere reorganization energy (inversely proportional to the molecular separation) can predict negative values of λ^{out} for small intramolecular separations, which are unphysical and hence should be used with care [45].

15.9 Electronic Coupling Elements

Electronic coupling elements, or transfer integrals , J_{ij}, entering the CT rate (Equations 15.23 and 15.24) are off-diagonal matrix elements, $J_{ij} = \left\langle \psi_i | \hat{H}^{\text{el}} | \psi_j \right\rangle$, of the electronic Hamiltonian, $\hat{H}^{\text{el}} = \hat{T}^{\text{el}} + \hat{V}^{\text{el-el}} + \hat{V}^{\text{nuc-el}}$, based on a diabatic (noninteracting) state, ψ_i [66]. A number of approaches can be used to evaluate electronic coupling elements. Their efficiency and accuracy depend on how the diabatic states and Hamiltonian are constructed, as well as how the matrix projection is performed.

Diabatic states are often approximated by the HOMO of monomers for hole transport, or the LUMO for electron transport ("frozen core" approximation) [105–107]. An approximate diabatic basis can also be constructed using constrained density functional theory [108].

The dimer Hamiltonian can be constructed using semiempirical methods, e.g., the Zerner's Intermediate Neglect of Differential Overlap (ZINDO) Hamiltonian [101,109–111]. This approach does not require a self-consistent evaluation of the dimer Hamiltonian and is therefore computationally very efficient [110]. One can also employ density functional theory (DFT) and use either the fully converged Hamiltonian or only the initial guess [105]. Another way to improve the efficiency is to reduce the number of orbitals for which electronic couplings are calculated. A detailed comparison of accuracy and efficiency of different approaches can be found in Refs. [105,112,113].

For (approximately) spherically shaped molecules the logarithm of the squared electronic coupling, $\log(J_{ij}^2)$, decays linearly as a function of the intermolecular separation r (at least for large distances), justifying the functional form of the Miller–Abrahams rate (see Equation 15.22). If molecular pairs are extracted from the respective dimers in the realistic atomistic morphology, $\log(J_{ij}^2)$ is often Gaussian distributed, with a mean and variance that depend on the molecular separation [55,114]. This observation can be used to parametrize coarse-grained (stochastic) charge transport models, as discussed in Section 15.11.1.

Since electronic couplings are related to the overlap of electronic orbitals participating in charge transport, they are very sensitive to the relative positions and orientations of molecules. Hence, an amorphous morphology of an organic material should be generated as precisely as possible. The corresponding methods are covered in the next section.

15.10 Morphology

In the previous sections, we explained how charge/exciton transport and recombination can be mapped on a series of events with specified rates. We have also emphasized that rates are very sensitive to both local and global molecular ordering. For computer-based predictions of material properties it is therefore important to simulate the material morphology as realistically as possible. In OLEDs, we usually deal with *amorphous* molecular arrangements—crystalline films are often less efficient in OLEDs because of the resulting large exciton diffusion distances, leading to enhanced exciton loss due to quenching processes. Amorphous films have a well-defined local structure, which depends on molecular interactions and processing conditions. In this section, we describe how amorphous morphologies of 10^4–10^5 molecules can be simulated using atomistic force fields or coarse-grained models. These systems can then be used to study small-scale charge transport [45,115–119], or to parameterize mesoscale models, as discussed in Section 15.11.1.

15.10.1 Classical Force Fields

The role of classical force fields in simulations of organic semiconductors is two-fold. First, they are used to generate atomistically resolved morphologies of molecular assemblies. Second, they are employed to evaluate the solid-state electrostatic and induction contributions to site energies (see Section 15.7). In both cases, these classical molecular representations should be appropriately parametrized.

For site energy calculations, it is important to evaluate electrostatic and induction contributions as accurately as possible. The corresponding parametrization is rather straightforward: A perturbative expansion of the intermolecular interaction energies is based on distributed multipoles (electrostatic interaction) and distributed polarizabilities (induction interaction) [92,97,120–122]. Van der Waals interactions are normally ignored since we are interested in free energy differences between charged and neutral states of the system. To improve computational efficiency, one can also use machine-learning techniques to devise simple structure–property relations for, e.g., atomic multipoles of molecular conformers [123].

Polarizable force fields based on distributed multipole expansions are still computationally prohibitive for simulating molecular arrangements of large systems. Moreover, parametrizations of effective *pairwise* potentials (partial charges and Lennard–Jones parameters), which mimic many-body van der Waals and induction interaction energies, are a nontrivial task, which largely relies on experimental input [124,125]. Standard forces fields, e.g., those suitable for biosystems, are often not transferable to organic molecules with large π-conjugated subsystems. A representative example is the comparison of the Williams 99 and the optimized potential for liquid simulations (OPLS) force fields for the amorphous mesophase of Alq_3 [126], which predict fairly different densities, radial distribution functions, and glass transition temperatures.

That being said, the development of new force fields for organics is currently limited to refitting of partial charges using the Merz–Singh–Kollman [127] scheme or the CHarges from ELectrostatic potentials using a grid-based method (CHELPG) [128] scheme, and a parametrization of missing bonded interactions from first-principles scans (see Figure 15.3). The remaining parameters are often taken from standard force fields, such as OPLS [124]. Reparametrized force fields are then verified against experimentally available densities and glass transition temperatures.

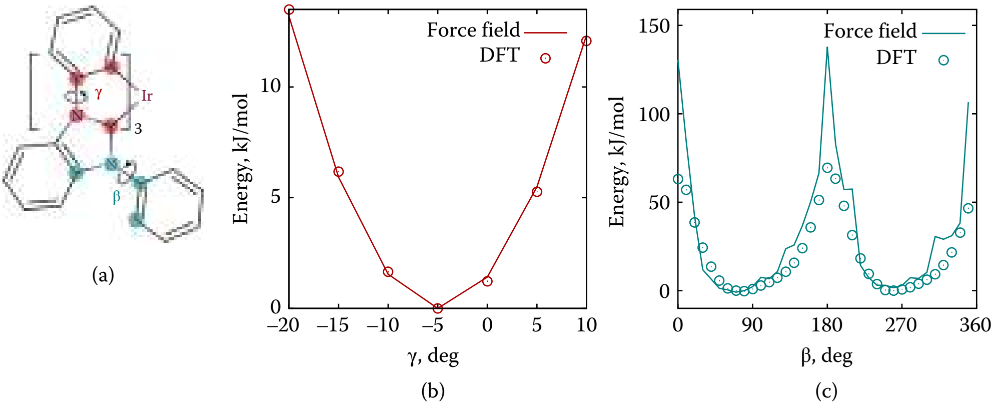


FIGURE 15.3 Potential energy scans of two dihedral angles of the DPBIC molecule, the chemical structure of which together with the definition of dihedral angles is shown in the inset (a). Symbols show DFT energies (B3LYP/6-311g(d,p)) and lines indicate the force-field energies after adjusting the force-field parameters. (Reproduced from P. Kordt et al., *Adv. Funct. Mater.*, 25, 1955–1971, 2015. With permission.)

15.10.2 Morphology Simulations

With the force field at hand, one can, in principle, simulate amorphous molecular assemblies. For systems with periodic boundaries in all directions this is done by first equilibrating the system above the glass transition temperature and then quenching it to room temperature in the *NPT* ensemble using molecular dynamics [126,129,130].

For thin slabs (2D-periodic systems) one can directly deposit molecules either using molecular dynamics [94] or Monte Carlo [131] techniques. In both cases the dynamics of deposition is not realistic. In other words, if surface diffusion plays an important role (e.g., in the case of guest aggregation in the host) these techniques cannot be applied in a straightforward manner, since the length scales and time scales required of host aggregation cannot be reached by atomistic molecular dynamics simulations. It is, however, possible to explore the fact that certain parts of molecules evolve on much slower time scales and larger-length scales and to combine several coherently moving atoms, connected via stiff degrees of freedom (e.g., bonds) into a single interaction site, as it is done, e.g., in the united atom force-fields with hydrogens incorporated into heavier atoms [132]. By doing this, we reduce the number of degrees of freedom to be propagated and, more importantly, obtain a much smoother potential energy landscape in terms of the *coarse-grained* degrees of freedom [133] (softer interaction potentials, less friction), allowing one to simulate 10–100 longer times and system sizes [132].

In order to perform correct statistical sampling of the coarse-grained degrees of freedom, the potential of mean force should be used as the interaction potential [134], which is inherently a many-body potential. To reduce computational cost, this potential is represented as the sum of a few functions, i.e., projected onto the basis functions of the force field. The accuracy of the coarse-grained model thus becomes sensitive to the way the projection is performed as well as the number of basis functions that are used to represent the coarse-grained force-field [135]. Existing projection schemes try to reproduce various pair distribution functions (structure-based coarse-graining [136–139]), to match the forces [134,140,141], to minimize the information loss in terms of relative entropy [142], or to use liquid state theory [143]. An extensive overview of such coarse-graining techniques is provided in Ref. [144]. Often, the accuracy of the coarse-grained model can be improved by explicitly incorporating information about macroscopic properties of the system, i.e., its equation of state or the symmetry of its mesophase [145–147]. Moreover, one can further reduce the number of degrees of freedom by introducing anisotropic interaction potentials. Eventually, atomistic details can be reintroduced into coarse-grained morphologies [148].

15.11 Scale Bridging

Now that we have seen how an OLED device can be modeled on different time and length scales, we eventually would like to transfer our knowledge about the system between the scales. For example, drift–diffusion equations (Section 15.3) can be used to calculate current–voltage curves of a device (micrometer scale). However, they require expressions for charge carrier mobility, the diffusion constant and, in the case recombination is taken into account, the recombination rate, all as a function of external field, carrier density, and temperature. Microscopic simulations (see Sections 15.4–15.10) can provide this information, at the same time retaining the link to the molecular structure. However, they are becoming computationally too demanding for high charge densities and large system sizes required to parametrize these dependencies. Our aim here is to provide several strategies, which can be used to link different time and length scales.

15.11.1 Stochastic Models

In an OLED, charges are inhomogeneously distributed [149] and charge density variations span several orders of magnitude. To cover the required density range in simulations, one needs to deal with relatively large systems—this quickly becomes computationally demanding if all rates are evaluated from first principles, as described in Section 15.6. To remedy the situation, one can devise a phenomenological algorithm to parametrize the master equation. In this section we outline how such an algorithm can be constructed for charge transport simulations.

We first note that the master equation (Section 15.4) is completely determined by the event rates. In the case of charge transport, these rates depend on site energy differences, electronic coupling elements, and reorganization energies. In order to evaluate observables of interest (e.g., charge mobility), one additionally needs site positions. Hence, the algorithm should be able to reproduce (statistically) several distribution and correlation functions.

Let us start with molecular positions. In an amorphous solid, both positions and orientations are completely defined by a set of (many-body) spatial correlation functions. For (approximately) spherically shaped molecules, the pair correlation function, or radial distribution function, $g(r)$, which quantifies the density of molecules at a separation r, contains the most relevant structural information. To reproduce this function approximately, one can use "thinning of a Poisson process" [114,150]. More accurate coarse-graining techniques, such as iterative Boltzmann inversion [138,151] or inverse Monte Carlo [136,137,152], allow a (numerically) exact reproduction of the radial distribution function [55]. These methods optimize a pair interaction potential, $U(r)$, in a way that the corresponding $g(r)$ is reproduced. They rely on the Henderson theorem [153], which states that there is a unique correspondence between $U(r)$ and $g(r)$. An illustration of this algorithm for an amorphous layer of DPBIC is presented in Figure 15.4a. The approach can also be applied to nonspherical molecules, by using several interacting sites per molecule [135].

The second ingredient of the stochastic model is the connectivity: In the atomistic model only molecules within a certain cutoff distance are used for calculating CT rates. The rest of the rates are set to zero. This is justified by the fact that electronic coupling elements decrease roughly exponentially with molecular separation [55], see, e.g., Equation 15.22. The distance that determines whether or not two molecules are connected is given by their two closest atoms. Since this information is not present in the coarse-grained model, the resulting probability of two sites to be connected, as a function of their center-of-mass separation, is given by the corresponding probability extracted from the reference data. Figure 15.4b shows this probability for an amorphous DPBIC layer.

As explained in Section 15.7, the stochastic generation of site energies should reproduce both their distribution function (DOS) and their spatial correlation. This can be achieved by mixing in site energy contributions of neighboring sites [89]. Figure 15.4c shows the spatial correlation function for an amorphous DPBIC layer.

Electronic coupling elements can also be generated using appropriate distributions. These distributions are, however, separation dependent: The logarithm of squared transfer integrals, $\log J^2$ (which is often Gaussian distributed), depends on molecular separation. For DPBIC, the distance dependence of the mean and the standard deviation is shown in Figure 15.4d. In the stochastic model, transfer integrals are then drawn from such distant-dependent distributions.

With all necessary rate ingredients, one can now validate the model, e.g., by evaluating the distribution of rates, see Figure 15.4e, or by directly comparing charge carrier mobilities, as shown in Figure 15.4f. Note that charge transport in systems with large energetic disorder has pronounced finite size-effects [130]. Therefore, similar system sizes should be used to compare stochastic and reference simulations.

Since stochastic models are computationally significantly less demanding, they can serve as an intermediate step between atomistic and macroscopic (drift–diffusion) descriptions, as discussed in the next section.

15.11.2 Parametrization of Gaussian Disorder Models

As mentioned in Section 15.3, macroscopic OLED modeling requires the charge mobility as a function of external field, temperature, and carrier density, $\mu(T, F, \rho)$. Analytic expressions of these dependencies can be provided by the EGDM and ECDM, as discussed in Section 15.7.1. These generic expressions include

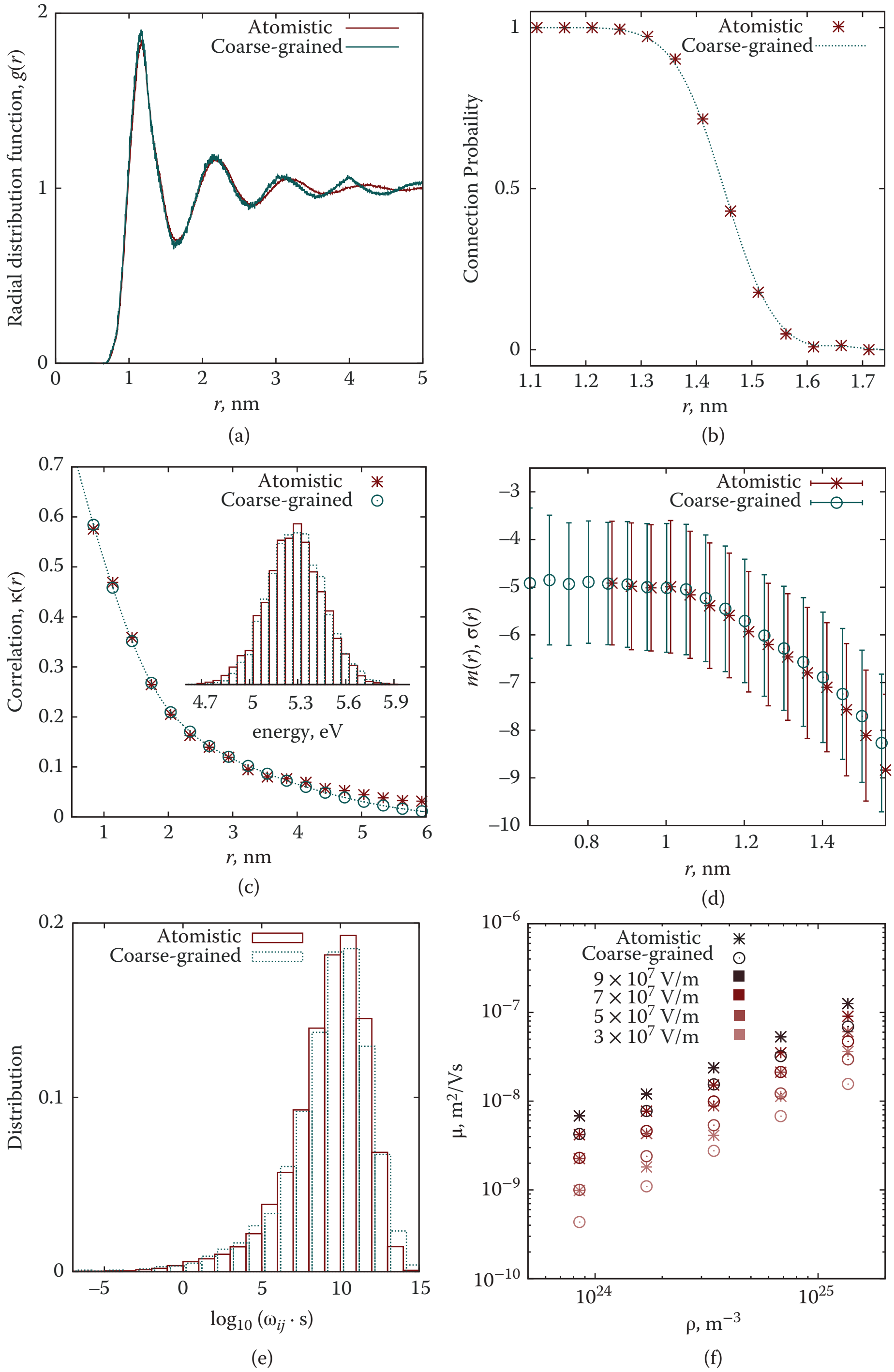


FIGURE 15.4 Comparison of hole states and hole mobilities in the atomistic reference system of DPBIC and in a stochastic model. (a) Radial distribution function, $g(r)$, where r is the center-of-mass distance. (b) Probability for sites to be connected. (c) Spatial correlation function, $\kappa(r)$, as defined in Equation 15.29; inset: site energy distribution with a mean (ionization potential) of 5.28 eV and an energetic disorder of $\sigma = 0.176$ eV. (d) Mean m and width σ of a distribution of the logarithm of squared electronic couplings, $\log_{10}(J^2/\text{eV}^2)$, for molecules at a fixed separation r. (e) Rate distributions. (f) Mobility as a function of hole density at different electric fields. (Reproduced from P. Kordt et al., *Adv. Funct. Mater.*, 25, 1955–1971, 2015, with permission.)

several material-specific parameters. In this section, we describe how to determine these parameters from simulations of small systems.

Both EGDM and ECDM depend parametrically on the lattice constant a, the energetic disorder σ (see Section 15.7), and a prefactor μ_0, which is related to the temperature-dependent mobility at zero field and charge density by $\mu_0(T) = 1.8 \times 10^{-9}\mu_0 \exp(-0.42\hat{\sigma}^2)$ for the EGDM (see Section 15.7.1) or $\mu_0(T) = 1.0 \times 10^{-9}\mu_0 \exp(-0.29\hat{\sigma}^2)$ in the case of ECDM.

In principle, both a and σ can be evaluated in a relatively small system: a as the mean distance between neighboring molecules and σ as the width of the DOS that results from perturbative energy calculations. μ_0 can be extracted from charge transport simulations performed at different temperatures. This approach, however, does not lead to reliable parameterizations [55,114]. Indeed, a multidimensional fit of simulated mobilities to the EGDM or ECDM expressions, for a wide range of temperatures, charge densities, and external fields, yields a very different set of parameters. A comparison of these two approaches for amorphous dicyanovinyl-substituted quaterthiophe (DCV4T) and DPBIC is given in Table 15.1. One can see, e.g., that the EGDM underestimates the energetic disorder, while the ECDM overestimates it. In both cases, spatial site energy correlations are responsible for this discrepancy: EGDM does not include correlations and compensates for higher mobility values by reducing the energetic disorder σ. On the other hand, ECDM overestimates spatial correlations and compensates this by reducing the lattice constant [89]. The discrepancy between microscopic values and fits to EGDM and ECDM teaches us that parameters of these models do not have a clear physical interpretation. Nevertheless, they still provide reasonable parametrizations and can eventually be used in conjunction with drift–diffusion equations; see Section 15.12.1. Figure 15.5 compares EGDM and ECDM fits to microscopic simulations for an amorphous mesophase of DCV4T.

On a technical side, the stochastic models described in Section 15.11.1 become very useful to perform the fits. They help cover the required range of charge carrier densities and reduce finite-size effects. Note, however, that finite-size effects in systems with small charge carrier densities and large disorder are so large that the actual value of mobility is overestimated by several orders of magnitude [130]. In this case, one needs to use the extrapolated mobility values [114].

15.11.3 Tabulated Mobilities

Fitting the results of KMC simulations to the parametrizations provided by the EGDM or ECDM imposes a constraint on the functional form of $\mu(\rho, \vec{F}, T)$. To avoid this, one can tabulate the mobility in a wide range of charge densities, temperatures, and electric fields. This tabulated function can then be used directly in the drift–diffusion equations solver [149]. The tabulation is computationally feasible only with the help of a stochastic model—otherwise it is not possible to reach the necessary system sizes and to span the wide density regime. Before using the tabulated function, it has to be interpolated and smoothed to ensure numeric stability [149]. Figure 15.6 shows the tabulated and smoothed mobility for amorphous DPBIC, which is eventually used to evaluate current–voltage characteristics of a DPBIC film; see Section 15.12.1.

TABLE 15.1 Lattice Spacing, Energetic Disorder, and Mobility at Zero Field and Density Extracted From a Microscopic System and From Fitting Simulated Hole Mobilities to Extend Gaussian Disorder Model (EGDM) and Extended Correlated Disorder Model (ECDM) for Amorphous Phases of (a) dicyanovinyl-substituted quaterthiophe (DCV4T) and (b) DPBIC

(a) DCV4T	a [nm]	σ [eV]	μ_0(300 K) [m^2/Vs]
Microscopic	0.86	0.253	2.0×10^{-21}
EGDM	1.79	0.232	2.1×10^{-21}
ECDM	0.34	0.302	3.3×10^{-22}

(b) DPBIC	a [nm]	σ [eV]	μ_0(300 K) [m^2/Vs]
Microscopic	1.06	0.176	3.4×10^{-12}
EGDM	1.67	0.134	2.1×10^{-11}
ECDM	0.44	0.211	1.8×10^{-13}

Source: DCV4T values: Reprinted from P. Kordt et al., *J. Chem. Theory Comput.*, 10, 2508–2513, 2014. With permission; DPBIC values: Reproduced from P. Kordt et al., *Adv. Funct. Mater.*, 25, 1955–1971, 2015. With permission.

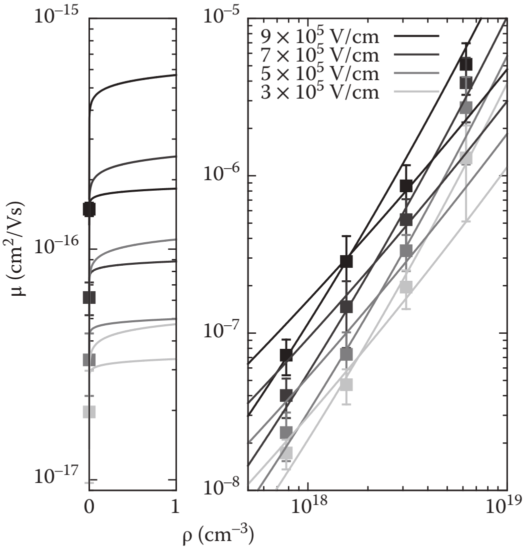


FIGURE 15.5 Parametrization of lattice models from simulated dicyanovinyl-substituted quaterthiophe (DCV4T) hole mobilities, μ, for different hole densities, ρ. Symbols are the simulated values for four different external fields, solid lines are the fit to the extended Gaussian disorder model (EGDM), and dashed lines are the fit to the extended correlated disorder model (ECDM). An extrapolation has been used to obtain non-dispersive values (i.e., without finite-size effects) in the limit of zero density. (Reprinted from P. Kordt et al., *J. Chem. Theory. Comput.*, 10, 2508–2513, 2014. With permission.)

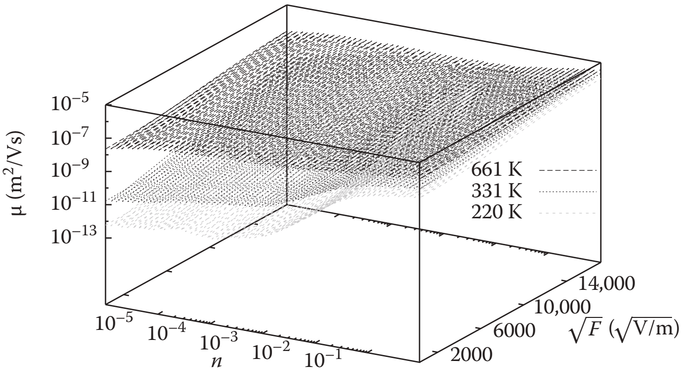


FIGURE 15.6 Hole mobilities for amorphous DPBIC as a function of the charge concentration *n* (number of charge carriers per site) after tabulation from simulations, smoothing, and interpolation. (Reprinted from P. Kordt et al., *Phys. Chem. Chem. Phys.*, 17, 22778–22783, 2015. With permission.) Published by the PCCP Owner Societies.

15.12 Case Studies

So far we have described various methodological developments and simulation approaches, which can be used to simulate multilayered OLED structures. In the following sections, we show how these methods can be used to simulate steady-state current–voltage characteristics of an OLED, perform impedance spectroscopy simulations, estimate OLED efficiency, and study electroluminescence of a white OLED, as well as to gain insight into OLED stability.

15.12.1 Current–Voltage Characteristics

We start by showing how the steady-state current–voltage characteristics of a single layer device can be simulated starting from the chemical structure of an organic semiconductor. As an example, we use a thin layer of DPBIC, a hole-conducting material, which is sandwiched between an ITO and aluminum electrode.

After parametrizing the DPBIC force field [55], amorphous boxes of 4000 DPBIC molecules are simulated by molecular dynamics (MD) simulations, as described in Section 15.10.2. The DOS of an amorphous solid state is then evaluated for holes using the perturbative scheme (see Section 15.7.2), yielding a mean value of 5.28 eV and an energetic disorder of $\sigma = 0.176$ eV. Density functional theory calculations (B3LYP/6-311g(d,p), see also Section 15.8) yield a hole reorganization energy of $\lambda = 0.068$ eV.

The reference system of 4000 molecules is further used to parametrize a stochastic algorithm (see Section 15.11.1) and to generate larger systems of 40,000 sites (details can be found in Ref. [149]). KMC simulations in large systems allow us to tabulate charge mobility as a function of charge density, electric field, and temperature. The interpolated and smoothed tabulated function, depicted in Figure 15.6, is then used to solve the drift–diffusion equations, as described in Section 15.3. For the electrode IPs we use average values of experimental reports: 4.73 eV for ITO [154–157] and 4.16 eV for aluminum [158]. These values, together with the DPBIC IPs, provide the value of the injection barrier, which is required to solve the Poisson Equation 15.3.

Without further microscopic calculations, it is now possible to solve the drift–diffusion equations for film thicknesses of 203, 257, and 314 nm and temperatures of 233, 293, and 313 K, corresponding to different experiments. Figure 15.7 shows a comparison of simulated and experimental current–voltage curves in these situations. The agreement is remarkable for higher temperatures (293 and 313 K). A possible reason for the larger differences at low temperatures are nonequilibrium processes, such as charge relaxation, which are not accounted for in our approach that relies on mobility parametrizations under stationary conditions. Another possibility is that the Marcus rate expression, Equation 15.23, is no longer valid at these temperatures, since it is derived assuming a classical promoting mode and is valid only for high enough temperatures.

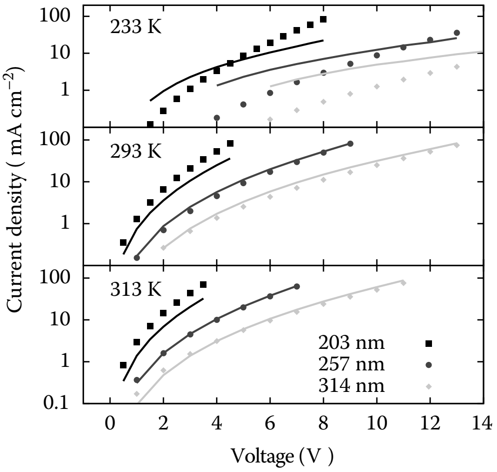


FIGURE 15.7 Current–voltage characteristics for DPBIC films of different thickness sandwiched between an indium tin oxide (ITO) anode and an aluminum cathode, measured at different temperatures. Theoretic predictions (lines) were obtained using parameter-free coupling of microscopic hole mobility data to drift–diffusion equations. Symbols are experimental results. (Reprinted from P. Kordt et al., *Phys. Chem. Chem. Phys.*, 17, 22778–22783, 2015. With permission.)

15.12.2 Impedance Spectroscopy

In addition to studies of steady-state current–voltage characteristics, valuable insight into the functioning of organic devices can be obtained from impedance spectroscopy studies. In particular, impedance spectroscopy can provide information about carrier relaxation in the DOS, can help distinguish between different trapping regimes in organic-semiconductor devices [36], and can be used to determine the width σ of the Gaussian DOS [159].

In impedance spectroscopy a dc bias V is applied over a device and, in addition to that, a small ac component $\Delta V(t) = \Delta V \exp(2\pi i f t)$ is added, where f is the frequency. The impedance $Z = Z' + iZ''$ is defined as the zero-amplitude limit of the ratio of $\Delta V(t)$ and the response $\Delta I \exp\left[2\pi i(f + \phi)t\right]$ in the current, with ϕ a phase difference. Of particular interest is the capacitance–voltage, C–V, characteristic, with the capacitance given by $C = -Z''/2\pi f|Z|^2$.

Applying KMC simulations to extract the small response $\Delta I(t)$ is extremely cumbersome because of the noise present in such simulations. It has been shown, however, that for single-carrier organic devices the current–voltage characteristics obtained by solving the master equation are practically the same as those obtained from KMC simulations [160]. The influence of small perturbations can be rather easily evaluated using the master equation. Within the framework of the time-dependent master equation, the small ac component of the voltage in a single-carrier device leads to a time-dependent probability, $p_i(t)$, of the occupation of a site i by a charge, obeying Equation 15.14:

$$\frac{\mathrm{d}p_i}{\mathrm{d}t} = \sum_{j \neq i} \left[\omega_{ji} p_j (1 - p_i) - \omega_{ij} p_i (1 - p_j)\right] \equiv g_i(\vec{p}), \tag{15.34}$$

where $\vec{p}$ is the vector of occupational probabilities of all sites.

Using a perturbative approach, first the steady-state solution, $\vec{p}_0$, for $\mathrm{d}p_i/\mathrm{d}t = 0$ at the applied static voltage, V, has to be evaluated. The procedure for doing so has been described in Ref. [99]. Sheets of sites representing the electrodes are introduced at either side of a simulation box representing the device. An additional small-amplitude ac voltage with frequency f induces a small change, $\vec{\Delta p}$. Linearizing, we write $\vec{p}(t) \approx \vec{p}_0 + \exp\left(2\pi i f t\right) \vec{\Delta p}$ and $\vec{g}(\vec{p}) \approx \vec{g}(\vec{p}_0) + \exp(2\pi i f t) \left[\Delta V \partial \vec{g}/\partial V + \hat{J}\vec{\Delta p}\right]$, with the matrix elements of the Jacobian, $\hat{J}$, given by $J_{ij} = \partial g_i/\partial p_j\big|_{\vec{p}_0}$. Substituting these expressions into Equation 15.34 and linearizing leads to the equation

$$(2\pi i f \hat{I} + \hat{J})\vec{\Delta p} = -\Delta V \frac{\partial \vec{g}}{\partial V}, \tag{15.35}$$

with $\hat{I}$ denoting the identity matrix. Equation 15.35 can be solved for $\vec{\Delta p}$ with standard techniques, and from this the current, ΔI, and the capacitance, C, are readily obtained.

As an example, we consider two hole-only devices with the structure glass, ITO (100 nm), poly(3,4-ethylenedioxythiophene) polystyrene sulfonate (PEDOT:PSS) (100 nm), light-emitting polymer (LEP), Pd (100 nm). The LEP consists of polyfluorene with 7.5 mol% copolymerized triarylamine units for hole transport; see the Figure 15.8a inset. The LEP-layer thicknesses are $L = 97$ and 121 nm for the two devices and their areas are $A = 9 \times 10^{-6}\ \mathrm{m}^2$. According to EGDM modeling studies of the current density–voltage, J–V, characteristics of these devices [161,162] no injection barrier is present at the anode (PEDOT:PSS) and injection barriers of 1.65 and 1.90 eV are present at the cathode (Pd) for the $L = 97$ and $L = 121$ nm device, respectively. These modeling studies gave best fits for the J–V characteristics with $\sigma = 0.13$ eV.

We apply the above method to calculate the C–V characteristics of these two devices. As in the EGDM studies [161,162] we assume an uncorrelated Gaussian DOS with standard deviation σ. The use of Marcus rates would require knowledge of the reorganization energy λ_{ij}, which is not available here. We therefore assume Miller–Abrahams nearest-neighbor hopping with rates given by Equation 15.22. In the absence

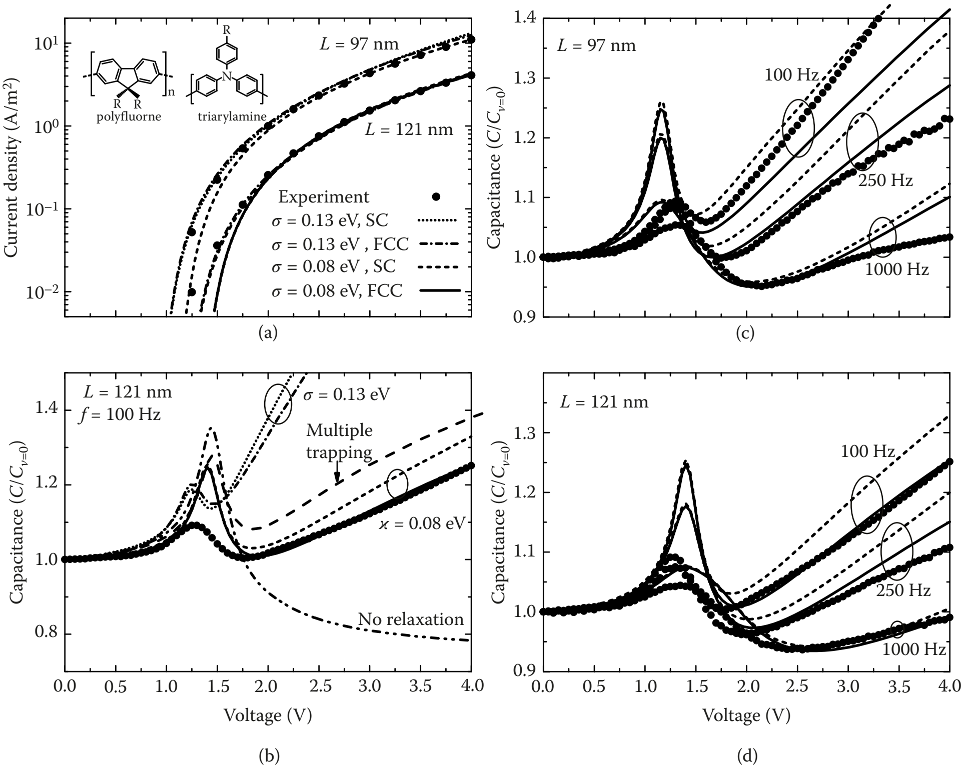


FIGURE 15.8 (a) Current density–voltage, J–V, characteristics of the devices with organic layer thicknesses $L = 97$ and $L = 121$ nm at $T = 295$ K. Dots: measurements [163]. Curves: solutions of the steady-state master equation with $\sigma = 0.13$ and 0.08 eV for simple cubic (SC) and face-centered cubic (FCC) lattices. Inset: the used hole-transporting copolymer. (b) Capacitance, C, normalized to its value at $V = 0$, as a function of V at a frequency $f = 100$ Hz and temperature $T = 295$ K for the $L = 121$ nm device. Dots: measurements [164]. Dash–dot–dotted curve: drift–diffusion calculation with an extended Gaussian disorder model (EGDM) mobility function for $\sigma = 0.13$ eV, neglecting relaxation. Long-dashed curve: multiple-trapping result, which includes relaxation [164]. Other curves: Solutions of the time-dependent master equation for $\sigma = 0.13$ and 0.08 eV and for SC and FCC lattices. (c) and (d) C–V characteristics at different frequencies for the $L = 97$ nm (c) and $L = 121$ (d) device. Dots: Measurements [164]. Curves: Solutions of the time-dependent master equation. (Reprinted with permission from M. Mesta, J. Cottaar, R. Coehoorn, and P. A. Bobbert, Study of charge-carrier relaxation in a disordered organic semiconductor by simulating impedance spectroscopy. *Appl. Phys. Lett.*, 104(21), 213301, 2014. Copyright [2014], American Institute of Physics.)

of any morphological information we assume a regular lattice of hole-transporting sites. To investigate a possible influence of morphology we investigate simple cubic (SC) as well as face-centered cubic (FCC) lattices with a lattice constant $a = 1.19$ nm for the SC lattice and $a = 1.88$ nm for the FCC lattice, in accordance with the known density 1.8×10^{26} m^{-3} of hole-transporting units. The simulation boxes have dimensions $L \times L_y \times L_z$, with $L_y = L_z = 50a$ and periodic boundary conditions in the y- and z-directions, yielding a sufficient lateral averaging. For further details we refer to Ref. [159].

As expected from the EGDM modeling [161,162] it can be observed in Figure 15.8a that the experimental J–V characteristics (dots) at room temperature ($T = 295$ K) of the two devices are very well described by the solution of the steady-state master equation for $\sigma = 0.13$ eV, both for the SC (dotted curve) and the FCC (dash-dotted curve) lattice (a prefactor in the hopping rates was adjusted in both cases to obtain an optimal fit). However, this is not at all true for the C–V characteristics. It is seen in Figure 15.8b that for the $L = 121$-nm device solving the time-dependent master equation for $\sigma = 0.13$ eV at a frequency $f = 100$ Hz yields results that deviate strongly from the experimental C–V characteristic. The fact that the master-equation results for the SC and FCC lattice are quite comparable shows that this deviation is probably not due to a morphological issue.

In order to understand the problem better we first distinguish the different regimes in the C–V characteristics. (1) At low voltage all characteristics converge to the geometrical capacitance, because almost no carriers are present in the device. (2) With increasing voltage, a sheet of holes builds up by diffusion from the anode, but these cannot yet move to the cathode because the electric field is still directed from cathode to anode. As a result, the effective thickness of the device decreases and the capacitance rises. (3) When approaching the built-in voltage V_{bi} these holes start to move to the cathode, leading to a decrease of the capacitance. The result is a peak in the C–V curve before V_{bi} is reached [165]. (4) In the regime beyond V_{bi} the C–V curve rises again. Here relaxation effects play a dominant role and therefore this is the regime we want to focus on.

In order to identify the effects of relaxation we display in Figure 15.8b (dash–dot–dotted curve) the C–V characteristic obtained by solving the time-dependent drift–diffusion equation with the EGDM mobility function corresponding to $\sigma = 0.13$ eV. In this case, the local mobility $\mu(x; \rho, F, T)$ depends on the instantaneous local charge density ρ and electric field F, and therefore contains no relaxation effects. It is seen that without relaxation effects the capacitance decreases after V_{bi} to a value that is even smaller than the geometrical capacitance. The long-dashed curve is the result of a multiple-trapping model for relaxation [164]. With a fitted conduction-level energy $E_{\mathrm{c}} = -0.75\sigma$, this model leads to a fair agreement with experiment.

Since solving the time-dependent master equation for $\sigma = 0.13$ eV apparently overestimates relaxation effects and since such effects decrease with decreasing σ we solved the time-dependent master equation for lower values of σ. With a value of $\sigma = 0.08$ eV we find a very satisfactory agreement with the experimental C–V characteristics, not only for the device and frequency considered in Figure 15.8b, but also for both devices and all considered frequencies; see Figure 15.8c and 15.8d. The only clear disagreement is in the peak, which is more pronounced in the calculations than in the experiment. This may be partially explained by lateral variations in V_{bi} of the devices [164]. The dashed (SC) and full (FCC) curves in Figure 15.8a are the corresponding J–V curves obtained by solving the steady-state master equation. It is observed that for high voltages the experimental J–V curves are very well described, but significant deviations occur at low voltages around V_{bi}.

The analysis brings up the question why there is such an apparent discrepancy between the description of steady-state and time-dependent charge transport. A possible explanation is that in steady-state transport the low-energy tail of the DOS is important, represented by a relatively large σ, while in time-dependent transport relaxing carriers probe a larger part of the DOS, represented by a smaller σ. This would mean that the shape of the DOS is more complicated than a single Gaussian. It would also explain the difference in the description of the J–V curves in Figure 15.8c. At low voltage, when carriers only occupy the low-energy tail of the DOS, $\sigma = 0.13$ eV gives a better description, while at higher voltage, when the DOS is filled up further, $\sigma = 0.08$ eV provides an excellent description, which is even slightly better than with $\sigma = 0.13$ eV. A value of $\sigma = 0.13$ eV at low voltage could also partially explain the lower peak in the C–V curve as compared to the calculations with $\sigma = 0.08$ eV in Figure 15.8b through d. We note that the position of the peak could be improved by adapting the used built-in voltages V_{bi}. These voltages were obtained from an EGDM fit of the J–V characteristics with $\sigma = 0.13$ eV [161,162], but should be optimized again in a fit with $\sigma = 0.08$ eV. In addition, the EGDM neglects spatial correlations of site energies. This can lead to discrepancies when analyzing an experimental system with correlations in terms of a model without correlations [89].

The present conclusion that $\sigma = 0.08$ eV should be used to describe carrier relaxation in the considered devices is fully in agreement with the conclusion that dark-injection experiments on the same devices, which also probe carrier relaxation, can be described by solving a time-dependent master equation with the same value of σ [163].

15.12.3 Efficiency

In OLEDs, electrical power is converted to a radiant flux (radiant energy emitted per unit time), Φ_e. The power efficiency, sometimes called the wall-plug efficiency, is given by

$$\eta_{\text{power}} = \frac{\Phi_e}{IV} = \frac{\int_0^\infty \Phi_{e,\lambda,\text{OLED}}(\lambda)d\lambda}{IV}, \tag{15.36}$$

where I is the current, V is the applied voltage, and $\Phi_{e,\lambda,\text{OLED}}$ is the total optical power that is emitted externally per unit wavelength λ. The power efficiency is generally limited by inevitable Ohmic losses in the electrodes and in the organic charge transport layers, and sometimes also by Ohmic losses due to the presence of internal organic–organic energy barriers outside the emissive layer. When judging the efficiency of the conversion process in the emissive layer, one therefore often focuses on a complementary quantity, the external quantum efficiency η_{EQE} (EQE), that is defined as the total number of externally emitted photons per charge carrier that has passed the device:

$$\eta_{\text{EQE}} = \frac{e}{I}\int_0^\infty \Phi_{e,\lambda,\text{OLED}}(\lambda)\frac{\lambda}{ch}d\lambda, \tag{15.37}$$

where e is the fundamental charge, c is the speed of light, and h is Planck's constant. Due to full or partial internal reflection of light, not all photons that are internally generated will escape from the microcavity that is formed by the OLED layer structure. It is therefore useful to introduce an additional quantity, the internal quantum efficiency (IQE) η_{IQE}, which is defined as the ratio of the total number of photons generated within the device and the number of electrons injected. The IQE is not directly measurable, but may be derived from the EQE using the expression

$$\eta_{\text{IQE}} = \frac{\eta_{\text{EQE}}}{\eta_{\text{out}}}, \tag{15.38}$$

where η_{out} is the light-outcoupling efficiency. For emission from a specific position in a planar OLED microcavity, under a specific angle and for a specific wavelength, the (s and p) polarization-dependent emitted light intensity may be obtained from optical simulations [166–169]. The light-outcoupling efficiency is thus an effective value, which is determined by averaging over the entire emission profile and the entire emission spectrum and which is sensitive to the precise angular dependence of the emission from the dye molecules. Application of advanced emission profile reconstruction techniques [170,171] and a measurement of the emitter orientation distribution [172] are required to determine η_{out} for a specific case with high precision. In the absence of such information, one often assumes that for well-designed phosphorescent OLEDs with a (glass | ITO | organic semiconductor | Al) layer structure and with a random emitter orientation η_{out} is approximately 0.2. Larger values, up to ~0.25–0.30, are possible by optimizing all layer thicknesses [173]. Several methods, including the use of a roughened external glass surface or the use of internal high-refractive index scattering layers, have been developed to enhance the light-outcoupling efficiency to values above 0.5 [174,175].

Recently, much progress has been made in advanced molecular-scale KMC calculations of the IQE [176–178]. We focus in this section on applications of KMC simulations to phosphorescent OLEDs based on a small concentration of metal-organic emitter molecules in a matrix material. In general, the IQE may be expressed as [174]

$$\eta_{\text{IQE}} = \eta_{\text{rec}}\eta_{\text{ST}}q_{\text{eff}}, \tag{15.39}$$

with η_{rec} the recombination efficiency, defined as the fraction of injected charges which contributes to exciton formation, η_{ST} the singlet–triplet factor, defined as the fraction of generated excitons which is quantum-mechanically allowed to decay radiatively, and q_{eff} the effective radiative decay efficiency, defined as the fraction of such excitons which actually decays radiatively. In phosphorescent OLEDs based on heavy metal-organic molecules, strong spin–orbit interaction gives rise to triplet states with some mixed-in singlet-character, so that also triplets are emissive and $\eta_{ST} = 1$. The recombination efficiency can be close to unity by making use of appropriate electron and hole blocking layers. A highly effective radiative decay efficiency may be obtained, first, by using emissive dye materials with a large radiative decay rate, Γ_{rad}, and a small nonradiative decay rate, Γ_{nr}. In the absence of other loss processes, the IQE is then equal to $\Gamma_{rad}/(\Gamma_{rad} + \Gamma_{nr}) \equiv \eta_{PL}$, the photoluminescence (PL) efficiency. Nonradiative decay is a result of the nuclear motion, so that the energy of the molecule in its excitonically excited state can be equienergetic with a highly vibrationally excited excitonic ground state [179]. Second, the matrix material and the adjacent blocking material should have a triplet energy level significantly larger than the dye triplet level so that the triplet excitons stay confined to the dye sites. These design rules are already relevant to the IQE at small current densities. Experimentally, the IQE is found to depend on the current density, J. At large J, η_{IQE} decreases with increasing J. For some devices, η_{IQE} is found to show a broad maximum before the decrease ("roll-off") sets in. A practical measure is the current density J_{90} at which the IQE has decreased to 90% of the maximum value. For efficient phosphorescent OLEDs, maximum reported values of J_{90} are approximately 300 $\mathrm{A/m^2}$ [180]. In commercial white OLEDs for lighting conditions, operated at high luminance levels, the efficiency loss due to roll-off can be indeed of the order of 10%.

Understanding the roll-off is not only important as a first step toward enhancing the efficiency. Loss processes that limit the IQE at high J can also trigger local degradation processes with a certain probability, as will be discussed in Section 15.12.5. By building "virtual OLEDs" in which the interplay of all charge transport and excitonic processes is included mechanistically using KMC simulations, the functioning of OLEDs can be studied with subnanosecond time and molecular-scale spatial resolution. The first demonstration of the feasibility of such an approach was presented by van Eersel et al. [176]. We discuss their simulation results for OLEDs based on the green-emitting metal-organic molecule tris[2-phenylpyridine]iridium ($\mathrm{Ir(ppy)_3}$) and the red phosphorescent dye platinum octaethylporphyrin (PtOEP).

The simulations were based on the three-dimensional (3D) KMC code Bumblebee (http:/simbeyond.com): For a detailed discussion of the model used, we refer to Refs. [176,177,181,182]. Ref. [176] also gives motivations for the parameter values used and provides analyses of the sensitivity of the simulation results to the parameter values. Briefly, the OLEDs were modeled as a collection of molecular sites on a simple cubic lattice. For each type of molecule, the site energies for electrons and holes were taken randomly from a Gaussian DOS with an average energy as given in Figure 15.9a. Charges were assumed to hop with a rate as described within the Miller–Abrahams formalism Equation 15.22. The hopping attempt frequency ω_0 and the wavefunction decay length $\lambda = 1/\gamma$ were taken equal for all pairs of sites. The simulations included the Coulomb interactions between all charge carrier pairs and with image charges in the metallic electrodes. In a natural way, the formation of space-charge layers near the injecting and organic–organic interfaces, and the resulting "band bending," was thus included. Cottaar et al. have demonstrated that in energetically disordered materials, as they are used in OLEDs, explicitly taking the individual 3D Coulomb interactions into account is important for properly treating charge accumulation near internal interfaces [183]. Instantaneous ISC was assumed, so that only triplet excitons were considered. Exciton generation and dissociation were treated in the same way as hops of electrons and holes, but including the triplet exciton binding energy. Radiative and nonradiative triplet exciton decay was included, as well as exciton transfer between the dye molecules, leading to exciton diffusion. The transfer rate was expressed as a sum of Förster-type and Dexter-type contributions, Equations 15.25 and 15.26, where $\tau = (\Gamma_{rad} + \Gamma_{nr})^{-1}$ is the effective decay time, R_F is the Förster radius for diffusion, and k_D is the Dexter prefactor. The two bimolecular loss processes that potentially contribute to the IQE roll-off, triplet–polaron quenching (TPQ) and triplet–triplet annihilation (TTA), were both included in a parameter-free manner, viz. by assuming an infinite (zero) rate when an exciton and a polaron or two excitons, respectively, are present on nearest-neighbor (more distant) sites. Table 15.2 gives an overview of the used parameter values.

Figure 15.10a and b shows a comparison of the calculated and experimental J–V and IQE roll-off curves, respectively, for a temperature of 300 K. The slope of the J–V curves is well described, but the absolute

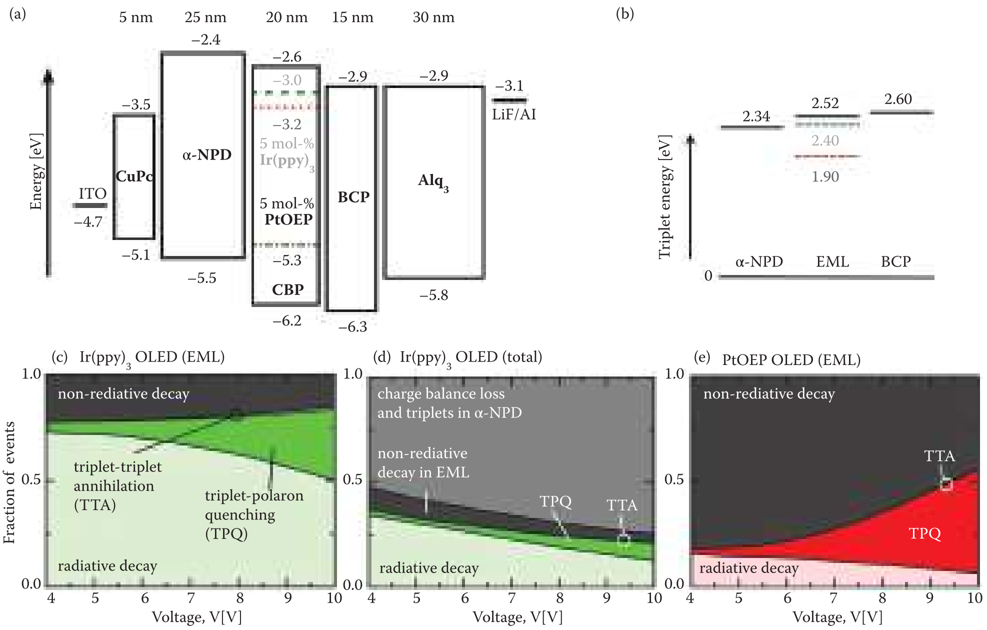


FIGURE 15.9 (a) Energy level and layer structure of the green and red organic light-emitting diodes (OLEDs) studied in Ref. [176]. The phosphorescent emissive layer (EML) is sandwiched in between materials facilitating hole and electron injection, transport and blocking: CuPc (copper phthalocyanine), α-NPD (4,4′-bis[N-(1-naphthyl)-N-phenyl-amino] biphenyl), BCP (2,9-dimethyl, 4,7-diphenyl, 1,10-phenanthroline), and Alq_3 (tris [8-hydroxyquinoline] aluminum). (b) Triplet energies for the materials used in the EML (solid line: CBP; dashed line: $Ir(ppy)_3$; dotted line: PtOEP) and the layers adjacent to the EML. (c)–(e) Contribution of the various exciton decay processes in the EML of the $Ir(ppy)_3$ device (b), in the entire $Ir(ppy)_3$ device (c), and in the EML of the PtOEP device (d). The figures show that even above 6 V only a small fraction of the efficiency loss (less than 2% (0.5%) for the $Ir(ppy)_3$ (PtOEP) devices) is due to triplet–triplet annihilation (TTA). (Reprinted with permission from H. van Eersel, P. A. Bobbert, R. A. J. Janssen, and R. Coehoorn. Monte Carlo study of efficiency roll-off of phosphorescent organic light-emitting diodes: Evidence for dominant role of triplet-polaron quenching. *Appl. Phys. Lett.*, 105(14):143303, 2014. Copyright [2014], American Institute of Physics.)

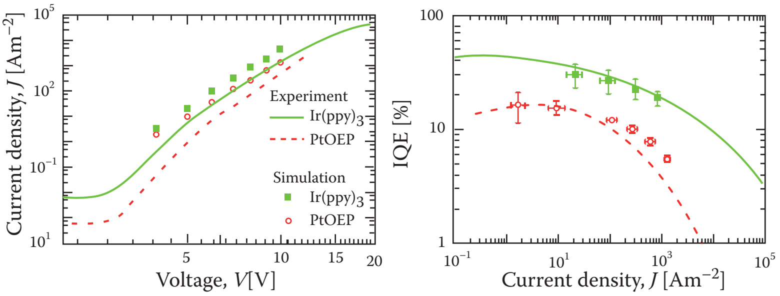
value of the current density is somewhat overestimated. As argued in Ref. [176], this might be related to an underestimation of the HOMO–LUMO gap, which was taken to be equal to the optical gap. This often-used approach neglects the exciton binding energy, which can be around 1 eV [184]. Such a correction would horizontally shift the *J*–*V* curves by 1 eV, giving rise to a significant reduction of the discrepancy. The roll-off curve would not be affected by such a correction. We note that (as mentioned above) the triplet exciton binding energy was included when calculating the rates of exciton generation and dissociation processes. For $Ir(ppy)_3$, the simulation results agree within the error margin with experiment, whereas for PtOEP the roll-off is slightly underestimated at high current densities. A sensitivity analysis was carried out to find out which uncertainties in the choice of parameter values have the largest impact. It was found, e.g., that the *J*–*V* characteristics are quite strongly determined by the energy-level differences at interfaces and between the host and guest states in the emissive layer (EML). The sensitivity to the hopping attempt frequency was found to be relatively small.

The simulations provided detailed views on the cause of the roll-off, as shown in Figure 15.9c through e. In both devices, most of the emission was found to occur near the anode-side of the EML. This may be understood from Figure 15.9a, from which the guest molecules are expected to give rise to stronger hole trapping than electron trapping. As a result, the effective electron mobility is larger than the hole mobility. In the emissive layer, most of the IQE loss was found to be due to TPQ, and only at high voltages

TABLE 15.2 Overview of the Simulation Parameters

Parameter	Description	Value	
Common			
ω_0	Hopping attempt frequency to the first neighbor	3.3×10^{10} s^{-1}	
σ	Width of the electron and hole Gaussian DOS	0.10 eV	
N_t	Site density	1.0×10^{27} m^{-3}	
$\lambda \equiv 1/\gamma$	Wavefunction decay length	0.3 nm	
ε_r	Relative dielectric permittivity	3.5	
σ_T	Width of the triplet exciton DOS	0.10 eV	
$E_{T,b}$	Triplet exciton binding energy	1.0 eV	
k_D	Prefactor for triplet exciton Dexter transfer	1.6×10^{10} s^{-1}	
Material-specific		Ir(ppy)$_3$	PtOEP
R_F	Förster radius for triplet exciton diffusion	1.5 nm	1.5 nm
Γ_{rad}	Radiative decay rate	0.816 μs^{-1}	0.1 μs^{-1}
Γ_{nr}	Nonradiative decay rate	0.249 μs^{-1}	0.525 μs^{-1}

Note: The highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energies are given in Figure 15.9a, and the triplet energies are given in Figure 15.9b.



(a)

(b)

FIGURE 15.10 (a) Simulated and experimental J–V characteristics. (b) Simulated internal quantum energy (IQE) and experimental IQE (assuming 20% outcoupling efficiency) as a function of the current–density. Experimental data from Ref. [185]. (From N. C. Giebink and S. R. Forrest, *Phys. Rev*. B., 77, 235215, 2008; Reprinted with permission from H. van Eersel et al., Monte Carlo study of efficiency roll-off of phosphorescent organic light-emitting diodes: Evidence for dominant role of triplet–polaron quenching. *Appl. Phys. Lett*., 105(14):143303, 2014. Copyright [2014], American Institute of Physics.)

was a small TTA contribution was found (see Figure 15.9c). For the Ir(ppy)$_3$ devices, the overall loss was found to be determined mostly by a nonideal recombination efficiency due to imperfect electron blocking and by triplet exciton diffusion to the α-NPD layer (see Figure 15.9d). The finding of electron loss to the α-NPD layer is consistent with the observation of some blue emission from that layer [185]. Due to subsequent TPQ, triplet transfer to the α-NPD layer also gives rise to a loss. From simulations, improving the devices by introducing perfect electron blocking was predicted to give rise to a 10% increase of the IQE at small voltages, from ~ 34% to ~ 44%. The IQE at small voltages was found to become equal to the PL efficiency assumed (77%) when the triplet transfer to the HTL also was eliminated. For the red devices, at low voltages no significant electron and triplet loss to the α-NPD layer was found, as may be understood from the lower LUMO energy and the smaller triplet energy of PtOEP. The IQE is then close to the PL efficiency assumed (16.4%).

15.12.4 Electroluminescence of a White OLED

The KMC simulations described in Section 15.5 can be employed to model all molecular-scale electronic processes that finally lead to electroluminescence of an OLED: injection, transport, and recombination of electrons and holes as well as diffusion and radiative decay of excitons. We consider here the white multilayer OLED stack of Figure 15.11a, which was studied experimentally and by KMC simulations in Ref. [181]. It concerns a so-called "hybrid" OLED, which combines red and green phosphorescent emission with blue fluorescent emission. Phosphorescent emission can be very efficient because of the harvesting of both singlet and triplet excitons (see Section 15.2.4). However, since stable blue phosphorescent emitters with long-term stability are to date unavailable, many commercial white OLEDs make use of blue fluorescent emission, despite the fact that then only singlet excitons are harvested.

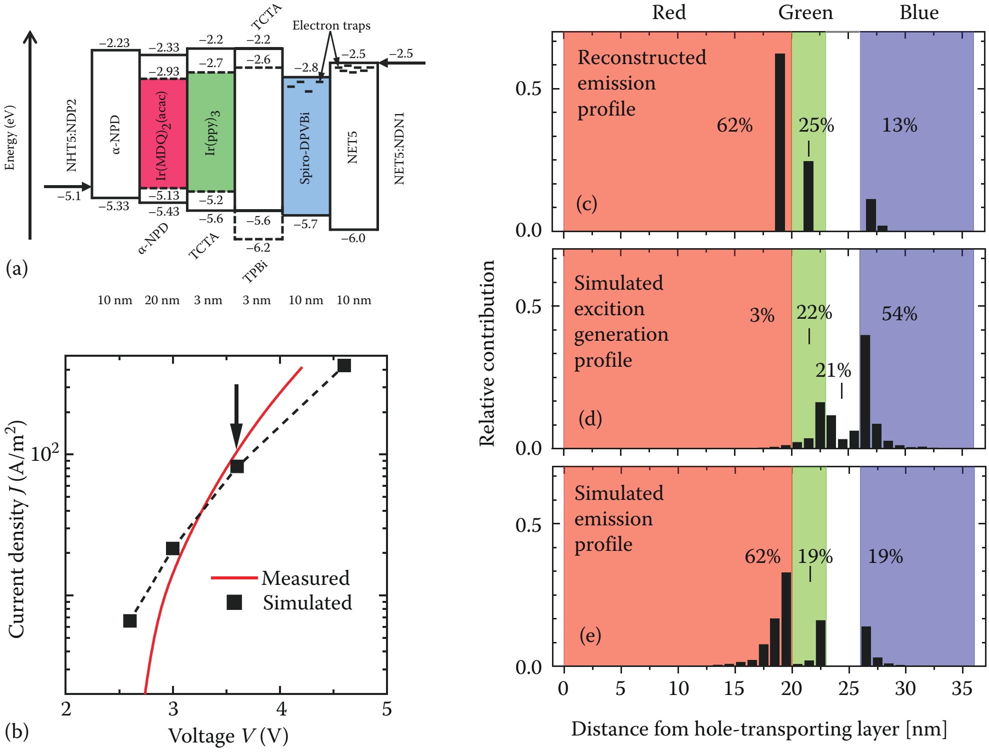


FIGURE 15.11 (a) Stack structure of the investigated white organic light-emitting diode (OLED) with the different layers and their thicknesses. The highest occupied molecular orbital (HOMO) and lowest unoccupied molecular orbital (LUMO) energy levels as used in the kinetic Monte Carlo (KMC) simulations are indicated by solid lines for host materials and dashed lines for guest materials. The horizontal arrows indicate the work functions used for the doped hole- and electron-injecting layers. Electron traps were assumed in the 2,2′,7,7′-tetrakis(2,2-diphenylvinyl)spiro-9,9′-bifluorene (spiro-DPVBi) and NET5 layers with concentrations of 10^{-3} and 5×10^{-3}, and characteristic energies $k_B T_0 = 0.2$ and 0.12 eV, respectively, of their exponential density of states (DOS). (b) Measured and simulated room-temperature current density-voltage, J–V characteristics. All other results presented are for the bias voltage $V = 3.6$ V, indicated by the arrow. The relative error in the simulated current density is estimated to be about 10%. (c) Emission profile reconstructed from the measured angle- and polarization-dependent emission spectrum [181]. The percentages of red, green, and blue emission are indicated. (d) Exciton generation profile as obtained from the KMC simulations. (e) Emission profile as obtained from the KMC simulations, after taking into account exciton diffusion and the radiative decay probabilities of the emitters. The error in the simulated profiles is about 1%. Reprinted by permission from Macmillan Publishers Ltd. *Nat. Mater.* [181] Copyright (2013).

The working principle of the OLED of Figure 15.11a is the following. Holes reach the light-emitting layers of the stack after being injected by an ITO layer into a 4 mol% *p*-doped injection layer of NHT5:NDP2 and transported through a hole-transporting and electron-blocking layer of α-NPD (N,N′-di(naphthalen-1-yl)-N,N′-diphenyl-benzidine). Electrons reach the emitting layers after being injected by an aluminum cathode into a 4 mol% *n*-doped electron-injection of NET5:NDN1 and transported through an electron-transporting and hole-blocking layer of NET5 (materials supplied by Novaled). Blue light is generated in a 10-nm-thick fluorescent layer of spiro-DPVBi (2,2′,7,7′-tetrakis(2,2-diphenylvinyl)spiro-9,9′-bi fluorene) adjacent to the NET5 layer. Green light is generated in a 3-nm-thick layer of TCTA (4,4′,4″-tris(N-carbazoyl)-triphenylamine) doped with 8 mol% of the green phosphorescent dye Ir(ppy)$_3$ (fac-tris(2-phenylpyridyl)iridium). Red light is generated in a 20-nm-thick α-NPD layer doped with 5 mol% of the red phosphorescent dye Ir(MDQ)$_2$(acac) ((acelylacetonate)bis(2-methyldibenzo[f,h]quinoxinalate)iridium).

The green phosphorescent layer is separated from the blue fluorescent layer by a thin (3-nm) interlayer consisting of a mixture of the hole-transporter TCTA with 33 mol% of the electron-transporter TPBi (1,3,5-tris(N-phenylbenzimidazol-2-yl)benzene). This interlayer has several purposes [186]. It should block the transfer of singlet excitons from the blue to the green layer and of triplet excitons from the green to the blue layer (spiro-DPVBi has a triplet energy lower than that of Ir(ppy)$_3$). Also, this interlayer should allow the passage of electrons from the blue to the green layer (by the TPBi) and of holes from the green to the blue layer (by the TCTA). The red phosphorescent layer is purposely in direct contact with the green phosphorescent layer, allowing triplet excitons formed on the phosphorescent dye in the green layer to diffuse to the phosphorescent dye in the red layer. This diffusion is an important process in establishing the right color balance.

The measured room-temperature current density–voltage characteristic of the OLED is shown in Figure 15.11b. The arrow indicates the bias voltage of 3.6 V for most of the reported results. The CIE 1931 color point of the perpendicularly emitted light at this bias was measured to be $[x, y] = [0.47, 0.45]$, which corresponds to warm-white emission. The EQE, i.e., the fraction of emitted photons per injected electron–hole pair, is measured to be $5 \pm 1\%$ [181]. Figure 15.11c shows the color-resolved emission profile, which was reconstructed with a precision on the order of a nanometer from the measured angle- and polarization-dependent emission spectra [170,181]. It is observed that in the blue layer, the emission occurs close to the interface with the interlayer and in the red layer close to the interface with the green layer.

The KMC simulations of the charge dynamics were carried out in the following way. Point sites are arranged on a cubic lattice with a lattice constant $a = 1$ nm, the typical intermolecular distance of the used molecular semiconductors represent the molecules in the stack. A simulation box of $50 \times 50 \times 56$ sites with periodic boundary conditions in the lateral (x- and y-) directions was used, which turned out to yield sufficiently accurate results. Since charge transport in various small-molecule materials was found to be described well by the ECDM [161,191], correlated disorder caused by random dipoles was assumed for the electron and hole energies. The energetic disorder was taken to be $\sigma = 0.1$ eV, corresponding to the value found for hole transport in α-NPD [161]. With this value the charge transport in all materials in the stack is expected to be reasonably described. Red and green emitting guests were introduced according to the known concentration of the emitters, with appropriately adapted energy levels. Electron traps were introduced in the layers in which electron transport is important: the blue fluorescent layer and the ETL. They were modeled with an exponential DOS of trap energies [192–194] with a concentration c_{trap} and a characteristic energy $k_B T_0$.

Nearest-neighbor hopping of charges on the lattice using the Miller–Abrahams rate (see Equation 15.22) was assumed. The energy differences in the hopping rates contain, apart from the random site energies, an electrostatic contribution due to the bias applied to the OLED and the Coulomb energy due to all present charges. The doped injection layers were treated as metallic, injecting and collecting charges with an energy according to their work function, indicated by the arrows in Figure 15.11a. Exciton generation was assumed to occur by hopping of an electron to a site where a hole resides, or vice versa, and was assumed to be always an energetically downward process.

Table 15.3 gives the parameters of the stack materials used in the KMC simulations of the charge dynamics. The parameters were determined from charge-transport and spectroscopic studies of the various materials [181]. Exciton diffusion within the green and red layers and from the green to the red layer was included in the simulations. Since the red and green emitters trap electrons as well as holes (see the energy level scheme in Figure 15.11a), almost all excitons in the red and green layers are generated on the emitters. The diffusion of excitons among the emitters was described by Förster transfer (see Equation 15.25), made possible by the spin–singlet character that is mixed into the exciton wave function by the spin-orbit coupling of the heavy iridium atoms. Apart from being transferred, excitons can decay radiatively with a rate $\Gamma_{\mathrm{rad},i} = 1/\tau_{\mathrm{r},i}$ or nonradiatively with a rate $\Gamma_{\mathrm{nr},i} = 1/\tau_{\mathrm{nr},i}$. These rates are related to the radiative decay probabilities η_{r} by $\eta_{\mathrm{r},i} = \Gamma_{\mathrm{rad},i}/(\Gamma_{\mathrm{rad},i} + \Gamma_{\mathrm{nr},i})$. We took $\eta_{\mathrm{r}} = 0.84$ and 0.76 for the red and green phosphorescent emitters in their respective hosts [169]. Other parameters used in the exciton dynamics are given in Table 15.4. Only exciton transfer from green to red was taken into account. Transfer from red to green should be negligible.

TABLE 15.3 Highest Occupied Molecular Orbital (HOMO) and Lowest Unoccupied Molecular Orbital (LUMO) Energies, Room Temperature Hole- and Electron-Mobilities $\mu_{0,\mathrm{h}}$ and $\mu_{0,\mathrm{e}}$ at Low Field and Low Carrier Density, Electron-Trap Concentration c_{trap}, and Trap Temperature T_0 of the Exponential Trap Density of States (DOS) in the Different Layers of the Stack, as Used in the Kinetic Monte Carlo (KMC) Simulations

Material	E_{HOMO} (eV)	E_{LUMO} (eV)	$\mu_{0,\mathrm{h}}$ (m^2/Vs)	$\mu_{0,\mathrm{e}}$ (m^2/Vs)	c_{trap}	T_0 (K)
NHT5:NDP2	−5.10					
α-NPD	−5.43	−2.33	6×10^{-9}	6×10^{-10}		
Ir(MDQ)$_2$(acac)	−5.13	−2.93	6×10^{-9}	6×10^{-10}		
TCTA	−5.60	−2.20	2×10^{-8}	2×10^{-9}		
Ir(ppy)$_3$	−5.20	−2.70	2×10^{-8}	2×10^{-9}		
TCTA	−5.60	−2.20	2×10^{-8}	2×10^{-9}		
TPBi	−6.20	−2.60	2×10^{-8}	2×10^{-9}		
Spiro-DPVBi	−5.70	−2.80	6×10^{-9}	8×10^{-9}	0.001	2350
NET5	−6.00	−2.50	1.5×10^{-11}	1.5×10^{-10}	0.005	1400
NET5:NDN1		−2.50				

Source: M. Mesta et al., *Nat. Mater.*, 12, 652–658, 2013.

Note: Spiro-DPVBi, 2,2′,7,7′-tetrakis(2,2-diphenylvinyl)Spiro-9,9′-bifluorene; TCTA, 4,4′,4″-tris(N-carbazoyl)-triphenylamine; α-NPD, N,N′-di(naphthalen-1-yl)-N,N′-diphenyl-benzidine; MDQ, acelylacetonate)bis(2-methyldibenzo[f,h]quinoxinalate)iridium; TPBI, 1,3,5-tris(N-phenylbenzimidazol-2-yl)benzene. NHT5:NDP2 and NET5:NDN1 are doped hole and electron transporter materials supplied by Novaled.

TABLE 15.4 Radiative and Nonradiative Exciton Decay Rates Γ_{rad} and Γ_{nr}, Förster Radii R_F for Exciton Transfer between Phosphorescent Emitter Molecules and Triplet Energies E_{T}

Material	Γ_{rad} (μs^{-1})	Γ_{nr} (μs^{-1})	R_{F} (nm)	E_{T} (eV)
Ir(MDQ)$_2$(acac)	0.588	0.112	1.5	2.0
Ir(ppy)$_3$	0.816	0.249	1.5	2.4

Note: For the green emitter Ir(ppy)$_3$ Γ_{rad} and Γ_{nr} were taken from Ref. [187]. For the red emitter Ir(MDQ)$_2$(acac), Γ_{rad} was obtained from Ref. [188], and Γ_{nr} from η_{r} is given in Ref. [169] and the relation $\eta_{\mathrm{r}} = \Gamma_{\mathrm{rad}}/(\Gamma_{\mathrm{rad}} + \Gamma_{\mathrm{nr}})$. The value $R_F = 1.5$ nm for transfer between equal emitter molecules is a typical value given in Ref. [189]. According to the estimate in Ref. [190] we took $R_{F,\mathrm{GR}} = 3.5$ nm for the transfer from a green to a red emitter. The triplet energies were taken from Ref. [186].

For each exciton generated in the red or green layer a separate simulation of its dye-to-dye diffusion and final radiative or nonradiative decay was performed. This diffusion was assumed to proceed independently from all other processes, which means that exciton quenching processes were neglected. Excitons generated on host sites in the red and green layers (a small fraction) were assumed to transfer instantaneously to an emitter in their neighborhood. Diffusion of excitons generated in the blue layer was not accounted for because the diffusion length of these excitons is short and because their transfer to the green layer is blocked by the interlayer. We assumed that in the blue fluorescent layer singlet and triplet excitons are generated in a quantum-statistical ratio of 1:3 and that the triplet excitons are lost. For the radiative decay probability of singlet excitons in the blue layer we took $\eta_r = 0.35$ [195]. Excitons generated in the interlayer were assumed to be lost by nonradiative decay or emission outside the visible spectrum.

The J–V characteristic following from the KMC simulations is given in Figure 15.11b. Considering the various simplifications that were made, the agreement with the measured characteristic is fair. The underestimation of J at high voltage could be due to a heating effect, while the overestimation of J at low voltage could result from a systematic underestimation of the LUMO energies due to neglect of the exciton binding energies. At the bias voltage of 3.6 V for most of the reported results, the experimental and simulated current density agree quite well.

Figure 15.11d and e present the simulated exciton generation profile and emission profile, respectively, at 3.6 V. The effect of exciton transfer from green to red is clearly observable and is very important for the color balance of this OLED. The simulated emission profile is in fair agreement with the reconstructed emission profile from Figure 15.11c. The broadening of the simulated emission profile over a few nanometers in the red layer found in the simulations is not seen in the reconstructed emission profile but this could be due to the limited resolution of the reconstruction procedure [170,181]. The total percentages of emission in the red, green, and blue agree quite well with the reconstructed emission profile.

The results presented in this section show that KMC simulations of all molecular-scale electronic processes leading to electroluminescence in quite complicated multilayer OLEDs with commercial relevance are feasible. The parameters in the present study were almost all obtained from experimental studies, but there is no obstacle for obtaining these from first-principles computational studies. Complete *in silico* studies of the functioning of commercial OLEDs therefore seem to be within reach.

15.12.5 Degradation

Developing an improved understanding of the mechanisms that limit the operational lifetime of OLEDs is of key importance toward the further adoption of OLED technology for display, lighting, and signage applications. Experimentally, given fixed current density and ambient temperature conditions, the luminance is often observed to decrease with time in an approximately exponential or stretched-exponential manner. For commercial white OLEDs for lighting applications, the time at which the luminance has dropped to 70% of the initial value (the so-called LT_{70} lifetime) can today be as large as 10,000 hours or more at a luminance of 8000 cd/m^2. However, long lifetimes are in practice often realized by making a trade-off with device efficiency (e.g., when using hybrid OLEDs, see Section 15.12.4) or production cost and ease of manufacturing (e.g., when using multiply vertically stacked OLEDs). Examples of other aspects of OLED reliability during prolonged operation are the voltage stability (at a fixed current density), the color point stability, the stability at high ambient temperatures, and the stability under prolonged exposure to the ambient atmosphere [196,197]. The latter issue, leading to so-called black-spot formation, has been mitigated by the development of improved encapsulation technologies [198]. Recently, an excellent overview of the degradation mechanisms and reactions in OLEDs has been given by Scholz et al. [199].

In this section, we focus on the use of molecular-scale OLED device modeling as a means to elucidate the role of various possible intrinsic degradation processes, i.e., processes that are caused by excitons (including those due to absorption of internally emitted photons), charges, and fields in the opto-electronically active organic semiconducting layer. When setting up an extension of the KMC simulations described in Sections 15.5 and 15.12.3, in order to include degradation, first an inventory should be made of the processes which for a specific system are expected to give rise to degradation. Formally, a distinction should be made between (1) monomolecular degradation processes, which occur when the state of a single-specific molecule is modified, e.g., due to the presence of a polaron or an exciton or due to a local electric field, and (2) bimolecular degradation processes, which occur upon an interaction between charges and/or excitons on two different molecules, e.g., exciton–exciton annihilation or exciton–polaron quenching.

Including these two types of degradation processes requires a rate and a probability, respectively, as well as a description of the resulting changes of the KMC parameter values. Monomolecular degradation has been observed, e.g., in OLEDs containing the electron transport material Alq_3, in which the presence of holes gives rise to a reduction of the emission by the formation of fluorescence quenchers [200]. Degradation due to the presence of singlet excitons has been reported for 4,4′-bis(N-carbazolyl)biphenyl (CBP), used as a matrix material in the EML of green-emitting phosphorescent OLEDs [201]. The degradation products can act as nonradiative recombination centers and trap states. Degradation due to singlet excitons has also been reported for OLEDs based on the fluorescent emitter material spiro-DPVBi [202]. An example of a bimolecular process which can be accompanied by degradation is triplet–polaron quenching (TPQ). When TPQ is due to excitation of the polaron by the triplet exciton, followed by nonradiative decay of the excited polaron due to internal conversion, the locally dissipated energy may, with a certain probability, lead to a chemical change of the molecule. Such a mechanism was found for a blue phosphorescent OLED using 4,4′-bis(3-methylcarbazol-9-yl)-2,2′-biphenyl (mCBP) as the host material in the EML [203]. The mCBP defect sites were argued to act as deep charge traps and the dissociation products were argued to damage the guest (emitter) so that it becomes a nonradiative center and a luminescence quencher.

A KMC degradation study starts in general with running a simulation without degradation until dynamic equilibrium is achieved under the operational conditions (voltage, temperature) of interest. Subsequently, the simulations are continued while degradation is switched on. Monomolecular degradation due to a charge or an exciton on a sensitive site is included as a new possible process which competes with all other possible processes. Degradation that accompanies a bimolecular process such as TPQ is included by branching the end result of that process, so that a defect site is formed with a certain probability. We note that in KMC simulations only the primary event needs to be described. All subsequent effects (e.g., a shift of the recombination zone to a less favorable position due to a changed mobility balance, or an efficiency loss due to TPQ at charges residing on defect molecules which act as traps [204]) will follow "automatically" from the simulations. KMC lifetime simulations are necessarily strongly accelerated, as practically realistic simulated times are usually at most of the order of 1 ms, 6–11 orders of magnitude smaller than actual lifetimes in the range of 1–100,000 hours. This can be accomplished by assuming an enhanced value of the degradation rate (for monomolecular degradation) or the degradation probability (for degradation accompanying a bimolecular process). As in the case of experimental OLED lifetime tests, it is also possible to accelerate the simulations by carrying them out for high current densities and/or elevated temperatures, followed by extrapolation to application-relevant operational conditions.

A first demonstration of the feasibility of 3D KMC OLED lifetime simulations was presented in Ref. [177]. The simulations were carried out for a symmetric OLED with an energy level structure as shown in Figure 15.12a, for a temperature of 300 K. The mixed-matrix emissive layer contains equal concentrations of the HTL and ETL material as a host, and 4 mol% of emitter molecules (guest). The hole and electron blocking is excellent, so that the recombination efficiency is 100%. The simulation parameters are as given in Table 15.2, with the following exceptions: $\varepsilon_r = 3$, $\sigma_T = 0$ eV (no triplet energy disorder), and $\Gamma_{rad} = 0.544\ \mu s^{-1}$ and $\Gamma_{nr} = 0.181\ \mu s^{-1}$ (values typical for the orange-red emitter $Ir(MDQ)_2(acac)$ in an α-NPD matrix [205]). As in Section 15.12.3, TPQ and TTA were treated in a parameter-free manner as instantaneous nearest-neighbor processes.

Van Eersel et al. [176] showed that for these symmetric devices and with a dye trap depth $\Delta = 0.2 - 0.3$ eV the emission profile at small voltages is quite uniform across the emissive layer. This is illustrated by Figure 15.12b, which shows the emission profile at 3 V for the case studied ($\Delta = 0.2$ eV). On the one hand,

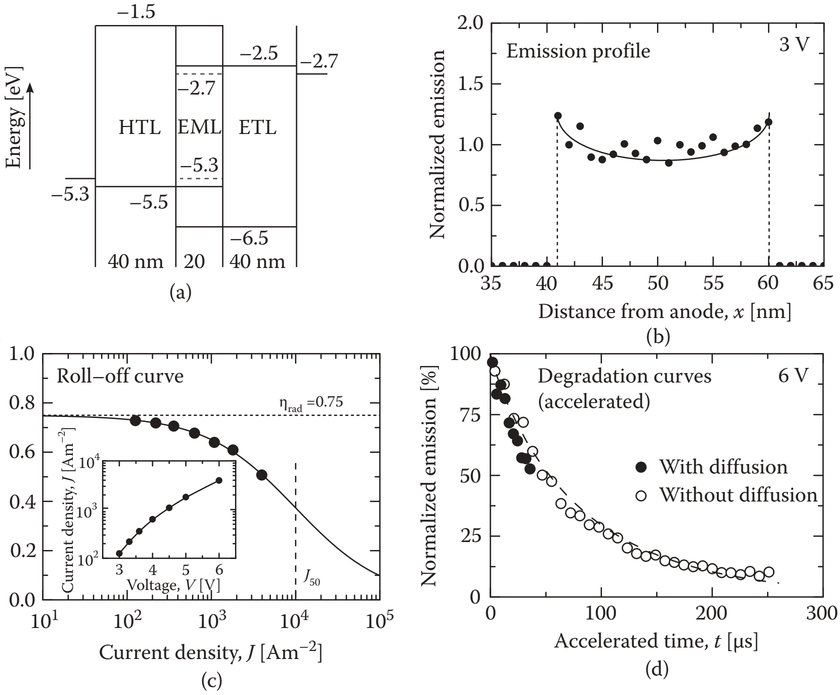


FIGURE 15.12 (a) Organic layer and energy level structure of the organic light-emitting diodes (OLEDs) considered in the kinetic Monte Carlo (KMC) degradation simulation studies. Dashed lines: Emitter energy levels (4 mol%). (b) Emission profile at 3 V. The full curve is a guide-to-the-eye. (c) Current density dependence of the internal quantum efficiency. The full curve is a fit using Equation 15.41 with m = 0.80 and J_{50} = 10 kA/m^2. The vertical dashed line indicates the current density at which the IQE has dropped to 50% (J_{50}) of the value at small current densities. (d) Time-dependence of the normalized emission at 6 V as obtained from a KMC degradation study which assumes that (1) the degradation occurs on triplet-excited dye sites upon triplet-polaron quenching processes (TPQ-t, see the main text) resulting from the displacement of a polaron on a neighbor site to the dye site, and which assumes that (2) each quenching is followed by a degradation process upon which the involved dye molecule becomes nonemissive. The simulations were carried out with and without exciton diffusion (closed and open symbols, respectively). The dashed curve is a stretched-exponential fit to the simulation results without diffusion (see the main text). (Reproduced from R. Coehoorn et al., *Adv. Funct. Mater.*, 25, 2024–2037, 2015. With permission.)

this choice avoids a large Ohmic loss due to deep trapping and a large overvoltage due to the enhanced built-in voltage (for large Δ). On the other hand, it also avoids a large loss due to strong roll-off caused by emission from thin zones near the blocking layer interfaces (for small Δ). The optimum value of Δ will depend on the application goal (see Ref. [176]) and is sensitive to the detailed mechanisms of the TPQ and TTA processes. Figure 15.12c shows the calculated IQE roll-off curve before degradation. The J–V curve is given in the inset. At small voltages, the IQE is equal to the assumed radiative decay (PL) efficiency ($\eta_{rad} = \Gamma_{rad}/(\Gamma_{rad} + \Gamma_{nr}) = 0.75$). The J_{90} current density, defined in Section 15.12.3, is approximately 640 A/m^2. This is larger than the largest experimental value obtained so far (~ 300 A/m^2, see Section 15.12.3). The full curve gives a fit through the data points, discussed below (see Equation 15.41). For the nearest-neighbor TPQ and TTA mechanisms considered, the roll-off is almost completely due to TPQ.

Figure 15.12d shows the dependence of the normalized emission on the simulated time, at constant voltage conditions (6 V), for a degradation scenario in which, upon a TPQ process, the polaron involved is displaced to the site at which the triplet exciton resides (a "TPQ-t process," see Figure 2 in Ref. [177]), which then becomes nonemissive with a degradation probability p_{degr} = 1. The simulations thus employ the largest possible acceleration factor. All other parameters are kept identical. The current density was found to remain essentially unchanged during the degradation process. Furthermore, it was found that choosing a smaller degradation probability does not significantly change the results, apart from changing the time scale. That indicates that the lifetime is still much larger than all other relevant time scales. Simulations including exciton diffusion are computationally more expensive, and were stopped when a 50% emission reduction had been obtained. Simulations without diffusion were continued until a reduction to only about 10% of the initial emission was reached. An approximate description of the decay is given by the stretched exponential curve, shown in the figure, with the form $I(t) = I(0)\exp[-(t/\tau_{\mathrm{sim,acc}})^{\beta}]$, with a simulated accelerated (1/e) lifetime $\tau_{\mathrm{sim,acc}} \cong 80$ μs and a stretching exponent $\beta = 0.91$. The actual (1/e) lifetime as predicted from the simulations is given by

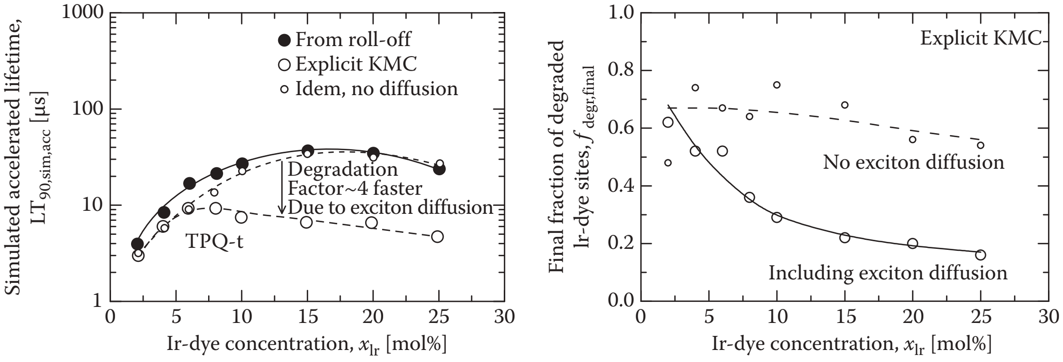
$$\tau_{\mathrm{sim}} = \frac{\tau_{\mathrm{sim,acc}}}{p_{\mathrm{degr}}}. \tag{15.40}$$

Conversely, using Equation 15.40 the value of p_{degr} could be deduced from a degradation simulation and a measurement of the lifetime. Such an analysis led in Ref. [177] to an estimated order-of-magnitude value of $p_{\mathrm{degr}} \sim 10^{-8}$ when assuming that the simulations discussed above are relevant to state-of-the-art white OLEDs.

Within a refined approach, the probability that a TPQ process gives rise to degradation could be treated stochastically, e.g., by treating it as a thermally activated process with an activation energy with a Gaussian distribution. Such an approach is expected to give rise to a smaller value of the stretching exponent β. Experimentally, values of β around 0.5 have been observed. The effect is so far generally explained in a rather phenomenological manner [206]. It should be noted that the decay can also become more stretched-exponential like for OLEDs with imbalanced electron and hole mobilities, resulting in a highly nonuniform emission profile. If this picture is correct, lifetime studies could also, albeit indirectly, provide information about the shape of the emission profile.

We envisage that, using Equation 15.40, lifetime predictions can be obtained from KMC lifetime simulations if the parameter p_{degr} (or its distribution) can be determined from a few well-chosen combined experimental and KMC calibration studies. Subsequently, KMC-based lifetime predictions can be obtained for other measurement conditions (e.g., current density and temperature), device architectures, and other dye concentrations in the same host. So far, studies which could validate this view have not been carried out. As a first step, KMC simulations were carried out of the iridium dye concentration dependence of the lifetime, for otherwise identical simulation parameters [177]. Figure 15.13a shows the accelerated LT_{90} lifetime obtained from the simulations (large open circles), and results from simulations in which exciton diffusion was switched off (small open circles). For iridium dye concentrations above about 7 mol%, exciton diffusion is found to yield a significant decrease of the lifetime, up to a factor ~4. The exciton diffusion length is then larger than the average distance to a degraded site, so that a large fraction of the excitons which have been generated on nondegraded sites is lost due to diffusion and subsequent nonradiative decay on degraded sites. Interestingly, the IQE obtained when exciton diffusion was switched off was found to be slightly reduced for all systems studied, namely by 1%–2%. The reduction cannot be due to switching off the transfer of triplets from matrix to guest sites, as due to the large energy gap of the matrix materials all excitons are generated directly on the dye sites. We surmise that the reduction is due to switching off the possibility that excitons diffuse to molecular sites in regions with a slightly smaller average polaron density, in which the loss due to TPQ is reduced. This effect may for the highest iridium dye concentrations considered be compensated in part by an increase of the IQE due to a smaller exciton diffusion contribution to the TPQ loss. For the devices and the parameter values employed, the latter effect was found to be almost negligible for iridium dye concentrations around and below 8 mol% [177].

The simulation data shown in Figure 15.13a may be analyzed more quantitatively with the help of a useful relationship which has been established between the lifetime and the IQE roll-off [177]. It is based on a model which assumes uniform electron, hole, and exciton densities in the emissive layer. If



(a) (b)

FIGURE 15.13 (a) Ir-dye concentration dependence of the LT_{90} lifetime as obtained from accelerated KMC lifetime simulations for the OLEDs shown in Figure 15.12a, at 6 V, for the TPQ-t degradation scenario discussed in the main text (large open circles) and when switching off exciton diffusion (small open circles). The closed circles show the accelerated lifetime which is expected from the simulated roll-off using Equation 15.42. The curves are guides-to-the-eye. The predicted actual LT_{90} lifetime is equal to $LT_{90,\text{sim,acc}}/p_{\text{degr}}$, with p_{degr} the probability that upon a TPQ event degradation takes place. Note that due to a variation of the current density with the Ir-dye concentration (see the main text) the lifetime as obtained using Equation 15.42 is not proportional to that concentration. (b) Ir-dye concentration dependence of the fraction of degraded Ir-dye molecules in the $t \rightarrow \infty$ limit, as obtained from the same KMC simulations including and without exciton diffusion. (Reproduced from R. Coehoorn et al., *Adv. Funct. Mater.*, 25, 2024–2037, 2015. With permission.)

the charge carrier mobility is charge carrier concentration (c) dependent and proportional to c^b, with b a positive exponent which increases with an increasing width of the polaron DOS, the IQE roll-off curve is given by

$$\eta_{\text{IQE}} = \frac{\eta_{\text{rad}}}{1 + \left(\frac{J}{J_{50}}\right)^m}, \tag{15.41}$$

with η_{rad} the PL efficiency (0.75 for the case studied in this section), J_{50} the current density at which the IQE has dropped to 50% of its low-voltage value, and $m = (1 + b)/(2 + b)$. In the absence of exciton diffusion, the simulated accelerated (1/e) lifetime is then given by

$$\tau_{\text{sim,acc}} = edn_{\text{dye}}J_{50}^{m}/J^{m+1}, \tag{15.42}$$

where d is the EML layer thickness and n_{dye} is the dye molecule volume density. From this formalism, a current density acceleration exponent $m + 1 = (3 + 2b)/(2 + b)$ in the range 1.5–2 is expected, as is indeed often observed for phosphorescent OLEDs. The linear n_{dye} dependence is due to the linear increase with increasing dye concentration of the probability that a dye is still emissive after a certain period of operation. The closed circles in Figure 15.13 show the simulated accelerated LT_{90} lifetime, obtained using Equation 15.42 under the assumption of exponential decay ($\beta = 1$), so that $LT_{90,\text{sim,acc}} = -\ln(0.9) \times \tau_{\text{sim,acc}} \cong 0.105 \times \tau_{\text{sim,acc}}$. These predictions from the roll-off curves agree quite well with the explicit KMC simulation results, obtained when exciton diffusion is switched off (small open circles). We note that at the constant voltage (6-V) condition employed, the current density shows a weak but nonnegligible nonmonotonic dependence on the Ir-dye concentration. It shows a broad minimum around a concentration of approximately 10 mol%, and is ~15% larger for the 2 and 25 mol% systems. Within the concentration range studied, the transport shows a cross-over from a low-concentration guest–host–guest hopping regime, in which the guest molecule states act as traps, to a high-concentration regime, in which the transport is predominantly due to direct guest–guest hopping (see, e.g., Figure 15.5a in Ref. [100]). The lifetime at a fixed voltage, as obtained from Equation 15.42, is therefore not proportional to the Ir-dye concentration.

Various optical and chemical analytical techniques have been used to investigate the degradation mechanisms of OLEDs and to quantify the concentrations of degraded molecules (see Ref. [199] and references therein). In future studies, it would be useful to compare such experimental results with the results of KMC simulations. Interestingly, the simulations for the model systems discussed in this section revealed that, to a good approximation, the fraction of degraded Ir-dye molecules increases with time as

$$f_{\mathrm{degr}}(t) = f_{\mathrm{degr,final}} \left[1 - \exp\left(-\left(t/\tau_{\mathrm{sim,acc}}\right)^{\beta}\right)\right], \tag{15.43}$$

with $f_{\mathrm{degr,final}}$ the final ($t \rightarrow \infty$) fraction of degraded Ir-dye molecules.

The values of $\tau_{\mathrm{sim,acc}}$ and β are, within the numerical uncertainty, equal to the values describing the luminance decay. Figure 15.13b shows the Ir-dye concentration dependence of $f_{\mathrm{degr,final}}$ as obtained from KMC simulations with and without exciton diffusion. The figure shows that in the absence of diffusion, only 60%–70% of the sites have degraded in the $t = \infty$ limit. This result indicates that, on a significant fraction (30%–40%) of the sites, excitons are either never formed, or that excitons on those sites are well protected against TPQ due to a position of those sites well outside the somewhat filamentary electron and hole current density pathways. The first explanation is consistent with the finding that in systems with a monomodal Gaussian DOS exciton generation preferentially takes place on sites with a low-lying electron or hole state [73]. When exciton diffusion is included, $f_{\mathrm{degr,final}}$ is found to decrease significantly with increasing Ir-dye concentration, to only approximately 0.16 for 25 mol% systems. This is consistent with the view that, due to the energetic disorder, the average polaron density and the local polaron diffusivity will be quite nonuniform [99,207], so that in the case of strong exciton diffusion, degradation will occur predominantly on the relatively small fraction of sites that are located in a region with a large average polaron density and diffusivity.

We emphasize that the degradation scenario assumed in the case study discussed in this section was only chosen for the purpose of giving a demonstration of the feasibility of KMC lifetime simulations. Including monomolecular or other bimolecular scenarios or including refinements (e.g., a degradation probability distribution, conversion of the degraded molecules to polaron trap sites, or an extension of the TPQ interaction range so that the role of the degraded molecule as an exciton quencher is enhanced) is straightforward. It also will be useful to develop analytical models as discussed above for other degradation scenarios and to extend these to a more realistic nonuniform emission from the EML, in which the lifetime becomes position-dependent, so that the emission decay becomes more stretched-exponential like.

15.13 Outlook

In this chapter, we have reviewed multiscale techniques used to simulate organic LEDs and demonstrated the feasibility of full 3D OLED modeling. As an outlook, we would like to mention areas where substantial method development is still required in order to achieve a parameter-free modeling of realistic devices. Refined studies that aim at developing a final view on the detailed performance of specific devices should consider the following: (1) first-principles evaluations of charge injection rates, (2) explicit treatment of the induction interaction when solving the master equation, (3) quantitative treatment of excited states embedded in a heterogeneous polarizable molecular environment, (4) more quantitative descriptions of charge–exciton and exciton–exciton interactions, and (5) descriptions of TTA and TPQ as longer-range Förster and Dexter-type interactions. Advancements in all these directions are absolutely vital for devising accurate structure–property relationships for organic semiconductors used in OLEDs.

Acknowledgments

This work was supported in part by the Federal Ministry of Education and Research BMBF grants MEDOS (FKZ 03EK3503B), MESOMERIE (FKZ 13N10723), and InterPhase (FKZ 13N13661). The project received funding from the NMP-20-2014—“Widening materials models” program under Grant Agreement No. 646259 (MOSTOPHOS). The project was also supported by NanoNextNL, a nanotechnology program of the Dutch Ministry of Economic Affairs. Part of the work of one of the authors (RC) was carried out at the Philips Research Laboratories (Eindhoven, The Netherlands).

References

1. I. Akasaki. Nobel lecture: Fascinated journeys into blue light. *Rev. Mod. Phys.*, 87(4):1119–1131, 2015.
2. H. Amano. Nobel lecture: Growth of GaN on sapphire via low-temperature de-posited buffer layer and realization of \$p\$-type GaN by Mg doping followed by low-energy electron beam irradiation. *Rev. Mod. Phys.*, 87(4):1133–1138, 2015.
3. S. Nakamura. Nobel lecture: Background story of the invention of efficient blue InGaN light emitting diodes. *Rev. Mod. Phys.*, 87(4):1139–1151, 2015.
4. S. R. Forrest. The path to ubiquitous and low-cost organic electronic appliances on plastic. *Nature*, 428(6986):911–918, 2004.
5. C. W. Tang and S. A. VanSlyke. Organic electroluminescent diodes. *Appl. Phys. Lett.*, 51(12):913–915, 1987.
6. R. Friend, J. Burroughes, and D. Bradley. Electroluminescent devices, Patent WO9013148 (A1). Cambridge Research and Innovation Limited; Cambridge Capital Management Limited; Lynxvale Limited, 1990.
7. R. Coehoorn, V. van Elsbergen, and C. Verschuren. High efficiency OLEDs for lighting applications. In E. Cantatore, editor, *Applications of Organic and Printed Electronics, Integrated Circuits and Systems*, pages 83–100. New York, NY: Springer, 2013.
8. T. Otani. Samsung shows new 4 flexible AMOLED that is so thin (0.05 mm) it “aps” in the wind textbar OLED-Info. 2008. https://www.oled-info.com/samsung-shows-new-4-flexible-amoled-so-thin-005mm-it-flaps-wind
9. R. Mertens. Continental shows a prototype dual-screen flexible AMOLED display for the automotive market—OLED-Info, 2014. https://www.oled-info.com/continental-shows-prototype-dual-screen-flexible-amoled-display-automotive-market
10. R. Olivares-Amaya, C. Amador-Bedolla, J. Hachmann, S. Atahan-Evrenk, R. S. Sanchez-Carrera, L. Vogt, and A. Aspuru-Guzik. Accelerated computational discovery of high-performance materials for organic photovoltaics by means of cheminformatics. *Energy Environ. Sci.*, 4(12):4849–4861, 2011.
11. P. Deglmann, A. Schaefer, and C. Lennartz. Application of quantum calculations in the chemical industry—An overview. *Int. J. Quantum Chem.*, 115(3):107–136, 2015. WOS:000346654700001.
12. H. Yersin. Triplet emitters for OLED applications. Mechanisms of exciton trapping and control of emission properties. In *Transition Metal and Rare Earth Compounds*, Yersin, Hartmut (Ed.) No. 241 in Topics in Current Chemistry, Hartmut Yersin (Ed.), pp. 1–26. Berlin: Springer, 2004.
13. L. Duan, K. Xie, and Y. Qiu. Review paper: Progress on efficient cathodes for organic light-emitting diodes. *J. Soc. Inf. Disp.*, 19(6):453–461, 2011.
14. A. Fukase, K. Luan Thanh Dao, and J. Kido. High-efficiency organic electroluminescent devices using iridium complex emitter and arylamine-containing polymer buffer layer. *Polymer. Adv. Tech.*, 13(8):601–604, 2002.
15. X. Zhou, D. S. Qin, M. Pfeiffer, J. Blochwitz-Nimoth, A. Werner, J. Drechsel, B. Maen-nig, K. Leo, M. Bold, P. Erk, and H. Hartmann. High-efficiency electrophosphorescent organic light-emitting diodes with double light-emitting layers. *Appl. Phys. Lett.*, 81(21):4070–4072, 2002.

16. C. Adachi, M. A. Baldo, S. R. Forrest, S. Lamansky, M. E. Thompson, and R. C. Kwong. High-efficiency red electrophosphorescence devices. *Appl. Phys. Lett.*, 78(11):1622–1624, 2001.
17. Y. Kawamura, S. Yanagida, and S. R. Forrest. Energy transfer in polymer electrophosphorescent light emitting devices with single and multiple doped luminescent layers. *J. Appl. Phys.*, 92(1):87–93, 2002.
18. B. A. Gregg, S.-G. Chen, and R. A. Cormier. Coulomb forces and doping in organic semiconductors. *Chem. Mater.*, 16(23):4586–4599, 2004.
19. K. Walzer, B. Maennig, M. Pfeiffer, and K. Leo. Highly efficient organic devices based on electrically doped transport layers. *Chem. Rev.*, 107(4):1233–1271, 2007.
20. M. A. Baldo, S. Lamansky, P. E. Burrows, M. E. Thompson, and S. R. Forrest. Very high-efficiency green organic light-emitting devices based on electrophosphorescence. *Appl. Phys. Lett.*, 75(1):4–6, 1999.
21. Md. K. Nazeeruddin, R. Humphry-Baker, D. Berner, S. Rivier, L. Zuppiroli, and M. Graetzel. Highly phosphorescence iridium complexes and their application in organic light-emitting devices. *J. Am. Chem. Soc.*, 125(29):8790–8797, 2003.
22. Q. Peng, W. Li, S. Zhang, P. Chen, F. Li, and Y. Ma. Evidence of the reverse intersystem crossing in intra-molecular charge transfer fluorescence based organic light-emitting devices through magneto-electroluminescence measurements. *Adv. Opt. Mater.*, 1(5):362–366, 2013.
23. S. Hirata, Y. Sakai, K. Masui, H. Tanaka, S. Youn Lee, H. Nomura, N. Nakamura, M. Yasumatsu, H. Nakanotani, Q. Zhang, K. Shizu, H. Miyazaki, and C. Adachi. Highly efficient blue electroluminescence based on thermally activated delayed fluorescence. *Nat. Mater.*, 14(3):330–336, 2015.
24. S. Reineke, K. Walzer, and K. Leo. Triplet–exciton quenching in organic phosphorescent light-emitting diodes with Ir-based emitters. *Phys. Rev. B*, 75(12):125328, 2007.
25. E. Knapp, R. Husermann, H. U. Schwarzenbach, and B. Ruhstaller. Numerical simulation of charge transport in disordered organic semiconductor devices. *J. Appl. Phys.*, 108(5):054504, 2010.
26. B. Ruhstaller, E. Knapp, B. Perucco, N. Reinke, D. Rezzonico, and F. Mueller. Advanced numerical simulation of organic light-emitting devices. In O. Sergiyenko, editor, *Optoelectronic Devices and Properties*. InTech, 2011. Rijeka, Croatia: DOI: 10.5772/14626.
27. G. Kaniadakis and P. Quarati. Kinetic equation for classical particles obeying an exclusion principle. *Phys. Rev. E*, 48(6):4263–4270, 1993.
28. Y. Roichman and N. Tessler. Generalized Einstein relation for disordered semiconductors-implications for device performance. *Appl. Phys. Lett.*, 80(11):1948–1950, 2002.
29. G. Lakhwani, A. Rao, and R. H. Friend. Bimolecular recombination in organic photovoltaics. *Annu. Rev. Phys. Chem.*, 65(1):557–581, 2014.
30. B. Perucco, N. A. Reinke, D. Rezzonico, E. Knapp, S. Harkema, and B. Ruhstaller. On the exciton profile in OLEDs-seamless optical and electrical modeling. *Org. Electron.*, 13(10):1827–1835, 2012.
31. G.-J. A. H. Wetzelaer and P. W. M. Blom. Diffusion-driven currents in organic-semiconductor diodes. *NPG Asia Mater.*, 6(7):e110, 2014.
32. J.-H. Lee, S. Lee, S.-J. Yoo, K.-H. Kim, and J.-J. Kim. Langevin and trap-assisted recombination in phosphorescent organic light emitting diodes. *Adv. Funct. Mater.*, 24(29):4681–4688, 2014.
33. R. R. Chance, A. Prock, and R. Silbey. Molecular fluorescence and energy transfer near interfaces. In I. Prigogine and S. A. Rice, editors, *Advances in Chemical Physics*, pp. 1–65. Hoboken, NJ: John Wiley & Sons, 1978.
34. H. K. Gummel. A self-consistent iterative scheme for one-dimensional steady state transistor calculations. *IEEE Trans. Electron Dev.*, 11(10):455–465, 1964.
35. D. L. Scharfetter and H. K. Gummel. Large-signal analysis of a silicon Read diode oscillator. *IEEE Trans. Electron Dev.*, 16(1):64–77, 1969.
36. E. Knapp and B. Ruhstaller. Numerical impedance analysis for organic semiconductors with exponential distribution of localized states. *Appl. Phys. Lett.*, 99(9), 093304, 2011.

37. E. Tuti, I. Batisti, and D. Berner. Injection and strong current channeling in organic disordered media. *Phys. Rev. B*, 70(16):161202, 2004.
38. K. D. Meisel, W. F. Pasveer, J. Cottaar, C. Tanase, R. Coehoorn, P. A. Bobbert, P. W. M. Blom, D. M. de Leeuw, and M. A. J. Michels. Charge-carrier mobilities in disordered semiconducting polymers: Effects of carrier density and electric field. *Phys. Status Solidi C*, 3(2):267–270, 2006.
39. N. Rappaport, Y. Preezant, and N. Tessler. Spatially dispersive transport: A mesoscopic phenomenon in disordered organic semiconductors. *Phys. Rev. B*, 76(23):235323, 2007.
40. J. Honerkamp. *Stochastische Dynamische Systeme : Konzepte, Numerische Methoden, Datenanalysen*. Weinheim: Wiley-VCH, 1990.
41. B. Derrida. Velocity and diffusion constant of a periodic one-dimensional hopping model. *J. Stat. Phys.*, 31(3):433–450, 1983.
42. K. Seki and M. Tachiya. Electric field dependence of charge mobility in energetically disordered materials: Polaron aspects. *Phys. Rev. B*, 65:014305, 2001.
43. S. D. Baranovskii. Theoretical description of charge transport in disordered organic semiconductors. *Phys. Status Solidi B*, 251(3):487–525, 2014.
44. J. Cottaar and P. A. Bobbert. Calculating charge-carrier mobilities in disordered semiconducting polymers: Mean field and beyond. *Phys. Rev. B*, 74(11):115204, 2006.
45. V. Rühle, A. Lukyanov, F. May, M. Schrader, T. Vehoff, J. Kirkpatrick, B. Baumeier, and D. Andrienko. Microscopic simulations of charge transport in disordered organic semiconductors. *J. Chem. Theory Comput.*, 7(10):3335–3345, 2011.
46. M. Heidernaetsch, M. Bauer, and G. Radons. Characterizing N-dimensional anisotropic Brownian motion by the distribution of diffusivities. *J. Chem. Phys.*, 139(18):184105, 2013.
47. J. L. Doob. Topics in the theory of Markoff chains. *Trans. Am. Math. Soc.*, 52(1):37–64, 1942.
48. J. L. Doob. Markoff chains—Denumerable case. *Trans. Am. Math. Soc.*, 58(3):455, 1945.
49. W M. Young and E. W. Elcock. Monte Carlo studies of vacancy migration in binary ordered alloys: I. *Proc. Phys. Soc.*, 89(3):735–746, 1966.
50. A. B. Bortz, M. H. Kalos, and J. L. Lebowitz. A new algorithm for Monte Carlo simulation of Ising spin systems. *J. Comput. Phys.*, 17(1):10–18, 1975.
51. D. T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.*, 22(4):403–434, 1976.
52. D. T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.*, 81(25):2340–2361, 1977.
53. K. A. Fichthorn and W. H. Weinberg. Theoretical foundations of dynamical Monte Carlo simulations. *J. Chem. Phys.*, 95(2):1090, 1991.
54. A. P. J. Jansen. Monte Carlo simulations of chemical reactions on a surface with time-dependent reaction-rate constants. *Comput. Phys. Commun.*, 86(1–2):1–12, 1995.
55. P. Kordt, J. J. M. van der Holst, M. Al Helwi, W. Kowalsky, F. May, A. Badinski, C. Lennartz, and D. Andrienko. Modeling of organic light emitting diodes: From molecular to device properties. *Adv. Funct. Mater.*, 25(13):1955–1971, 2015.
56. M. A. Gibson and J. Bruck. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A.*, 104(9):1876–1889, 2000.
57. B. D. Lubachevsky. Efficient parallel simulations of dynamic Ising spin systems. *J. Comput. Phys.*, 75(1):103–122, 1988.
58. Y. Shim and J. G. Amar. Semirigorous synchronous sublattice algorithm for parallel kinetic Monte Carlo simulations of thin film growth. *Phys. Rev. B*, 71(12), 125432, 2005.
59. M. Merrick and K. A. Fichthorn. Synchronous relaxation algorithm for parallel kinetic Monte Carlo simulations of thin film growth. *Phys. Rev. E*, 75(1), 011606, 2007.
60. E. Martnez, J. Marian, M. H. Kalos, and J. M. Perlado. Synchronous parallel kinetic Monte Carlo for continuum diffusion-reaction systems. *J. Comput. Phys.*, 227(8):3804–3823, 2008.

61. N. J. van der Kaap and L. J. A. Koster. Massively parallel kinetic Monte Carlo simulations of charge carrier transport in organic semiconductors. *J. Comput. Phys.*, 307:321–332, 2016.
62. A. Miller and E. Abrahams. Impurity conduction at low concentrations. *Phys. Rev.*, 120(3):745–755, 1960.
63. H. Baessler. Charge transport in disordered organic photoconductors: A Monte Carlo simulation study. *Phys. Status Solidi B*, 175(1):15–56, 1993.
64. R. A. Marcus. Electron transfer reactions in chemistry. Theory and experiment. *Rev. Mod. Phys.*, 65(3):599–610, 1993.
65. G. R. Hutchison, M. A. Ratner, and T. J. Marks. Hopping transport in conductive heterocyclic oligomers: Reorganization energies and substituent effects. *J. Am. Chem. Soc.*, 127(7):2339–2350, 2005.
66. V. May and O. Kuehn. *Charge and Energy Transfer Dynamics in Molecular Systems*, 3rd ed. Weinheim: Wiley-VCH, 2011.
67. M. Bixon and J. Jortner. Electron transfer from isolated molecules to biomolecules. In I. Prigogine and S. A. Rice, editors, *Advances in Chemical Physics*, pp. 35–202. Hoboken, NJ: John Wiley & Sons, 2007.
68. R. Egger, C. H. Mak, and U. Weiss. Quantum rates for nonadiabatic electron transfer. *J. Chem. Phys.*, 100(4):2651–2660, 1994.
69. M. P. A. Fisher and A. T. Dorsey. Dissipative quantum tunneling in a biased double-well system at finite temperatures. *Phys. Rev. Lett.*, 54(15):1609–1612, 1985.
70. H. Grabert and U. Weiss. Quantum tunneling rates for asymmetric double-well systems with Ohmic dissipation. *Phys. Rev. Lett.*, 54(15):1605–1608, 1985.
71. A. J. Leggett, S. Chakravarty, A. T. Dorsey, M. P. A. Fisher, A. Garg, and W. Zwerger. Dynamics of the dissipative two-state system. *Rev. Mod. Phys.*, 59(1):1–85, 1987.
72. K. Asadi, A. J. Kronemeijer, T. Cramer, L. J. A. Koster, P. W. M. Blom, and D. M. de Leeuw. Polaron hopping mediated by nuclear tunnelling in semiconducting polymers at high carrier density. *Nat. Commun.*, 4:1710, 2013.
73. J. J. M. van der Holst, F. W. A. van Oost, R. Coehoorn, and P. A. Bobbert. Electron–hole recombination in disordered organic semiconductors: Validity of the Langevin formula. *Phys. Rev. B*, 80(23):235202-1–235202-8, 2009.
74. T. H. Foerster. Zwischenmolekulare Energiewanderung und Fluoreszenz. *Ann. Phys.*, 437(1–2):55–75, 1948.
75. D. L. Dexter. A theory of sensitized luminescence in solids. *J. Chem. Phys.*, 21(5):836–850, 1953.
76. G. D. Scholes and K. P. Ghiggino. Rate expressions for excitation transfer I. Radiationless transition theory perspective. *J. Chem. Phys.*, 101(2):1251–1261, 1994.
77. G. D. Scholes, R. D. Harcourt, and K. P. Ghiggino. Rate expressions for excitation transfer. III. An ab initio study of electronic factors in excitation transfer and exciton resonance interactions. *J. Chem. Phys.*, 102(24):9574–9581, 1995.
78. R. D. Harcourt, G. D. Scholes, and K. P. Ghiggino. Rate expressions for excitation transfer. II. Electronic considerations of direct and through configuration exciton resonance interactions. *J. Chem. Phys.*, 101(12):10521–10525, 1994.
79. S. Speiser. Photophysics and mechanisms of intramolecular electronic energy transfer in bichromophoric molecular systems: Solution and supersonic jet studies. *Chem. Rev.*, 96(6):1953–1976, 1996.
80. R. F. Fink, J. Pfister, H. Mei Zhao, and B. Engels. Assessment of quantum chemical methods and basis sets for excitation energy transfer. *Chem. Phys.*, 346(1–3):275–285, 2008. WOS:000256142000032.
81. C. Tanase, P. W. M. Blom, D. M. de Leeuw, and E. J. Meijer. Charge carrier density dependence of the hole mobility in poly(p-phenylene vinylene). *Phys. Status Solidi A*, 201(6):1236–1245, 2004.

82. F. Laquai and D. Hertel. Influence of hole transport units on the efficiency of polymer light emitting diodes. *Appl. Phys. Lett.*, 90(14):142109, 2007.
83. S. V. Novikov, D. H. Dunlap, V. M. Kenkre, P. E. Parris, and A. V. Vannikov. Essential role of correlations in governing charge transport in disordered organic materials. *Phys. Rev. Lett.*, 81(20):4472–4475, 1998.
84. D. H. Dunlap, P. E. Parris, and V. M. Kenkre. Charge-dipole model for the universal field dependence of mobilities in molecularly doped polymers. *Phys. Rev. Lett.*, 77(3):542–545, 1996.
85. S. W. de Leeuw, J. W. Perram, and E. R. Smith. Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proc. Royal Soc. A: Math. Phys. Eng. Sci.*, 373(1752):27–56, 1980.
86. P. P. Ewald. Die Berechnung optischer und elektrostatischer Gitterpotentiale. *Ann. Phys.*, 369(3):253–287, 1921.
87. R. H. Young. Dipolar lattice model of disorder in random media analytical evaluation of the Gaussian disorder model. *Philos. Mag. B*, 72(4):435–457, 1995.
88. S. V. Novikov and A. V. Vannikov. Cluster structure in the distribution of the electrostatic potential in a lattice of randomly oriented dipoles. *J. Phys. Chem.*, 99(40):14573–14576, 1995.
89. P. Kordt and D. Andrienko. Modeling of spatially correlated energetic disorder in organic semiconductors. *J. Chem. Theory Comput.*, 12(1):36–40, 2016.
90. W. F. Pasveer, J. Cottaar, C. Tanase, R. Coehoorn, P. A. Bobbert, P. W. M. Blom, D. M. de Leeuw, and M. A. J. Michels. Unified description of charge-carrier mobilities in disordered semiconducting polymers. *Phys. Rev. Lett.*, 94(20), 2005.
91. M. Bouhassoune, S. L. M. van Mensfoort, P. A. Bobbert, and R. Coehoorn. Carrier-density and field-dependent charge-carrier mobility in organic semiconductors with correlated Gaussian disorder. *Org. Electron.*, 10(3):437–445, 2009.
92. A. J. Stone. *The Theory of Intermolecular Forces*. Oxford: Clarendon Press, 1997.
93. C. Poelking and D. Andrienko. Long-range embedding of molecular ions and excitations in a polarizable molecular environment. *J. Chem. Theory Comput.*, 12(9):4516–4523, 2016.
94. C. Poelking, M. Tietze, C. Elschner, S. Olthof, D. Hertel, B. Baumeier, F. Wrthner, K. Meerholz, K. Leo, and D. Andrienko. Impact of mesoscale order on open-circuit voltage in organic solar cells. *Nat. Mater.*, 14(4):434–439, 2014.
95. C. Poelking and D. Andrienko. Design rules for organic donor-acceptor heterojunctions: Pathway for charge splitting and detrapping. *J. Am. Chem. Soc.*, 137(19):6320–6326, 2015.
96. D. Andrienko. Simulations of Morphology and Charge Transport in Supramolecular Organic Materials. In *Supramolecular Materials for Opto-Electronics*, Norbert Koch (Ed.). Cambridge, UK: The Royal Society of Chemistry, 2014.
97. B. T. Thole. Molecular polarizabilities calculated with a modified dipole interaction. *Chem. Phys.*, 59(3):341–350, 1981.
98. S. L. M. van Mensfoort and R. Coehoorn. Effect of Gaussian disorder on the voltage dependence of the current density in sandwich-type devices based on organic semiconductors. *Phys. Rev. B*, 78(8):085207, 2008.
99. J. J. M. van der Holst, M. A. Uijttewaal, B. Ramachandhran, R. Coehoorn, P. A. Bobbert, G. A. de Wijs, and R. A. de Groot. Modeling and analysis of the three-dimensional current density in sandwich-type single-carrier devices of disordered organic semiconductors. *Phys. Rev. B*, 79(8):085203, 2009.
100. R. Coehoorn and P. A. Bobbert. Effects of Gaussian disorder on charge carrier transport and recombination in organic semiconductors. *Phys. Status Solidi A*, 209(12):2354–2377, 2012.
101. J.-L. Bredas, D. Beljonne, V. Coropceanu, and J. Cornil. Charge-transfer and energy-transfer processes in pi-conjugated oligomers and polymers: A molecular picture. *Chem. Rev.*, 104(11):4971–5004, 2004.
102. J. R. Reimers. A practical method for the use of curvilinear coordinates in calculations of normal-mode-projected displacements and Duschinsky rotation matrices for large molecules. *J. Chem. Phys.*, 115(20):9103–9109, 2001.

103. D. P. McMahon and A. Troisi. Evaluation of the external reorganization energy of polyacenes. *J. Phys. Chem. Lett.*, 1(6):941–946, 2010.
104. J. E. Norton and J.-L. Bredas. Polarization energies in oligoacene semiconductor crystals. *J. Am. Chem. Soc.*, 130(37):12377–12384, 2008.
105. B. Baumeier, J. Kirkpatrick, and D. Andrienko. Density-functional based determination of intermolecular charge transfer properties for large-scale morphologies. *Phys. Chem. Chem. Phys.*, 12(36):11103, 2010.
106. J. Huang and M. Kertesz. Validation of intermolecular transfer integral and bandwidth calculations for organic molecular materials. *J. Chem. Phys.*, 122(23):234707, 2005.
107. E. F. Valeev, V. Coropceanu, D. A. da Silva Filho, S. Salman, and J.-L. Bredas. Effect of electronic polarization on charge-transport parameters in molecular organic semiconductors. *J. Am. Chem. Soc.*, 128(30):9882–9886, 2006.
108. T. Van Voorhis, T. Kowalczyk, B. Kaduk, L.-P. Wang, C.-L. Cheng, and Q. Wu. The diabatic picture of electron transfer, reaction barriers, and molecular dynamics. *Annu. Rev. Phys. Chem.*, 61(1):149–170, 2010.
109. J. Ridley and M. Zerner. An intermediate neglect of differential overlap technique for spectroscopy: Pyrrole and the azines. *Theor. Chim. Acta*, 32(2):111–134, 1973.
110. J. Kirkpatrick. An approximate method for calculating transfer integrals based on the ZINDO Hamiltonian. *Int. J. Quantum Chem.*, 108(1):51–56, 2008.
111. V. Coropceanu, J. Cornil, D. A. da Silva Filho, Y. Olivier, R. Silbey, and J.-L. Bredas. Charge transport in organic semiconductors. *Chem. Rev.*, 107(4):926–952, 2007.
112. A. Kubas, F. Hoffmann, A. Heck, H. Oberhofer, M. Elstner, and J. Blumberger. Electronic couplings for molecular charge transfer: Bench-marking CDFT, FODFT, and FODFTB against high-level ab initio calculations. *J. Chem. Phys.*, 140(10):104105, 2014.
113. F. Gajdos, S. Valner, F. Hoffmann, J. Spencer, M. Breuer, A. Kubas, M. Dupuis, and J. Blumberger. Ultrafast estimation of electronic couplings for electron transfer between pi-conjugated organic molecules. *J. Chem. Theory Comput.*, 10(10):4653–4660, 2014.
114. P. Kordt, O. Stenzel, B. Baumeier, V. Schmidt, and D. Andrienko. Parametrization of extended Gaussian disorder models from microscopic charge transport simulations. *J. Chem. Theory Comput.*, 10(6):2508–2513, 2014.
115. S. Athanasopoulos, J. Kirkpatrick, D. Martnez, J. M. Frost, C. M. Foden, A. B. Walker, and J. Nelson. Predictive study of charge transport in disordered semiconducting polymers. *Nano Lett.*, 7(6):1785–1788, 2007.
116. J. Kirkpatrick, V. Marcon, J. Nelson, K. Kremer, and D. Andrienko. Charge mobility of discotic mesophases: A multiscale quantum and classical study. *Phys. Rev. Lett.*, 98(22):227402, 2007.
117. J. Nelson, J. J. Kwiatkowski, J. Kirkpatrick, and J. M. Frost. Modeling charge transport in organic photovoltaic materials. *Acc. Chem. Res.*, 42(11):1768–1778, 2009.
118. Y. Nagata and C. Lennartz. Atomistic simulation on charge mobility of amorphous tris(8-hydroxyquinoline) aluminum (Alq[sub 3]): Origin of Poole–Frenkel-type behavior. *J. Chem. Phys.*, 129(3):034709, 2008.
119. A. Fuchs, T. Steinbrecher, M. S. Mommer, Y. Nagata, M. Elstner, and C. Lennartz. Molecular origin of differences in hole and electron mobility in amorphous Alq3—A multiscale simulation study. *Phys. Chem. Chem. Phys.*, 14(12):4259–4270, 2012.
120. A. J. Stone. Distributed polarizabilities. *Mol. Phys.*, 56(5):1065–1082, 1985.
121. A. J. Stone. Distributed multipole analysis—Stability for large basis sets. *J. Chem. Theory Comput.*, 1(6):1128–1132, 2005.
122. A. J. Misquitta and A. J. Stone. Distributed polarizabilities obtained using a constrained density-fitting algorithm. *J. Chem. Phys.*, 124(2):024111, 2006.
123. T. Bereau, D. Andrienko, and O. Anatole von Lilienfeld. Transferable atomic multipole machine learning models for small organic molecules. *J. Chem. Theory Comput.*, 11(7):3225–3233, 2015.

124. W. L. Jorgensen and J. Tirado-Rives. The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *J. Am. Chem. Soc.*, 110(6):1657–1666, 1988.
125. W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.*, 118(45):11225–11236, 1996.
126. A. Lukyanov, C. Lennartz, and D. Andrienko. Amorphous films of tris(8-hydroxyquinolinato)aluminium: Force-field, morphology, and charge transport. *Phys. Status Solidi (A)*, 206(12):2737–2742, 2009.
127. B. H. Besler, K. M. Merz, and P. A. Kollman. Atomic charges derived from semiempirical methods. *J. Comput. Chem.*, 11(4):431–439, 1990.
128. C. M. Breneman and K. B. Wiberg. Determining atom-centered monopoles from molecular electrostatic potentials. The need for high sampling density in formamide conformational analysis. *J. Comput. Chem.*, 11(3):361–373, 1990.
129. F. May, M. Al-Helwi, B. Baumeier, W. Kowalsky, E. Fuchs, C. Lennartz, and D. Andrienko. Design rules for charge-transport efficient host materials for phosphorescent organic light-emitting diodes. *J. Am. Chem. Soc.*, 134(33):13818–13822, 2012. WOS:000307699000042.
130. A. Lukyanov and D. Andrienko. Extracting nondispersive charge carrier mobilities of organic semiconductors from simulations of small systems. *Phys. Rev. B*, 82(19):193202, 2010.
131. T. Neumann, D. Danilov, C. Lennartz, and W. Wenzel. Modeling disordered morphologies in organic semiconductors. *J. Comput. Chem.*, 34(31):2716–2725, 2013.
132. M. Moral, W.-J. Son, J. C. Sancho-Garca, Y. Olivier, and L. Muccioli. Cost-effective force field tailored for solid-phase simulations of OLED materials. *J. Chem. Theory Comput.*, 11(7):3383–3392, 2015.
133. O. Anatole von Lilienfeld, I. Tavernelli, U. Rothlisberger, and D. Sebastiani. Optimization of effective atom centered potentials for London dispersion forces in density functional theory. *Phys. Rev. Lett.*, 93(15):153004, 2004.
134. W. G. Noid, J.-W. Chu, G. S. Ayton, V. Krishna, S. Izvekov, G. A. Voth, A. Das, and H. C. Andersen. The multiscale coarse graining method. 1. A rigorous bridge between atomistic and coarse-grained models. *J. Chem. Phys.*, 128:244114, 2008.
135. V. Rühle, C. Junghans, A. Lukyanov, K. Kremer, and D. Andrienko. Versatile object-oriented toolkit for coarse-graining applications. *J. Chem. Theory Comput.*, 5(12):3211–3223, 2009.
136. A. K. Soper. Empirical potential Monte Carlo simulation of fluid structure. *Chem. Phys.*, 202(2–3):295–306, 1996.
137. A. P. Lyubartsev and A. Laaksonen. Calculation of effective interaction potentials from radial distribution functions: A reverse Monte Carlo approach. *Phys. Rev. E*, 52(4):3730–3737, 1995.
138. D. Reith, M. Pütz, and F. Müller-Plathe. Deriving effective mesoscale potentials from atomistic simulations. *J. Comput. Chem.*, 24(13):1624–1636, 2003.
139. M. Jochum, D. Andrienko, K. Kremer, and C. Peter. Structure-based coarse-graining in liquid slabs. *J. Chem. Phys.*, 137(6):064102, 2012.
140. F. Ercolessi and J. B. Adams. Interatomic potentials from first-principles calculations: The force-matching method. *EPL*, 26(8):583, 1994.
141. S. Izvekov, M. Parrinello, C. J. Burnham, and G. A. Voth. Effective force fields for condensed phase systems from ab initio molecular dynamics simulation: A new method for force-matching. *J. Chem. Phys.*, 120(23):10896–10913, 2004.
142. M. Scott Shell. The relative entropy is fundamental to multiscale and inverse thermo-dynamic problems. *J. Chem. Phys.*, 129(14):144108, 2008.
143. W. G. Noid, J.-W. Chu, G. S. Ayton, and G. A. Voth. Multiscale coarse-graining and structural correlations: Connections to liquid-state theory. *J. Phys. Chem. B*, 111(16):4116–4127, 2007.

144. W. G. Noid. Perspective: Coarse-grained models for biomolecular systems. *J. Chem. Phys.*, 139(9):090901, 2013.
145. K. Ch. Daoulas and M. Müller. Comparison of simulations of lipid membranes with membranes of block copolymers. In W. Peter Meier and W. Knoll, editors, *Polymer Membranes/Biomembranes*, No. 224 in Advances in Polymer Science, pp. 43–85. Berlin: Springer, 2010.
146. P. Gemnden, C. Poelking, K. Kremer, D. Andrienko, and K. Ch. Daoulas. Nematic ordering, conjugation, and density of states of soluble polymeric semiconductors. *Macromolecules*, 46(14):5762–5774, 2013.
147. C. Poelking, K. Daoulas, A. Troisi, and D. Andrienko. Morphology and charge transport in P3ht: A theorist's perspective. In S. Ludwigs, editor, *P3HT Revisited—From Molecular Scale to Solar Cell Devices*, No. 265 in Advances in Polymer Science, pp. 139–180. Berlin: Springer, 2014. doi:10.1007/12-2014-277.
148. P. Gemnden, C. Poelking, K. Kremer, K. Daoulas, and D. Andrienko. Effect of mesoscale ordering on the density of states of polymeric semiconductors. *Macromol. Rapid Commun.*, 36(11):1047–1053, 2015.
149. P, Kordt, S. Stodtmann, A. Badinski, M. Al Helwi, C. Lennartz, and D. Andrienko. Parameter-free continuous drift–diffusion models of amorphous organic semiconductors. *Phys. Chem. Chem. Phys.*, 17(35):22778–22783, 2015.
150. B. Baumeier, O. Stenzel, C. Poelking, D. Andrienko, and V. Schmidt. Stochastic modeling of molecular charge transport networks. *Phys. Rev. B*, 86(18):184202, 2012.
151. W. Tschoep, K. Kremer, J. Batoulis, T. Buerger, and O. Hahn. Simulation of polymer melts. I. Coarse-graining procedure for polycarbonates. *Acta Polym.*, 49(2–3):61–74, 1998.
152. T. Murtola, A. Bunker, I. Vattulainen, M. Deserno, and M. Karttunen. Multiscale modeling of emergent materials: Biological and soft matter. *Phys. Chem. Chem. Phys.*, 11(12):1869–1892, 2009.
153. R. L. Henderson. A uniqueness theorem for fluid pair correlation functions. *Phys. Lett. A*, 49(3):197–198, 1974.
154. W.-K. Oh, S. Qamar Hussain, Y.-J. Lee, Y. Lee, S. Ahn, and J. Yi. Study on the ITO work function and hole injection barrier at the interface of ITO/a-Si:H(p) in amorphous/crystalline silicon heterojunction solar cells. *Mater. Res. Bull.*, 47(10):3032–3035, 2012.
155. R. Schlaf, H. Murata, and Z. H. Kafafi. Work function measurements on indium tin oxide films. *J. Electron Spectrosc. Relat. Phenom.*, 120(1–3):149–154, 2001.
156. G. M. Wu, H. H. Lin, and H. C. Lu. Work function and valence band structure of tin-doped indium oxide thin films for OLEDs. *Vacuum*, 82(12):1371–1374, 2008.
157. C. Hyung Kim, C. Deuck Bae, K. Hee Ryu, B. Ki Lee, and H. Jung Shin. Local work function measurements on various inorganic materials using Kelvin probe force spectroscopy. *Solid State Phenom.*, 124–126:607–610, 2007.
158. D. R. Lide. *CRC Handbook of Chemistry and Physics: A Ready-Reference Book of Chemical and Physical Data*, 79th ed. Boca Raton, FL: CRC Press, 1998.
159. M. Mesta, J. Cottaar, R. Coehoorn, and P. A. Bobbert. Study of charge-carrier relaxation in a disordered organic semiconductor by simulating impedance spectroscopy. *Appl. Phys. Lett.*, 104(21):213301-1–213301-4, 2014.
160. J. J. M. van der Holst, F. W. A. van Oost, R. Coehoorn, and P. A. Bobbert. Monte Carlo study of charge transport in organic sandwich-type single-carrier devices: Effects of Coulomb interactions. *Phys. Rev. B*, 83(8):085206, 2011.
161. S. L. M. van Mensfoort, S. I. E. Vulto, R. A. J. Janssen, and R. Coehoorn. Hole transport in polyfluorene-based sandwich-type devices: Quantitative analysis of the role of energetic disorder. *Phys. Rev. B*, 78(8):085208, 2008.
162. R. J. de Vries, S. L. M. van Mensfoort, R. A. J. Janssen, and R. Coehoorn. Relation between the built-in voltage in organic light-emitting diodes and the zero-field voltage as measured by electroabsorption. *Phys. Rev. B*, 81(12):125203, 2010.

163. M. Mesta, C. Schaefer, J. de Groot, J. Cottaar, R. Coehoorn, and P. A. Bobbert. Charge-carrier relaxation in disordered organic semiconductors studied by dark injection: Experiment and modeling. *Phys. Rev. B*, 88(17):174204, 2013.
164. W. C. Germs, J. J. M. van der Holst, S. L. M. van Mensfoort, P. A. Bobbert, and R. Coehoorn. Modeling of the transient mobility in disordered organic semiconductors with a Gaussian density of states. *Phys. Rev. B*, 84(16):165210, 2011.
165. S. L. M. van Mensfoort and R. Coehoorn. Determination of injection barriers in organic semiconductor devices from capacitance measurements. *Phys. Rev. Lett.*, 100(8):086802, 2008.
166. K. A. Neyts. Simulation of light emission from thin-film microcavities. *J. Opt. Soc. Am. A*, 15(4):962–971, 1998.
167. V. Bulovi, V. B. Khalfin, G. Gu, P. E. Burrows, D. Z. Garbuzov, and S. R. Forrest. Weak microcavity effects in organic light-emitting devices. *Phys. Rev. B*, 58(7):3730–3740, 1998.
168. Z. B. Wang, M. G. Helander, X. F. Xu, D. P. Puzzo, J. Qiu, M. T. Greiner, and Z. H. Lu. Optical design of organic light emitting diodes. *J. Appl. Phys.*, 109(5):053107, 2011.
169. M. Furno, R. Meerheim, S. Hofmann, B. Luessem, and K. Leo. Efficiency and rate of spontaneous emission in organic electroluminescent devices. *Phys. Rev. B*, 85(11):115205, 2012.
170. S. L. M. van Mensfoort, M. Carvelli, M. Megens, D. Wehenkel, M. Bartyzel, H. Greiner, R. A. J. Janssen, and R. Coehoorn. Measuring the light emission profile in organic light-emitting diodes with nanometre spatial resolution. *Nat. Photon.*, 4(0):329–335, 2010.
171. M. Flaemmich, D. Michaelis, and N. Danz. Accessing OLED emitter properties by radiation pattern analyses. *Org. Electron.*, 12(1):83–91, 2011.
172. J. Frischeisen, D. Yokoyama, C. Adachi, and W. Bruetting. Determination of molecular dipole orientation in doped fluorescent organic thin films by photoluminescence measurements. *Appl. Phys. Lett.*, 96(7):073302, 2010.
173. S.-Y. Kim and J.-J. Kim. Outcoupling efficiency of organic light emitting diodes and the effect of ITO thickness. *Org. Electron.*, 11(6):1010–1015, 2010.
174. W. Bruetting, J. Frischeisen, T. D. Schmidt, B. J. Scholz, and C. Mayr. Device efficiency of organic light-emitting diodes: Progress by improved light outcoupling. *Phys. Status Solidi A*, 210(1):44–65, 2013.
175. M. C. Gather and S. Reineke. Recent advances in light outcoupling from white organic light-emitting diodes. *J. Photon. Energy*, 5(1):057607–057607, 2015.
176. H. van Eersel, P. A. Bobbert, R. A. J. Janssen, and R. Coehoorn. Monte Carlo study of efficiency roll-offof phosphorescent organic light-emitting diodes: Evidence for dominant role of triplet–polaron quenching. *Appl. Phys. Lett.*, 105(14):143303, 2014.
177. R. Coehoorn, H. van Eersel, P. Bobbert, and R. Janssen. Kinetic Monte Carlo study of the sensitivity of OLED efficiency and lifetime to materials parameters. *Adv. Funct. Mater.*, 25(13):2024–2037, 2015.
178. Y. Shen and N. C. Giebink. Monte Carlo simulations of nanoscale electrical in-homogeneity in organic light-emitting diodes and its impact on their efficiency and lifetime. *Phys. Rev. Appl.*, 4(5):054017, 2015.
179. G. Lanzani. *The Photophysics behind Photovoltaics and Photonics*. Weinheim: Wiley-VCH, 2012. OCLC: ocn768072884.
180. C. Murawski, K. Leo, and M. C. Gather. Efficiency roll-off in organic light-emitting diodes. *Adv. Mater.*, 25(47):6801–6827, 2013.
181. M. Mesta, M. Carvelli, R. J. de Vries, H. van Eersel, J. J. M. van der Holst, M. Schober, M. Furno, B. Luessem, K. Leo, P. Loebl, R. Coehoorn, and P. A. Bobbert. Molecular-scale simulation of electroluminescence in a multilayer white organic light-emitting diode. *Nat. Mater.*, 12(7):652–658, 2013.
182. H. van Eersel, P. A. Bobbert, and R. Coehoorn. Kinetic Monte Carlo study of triplet–triplet annihilation in organic phosphorescent emitters. *J. Appl. Phys.*, 117(11):115502, 2015.

183. J. Cottaar, R. Coehoorn, and P. A. Bobbert. Modeling of charge transport across disordered organic heterojunctions. *Org. Electron.*, 13(4):667–672, 2012.
184. H. Yoshida and K. Yoshizaki. Electron affinities of organic materials used for organic light-emitting diodes: A low-energy inverse photoemission study. *Org. Electron.*, 20:24–30, 2015.
185. N. C. Giebink and S. R. Forrest. Quantum efficiency roll-off at high brightness in fluorescent and phosphorescent organic light emitting diodes. *Phys. Rev. B*, 77(23):235215, 2008.
186. G. Schwartz, S. Reineke, T. Conrad Rosenow, K. Walzer, and K. Leo. Triplet harvesting in hybrid white organic light-emitting diodes. *Adv. Funct. Mater.*, 19(9):1319–1333, 2009.
187. S. Mladenovski, S. Reineke, and K. Neyts. Measurement and simulation of exciton decay times in organic light-emitting devices with different layer structures. *Opt. Lett.*, 34(9):1375–1377, 2009.
188. K.-C. Tang, K. Lin Liu, and I.-C. Chen. Rapid intersystem crossing in highly phosphorescent iridium complexes. *Chem. Phys. Lett.*, 386(4–6):437–441, 2004.
189. Y. Kawamura, J. Brooks, J. J. Brown, H. Sasabe, and C. Adachi. Intermolecular interaction and a concentration-quenching mechanism of phosphorescent Ir(III) complexes in a solid film. *Phys. Rev. Lett.*, 96(1):017404, 2006.
190. F. S. Steinbacher, R. Krause, A. Hunze, and A. Winnacker. Triplet exciton transfer mechanism between phosphorescent organic dye molecules. *Phys. Stat. Sol. (A)*, 209(2):340–346, 2012.
191. S. L. M. van Mensfoort, R. J. de Vries, V. Shabro, H. P. Loebl, R. A. J. Janssen, and R. Coehoorn. Electron transport in the organic small-molecule material BAlq—The role of correlated disorder and traps. *Org. Electron.*, 11(8):1408–1413, 2010.
192. M. M. Mandoc, B. de Boer, G. Paasch, and P. W. M. Blom. Trap-limited electron transport in disordered semiconducting polymers. *Phys. Rev. B*, 75(19):193202, 2007.
193. F. May, B. Baumeier, C. Lennartz, and D. Andrienko. Can lattice models predict the density of states of amorphous organic semiconductors? *Phys. Rev. Lett.*, 109(13):136401, 2012.
194. S. Olthof, S. Mehraeen, S. K. Mohapatra, S. Barlow, V. Coropceanu, J.-L. Bredas, S. R. Marder, and A. Kahn. Ultralow doping in organic semiconductors: Evidence of trap filling. *Phys. Rev. Lett.*, 109(17):176601, 2012.
195. G. Li, C. H. Kim, Z. Zhou, J. Shinar, K. Okumoto, and Y. Shirota. Combinatorial study of exciplex formation at the interface between two wide band gap organic semiconductors. *Appl. Phys. Lett.*, 88(25):253505, 2006.
196. H. Aziz and Z. D. Popovic. Degradation phenomena in small-molecule organic light-emitting devices. *Chem. Mater.*, 16(23):4522–4532, 2004.
197. F. So and D. Kondakov. Degradation mechanisms in small-molecule and polymer organic light-emitting diodes. *Adv. Mater.*, 22(34):3762–3777, 2010.
198. J.-S. Park, H. Chae, H. Kyoon Chung, and S. I. Lee. Thin film encapsulation for flexible AM-OLED: A review. *Semicond. Sci. Technol.*, 26(3):034001, 2011.
199. S. Scholz, D. Kondakov, B. Luessem, and K. Leo. Degradation mechanisms and reactions in organic light-emitting devices. *Chem. Rev.*, 115(16):8449–8503, 2015.
200. H. Aziz, Z. D. Popovic, N.-X. Hu, A.-M. Hor, and G. Xu. Degradation mechanism of small molecule-based organic light-emitting devices. *Science*, 283(5409):1900–1902, 1999.
201. D. Y. Kondakov, W. C. Lenhart, and W. F. Nichols. Operational degradation of organic light-emitting diodes: Mechanism and identification of chemical products. *J. Appl. Phys.*, 101(2):024512, 2007.
202. R. Seifert, S. Scholz, B. Luessem, and K. Leo. Comparison of ultraviolet- and charge-induced degradation phenomena in blue fluorescent organic light emitting diodes. *Appl. Phys. Lett.*, 97(1):013308, 2010.
203. N. C. Giebink, B. W. D'Andrade, M. S. Weaver, P. B. Mackenzie, J. J. Brown, M. E. Thompson, and S. R. Forrest. Intrinsic luminance loss in phosphorescent small-molecule organic light emitting devices due to bimolecular annihilation reactions. *J. Appl. Phys.*, 103(4):044509, 2008.

204. T. D. Schmidt, L. Jaeger, Y. Noguchi, H. Ishii, and W. Bruetting. Analyzing degradation effects of organic light-emitting diodes via transient optical and electrical measurements. *J. Appl. Phys.*, 117(21):215502, 2015.
205. T. D. Schmidt, D. S. Setz, M. Flaemmich, J. Frischeisen, D. Michaelis, B. C. Krummacher, N. Danz, and W. Bruetting. Evidence for non-isotropic emitter orientation in a red phosphorescent organic light-emitting diode and its implications for determining the emitter's radiative quantum efficiency. *Appl. Phys. Lett.*, 99(16):163302, 2011.
206. C. Fry, B. Racine, D. Vaufrey, H. Doyeux, and S. Cin. Physical mechanism responsible for the stretched exponential decay behavior of aging organic light-emitting diodes. *Appl. Phys. Lett.*, 87(21):213502, 2005.
207. A. Mass, R. Coehoorn, and P. A. Bobbert. Universal size-dependent conductance fluctuations in disordered organic semiconductors. *Phys. Rev. Lett.*, 113(11), 2014.

16

Tunnel-Junction Light-Emitting Diodes

Yen-Kuang Kuo
Jih-Yuan Chang
Ya-Hsuan Shih
Fang-Ming Chen
and
Miao-Chan Tsai

16.1 Introduction 523
Polarization-Assisted Tunneling Effect • Efficiency Droop in III-Nitride LEDs
16.2 Physical Models and Simulation Parameters 524
Bulk Band Structure • MQW Model: Approximation of Effective Mass • Incomplete Ionization of Impurities • Electronic Polarization and Interface Charges • Interband Tunneling Model
16.3 III-Nitride Tunnel Junction 528
16.4 Tunnel-Junction LED 530
Single LED • Blue Tunnel-Junction LED • Green Tunnel-Junction LED
16.5 Conclusion 536

16.1 Introduction

The use of p-n tunnel junctions for applications in GaN-based electronic and optoelectronic devices has become increasingly attractive. The use of tunnel junctions has conceptually been desired for the reuse of carriers for coupled active regions, enabling high quantum efficiencies and improved vertical transport (Ozden et al., 2001). Concepts of the tunnel field-effect transistors, multiple active region light-emitting diodes (LEDs) and laser diodes, high-conductivity hole injection layers, and multijunction solar cells may be realized with the III-nitride material system providing that a sufficiently low-resistivity tunnel junction can be obtained. However, the feasibility of forming low-resistivity tunnel junctions in the wide-bandgap III-nitride devices has been a challenge due to the high hole and electron concentrations required for band alignment. Substantial effort is still required to resolve this critical issue.

16.1.1 Polarization-Assisted Tunneling Effect

The probability of interband tunneling across a potential barrier is mainly governed by the tunneling barrier height determined by the bandgap and the tunneling barrier thickness (Simon et al., 2009). For wide-bandgap III-nitrides, tunneling is difficult due to the high barrier height and the difficulty in achieving degenerate p-type impurity doping (Simon et al., 2009; Krishnamoorthy et al., 2010). However, the strong polarization-induced electric field available in the III-nitrides and other highly polar semiconductor materials provides a new design approach for tunneling structures. In hetero-structures of highly polar materials, the polarization-induced sheet charges can create significantly high electric fields resulting in large band bending over a small distance. The insufficient band-bending for interband tunneling from doping-induced built-in electric field (E_{bi}) can thus be supplemented by the polarization-induced electric field (E_p) assuming that the two electric fields are oriented in the same direction. The probability of interband tunneling can therefore be increased.

Recently, GaN-based tunnel junctions, which are supplemented by the polarization-induced electric field, have been demonstrated (Simon et al., 2009; Krishnamoorthy et al., 2010, 2011, 2013; Schubert, 2010a; Grundmann and Mishra, 2007). The use of large spontaneous and piezoelectric polarization fields present in the III-nitrides enables the formation of thin depletion region and relaxes the strict requirement for degenerate doping. Structures including the GaN/AlN/GaN, AlN/GaN/AlN, and GaN/InGaN/GaN have been investigated theoretically and experimentally (Simon et al., 2009; Krishnamoorthy et al., 2010; Schubert, 2010a). In particular, tunnel-junction structures using an intermediate InGaN layer have been developed due to the advantages of low bandgap and large piezoelectric polarization. It was shown by estimation of tunneling probability using the Wentzel–Kramers–Brillouin (WKB) approximation that the optimal performance, with n-doping $N_D = 5\times10^{19}$ cm^{-3} and p-doping $N_A = 1\times10^{19}$ cm^{-3}, would occur when the indium composition in InGaN is greater than 30% and the thickness is larger than 3 nm, which was achieved on the N-polar GaN with molecular beam epitaxy (MBE) growth method (Grundmann and Mishra, 2007; Krishnamoorthy et al., 2011, 2013). In many of the envisioned applications in optoelectronics, an indium composition of greater than approximately 20% in InGaN is not favorable due to the high optical absorption caused by the InGaN and the problem in crystal growth for the layers subsequent to the tunnel junction. Moreover, the doping levels are limited by the Si-induced roughening and strain, ionization efficiency, and solubility of Mg dopants. In this chapter, design and analysis for the low-resistivity tunnel junctions, which are with adequate indium composition and doping density for a typical Ga-polar InGaN multiple-quantum well (MQW) LED structure, are introduced.

16.1.2 Efficiency Droop in III-Nitride LEDs

For the application in solid-state lighting, the development of high-efficiency and high-power III-nitride LEDs is required. In this demand, the major issue that needs to be resolved might be the problem of efficiency droop, i.e., the reduction in illumination efficiency at high current density. It is believed that heating is not the major cause for efficiency droop due to the fact that the efficiency droop occurs in both pulsed and continuous wave (CW) conditions and the droop becomes severe with the decrease of ambient temperature (Laubsch et al., 2009; Kim et al., 2007). To date, the physical origin of efficiency droop remains debatable even though numerous possible physical mechanisms, such as the carrier delocalization (Monemar and Sernelius, 2007; Chichibu et al., 2006), Auger recombination (Shen et al., 2007; Delaney et al., 2009), insufficient hole injection (Rozhansky and Zakheim, 2006), carrier leakage (Shim et al., 2012; Lin et al., 2012), and polarization effect (Kim et al., 2007; Kuo et al., 2009), have been proposed and demonstrated to be related to the efficiency droop in GaN-based LEDs. Even though continuous efforts have been made, an overall solution for the efficiency droop of III-nitride LEDs is still lacking. In this chapter, instead of trying to promote the quantum efficiency at high current density, another approach in circumventing the issue of efficiency droop is proposed, which is based on the use of GaN-based tunnel junctions. Specifically, in this approach it is suggested to operate the III-nitride LEDs within the range of high efficiency, i.e., to operate the LEDs at low current density, with the insertion of a low-loss tunnel junction between two active regions, which is beneficial in reducing the carrier density in each active region and mitigating the problem of efficiency droop.

16.2 Physical Models and Simulation Parameters

16.2.1 Bulk Band Structure

The characteristics of III-nitride tunnel junction and tunnel-junction LEDs constructed along the *c*-axis are numerically studied using a self-consistent simulation program Advanced Physical Models of Semiconductor Devices (APSYS). APSYS employs the 6×6 $\mathbf{k}\cdot\mathbf{p}$ model, which was developed for the strained wurtzite semiconductor by Chuang and Chang (1996), to calculate the energy-band structures of the nonactive bulk regions. The unstrained bandgap energies of the $\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}$ and $\mathrm{Al}_y\mathrm{Ga}_{1-y}\mathrm{N}$ ternary alloys can be expressed by the following formula (Vurgaftman et al., 2001)

$$E_\mathrm{g}(\mathrm{In}_x\mathrm{Ga}_{1-x}\mathrm{N}) = x \cdot E_\mathrm{g}(\mathrm{GaN}) + (1-x) \cdot E_\mathrm{g}(\mathrm{InN}) - x \cdot (1-x) \cdot B(\mathrm{InGaN}), \tag{16.1}$$

$$E_\mathrm{g}(\mathrm{Al}_y\mathrm{Ga}_{1-y}\mathrm{N}) = y \cdot E_\mathrm{g}(\mathrm{GaN}) + (1-y) \cdot E_\mathrm{g}(\mathrm{AlN}) - y \cdot (1-y) \cdot B(\mathrm{AlGaN}), \tag{16.2}$$

where $E_\mathrm{g}(\mathrm{GaN})$, $E_\mathrm{g}(\mathrm{InN})$, and $E_\mathrm{g}(\mathrm{AlN})$ are the bandgap energies of GaN, InN, and AlN, which have values of 3.42, 0.64, and 6.0 eV at 300 K, respectively (Nepal et al., 2005; Wu et al., 2003). $B(\mathrm{InGaN})$ and $B(\mathrm{AlGaN})$ are the bandgap bowing parameters of InGaN and AlGaN, respectively. The bandgap bowing parameters of InGaN and AlGaN are assumed to be 2.1 and 1.0 eV, respectively (Gorczyca et al., 2011a, b). The band-offset ratio is assumed to be 0.7/0.3 for the III-nitride material systems.

16.2.2 MQW Model: Approximation of Effective Mass

A model of effective mass for the computation of density of states (DOS) and for a simplified MQW model is needed because, in general, the dispersion of bulk valence bands is nonparabolic and anisotropic with strong mixing or anticrossing behavior in the direction perpendicular to the c-axis. Based on the method of Chuang and Chang (1996), an analytical model of effective mass is implemented for the valence band.

Within a range of small k, a situation when the valence band is lightly populated by holes, the following effective masses hold:

$$m_\mathrm{hh}^\mathrm{z} = -m_0(A_1 + A_3)^{-1}, \tag{16.3}$$

$$m_\mathrm{lh}^\mathrm{z} = -m_0\left[A_1 + \left(\frac{E_2^0 - \lambda_\mathrm{e}}{E_2^0 - E_3^0}\right)A_3\right]^{-1}, \tag{16.4}$$

$$m_\mathrm{ch}^\mathrm{z} = -m_0\left[A_1 + \left(\frac{E_3^0 - \lambda_\mathrm{e}}{E_3^0 - E_2^0}\right)A_3\right]^{-1}, \tag{16.5}$$

$$m_\mathrm{hh}^\mathrm{t} = -m_0(A_2 + A_4)^{-1}, \tag{16.6}$$

$$m_\mathrm{lh}^\mathrm{t} = -m_0\left[A_2 + \left(\frac{E_2^0 - \lambda_\mathrm{e}}{E_2^0 - E_3^0}\right)A_4\right]^{-1}, \tag{16.7}$$

$$m_\mathrm{ch}^\mathrm{t} = -m_0\left[A_2 + \left(\frac{E_3^0 - \lambda_\mathrm{e}}{E_3^0 - E_2^0}\right)A_4\right]^{-1}, \tag{16.8}$$

where E_i^0 $(i = 1, 2, 3)$ are the valence-band edge (values) at $k = 0$.

For a large range of k, i.e., a situation when the valence band is heavily populated by holes, the following effective masses are valid:

$$m_\mathrm{hh}^\mathrm{z} = -m_0(A_1 + A_3)^{-1}, \tag{16.9}$$

$$m_\mathrm{lh}^\mathrm{z} = -m_0(A_1 + A_3)^{-1}, \tag{16.10}$$

$$m_\mathrm{ch}^\mathrm{z} = -m_0 A_1^{-1}, \tag{16.11}$$

$$m_\mathrm{hh}^\mathrm{t} = -m_0(A_2 + A_4 - A_5)^{-1}, \tag{16.12}$$

$$m_\mathrm{lh}^\mathrm{t} = -m_0(A_2 + A_4 - A_5)^{-1}, \tag{16.13}$$

$$m_\mathrm{ch}^\mathrm{t} = -m_0 A_2^{-1}. \tag{16.14}$$

In the present simulation, a compromise of the above two models is used, i.e., to average them. Note that the choice of model is based on how heavily the valence bands are populated by holes. Other parameters of band structure for the binary nitride wurtzite semiconductors employed in the simulation are listed in Table 16.1 (Wu, 2009; Vurgaftman and Meyer, 2007).

16.2.3 Incomplete Ionization of Impurities

The degree of ionization is described by the occupancies f_D and f_A. Shallow impurities are assumed to be in equilibrium with the local carriers and therefore the occupancy of the shallow impurities can be described by

$$f_\mathrm{D} = \frac{1}{1 + g_\mathrm{d}^{-1} \exp[(E_\mathrm{D} - E_\mathrm{fn})/kT]}, \tag{16.15}$$

$$f_\mathrm{A} = \frac{1}{1 + g_\mathrm{a} \exp[(E_\mathrm{A} - E_\mathrm{fp})/kT]}, \tag{16.16}$$

where the subscripts D and A are used to denote shallow donors and acceptors, respectively. The degeneracy levels are $g_\mathrm{d} = 2$ and $g_\mathrm{a} = 4$ (Tiwari, 1992).

TABLE 16.1 Material Parameters of the Binary Semiconductors GaN, AlN, and InN (Wu, 2009; Vurgaftman and Meyer, 2007)

Parameter	Symbol (Unit)	GaN	AlN	InN
Electron effective mass (*c*-axis)	$m_c^z\ (\mathrm{m}_0)$	0.21	0.33	0.068
Electron effective mass (transverse)	$m_c^t\ (\mathrm{m}_0)$	0.19	0.32	0.065
Hole effective mass parameter	A_1	−7.21	−3.86	−8.21
	A_2	−0.44	−0.25	−0.68
	A_3	6.68	3.58	7.57
	A_4	−3.46	−1.32	−5.23
	A_5	−3.40	−1.47	−5.11
	A_6	−4.90	−1.64	−5.96
Spin–orbit split energy	Δ_so(eV)	0.017	0.019	0.005
Crystal-field split energy	Δ_cr(eV)	0.010	−0.169	0.040
Lattice constant *a*	a_0(Å)	3.189	3.112	3.533
Lattice constant *c*	c_0(Å)	5.185	4.982	5.693
Elastic stiffness constant	C_{11} (GPa)	390	396	223
Elastic stiffness constant	C_{12} (GPa)	145	137	115
Elastic stiffness constant	C_{13} (GPa)	106	108	92
Elastic stiffness constant	C_{33} (GPa)	398	373	224
Elastic stiffness constant	C_{44} (GPa)	105	116	48
Hydrost. deform. potential (*c*-axis)	a_z (eV)	−7.1	−3.4	−4.2
Hydrost. deform. potential (transverse)	a_t (eV)	−9.9	−11.8	−4.2
Shear deform. potential	D1 (eV)	−3.6	−2.9	−3.6
	D2 (eV)	1.7	4.9	1.7
	D3 (eV)	5.2	9.4	5.2
	D4 (eV)	−2.7	−4.0	−2.7
	D5 (eV)	−2.8	−3.3	−2.8
	D6 (eV)	−4.3	−2.7	−4.3
Dielectric constant	ε	8.9	8.5	10.5

Note: $\Delta_\mathrm{cr} = \Delta_1$, $\Delta_\mathrm{so} = 3\Delta_2 = 3\Delta_3$.

In the simulation, the acceptor activation energy (E_A) of Mg doped in InGaN is scaled linearly from 60 meV (InN) (Khan et al., 2007) to 166 meV (GaN) (Kumakura et al., 2003). The acceptor activation energy of Mg doped in AlGaN is scaled linearly from 166 meV (GaN) (Kumakura et al., 2003) to 510 meV (AlN) (Nam et al., 2003).

16.2.4 Electronic Polarization and Interface Charges

In typical III-nitride semiconductor devices, the device structure is usually constructed of ternary AlGaN and InGaN alloys. In order to consider the influences of internal polarization in tunnel junction and tunnel-junction LEDs, the method developed by Fiorentini et al. (2002) is employed to estimate the internal polarization in Ga-face configuration, which is represented by fixed surface charges at hetero-interfaces (Fiorentini et al., 2002). The composition of the compound determines the net polarization charges that remain at each hetero-interface. The nonlinear model described in the papers by Fiorentini et al. (2002) and Vurgaftman and Meyer (2003) is used here. The spontaneous polarization P_{sp} (Cm^{-2}) is calculated as

$$P_{sp}(\text{Al}_x\text{Ga}_{1-x}\text{N}) = -0.090 \cdot x - 0.034 \cdot (1-x) + 0.019 \cdot x \cdot (1-x), \tag{16.17}$$

$$P_{sp}(\text{In}_x\text{Ga}_{1-x}\text{N}) = -0.042 \cdot x - 0.034 \cdot (1-x) + 0.038 \cdot x \cdot (1-x). \tag{16.18}$$

For binary compounds, the piezoelectric polarization P_{pz} (Cm^{-2}) is given as the nonlinear function of the transverse strain ε_{xx} by

$$P_{pz}(\text{GaN}) = -0.918 \cdot \varepsilon_{xx} + 9.541 \cdot \varepsilon_{xx}^2, \tag{16.19}$$

$$P_{pz}(\text{InN}) = -1.373 \cdot \varepsilon_{xx} + 7.559 \cdot \varepsilon_{xx}^2, \tag{16.20}$$

$$P_{pz}(\text{AlN}) = -1.808 \cdot \varepsilon_{xx} - 7.888 \cdot \varepsilon_{xx}^2 \; (\varepsilon_{xx} > 0), \tag{16.21}$$

$$P_{pz}(\text{AlN}) = -1.808 \cdot \varepsilon_{xx} + 5.624 \cdot \varepsilon_{xx}^2 \; (\varepsilon_{xx} < 0). \tag{16.22}$$

The piezoelectric polarization of a specific ternary compound is linearly interpolated by the piezoelectric polarizations of the corresponding binary compounds.

The built-in electric fields of III-nitride devices obtained from experimental measurements are usually smaller than the values obtained from theoretical calculation, ranging from 20% to 80% (Piprek, 2007), due mainly to partial compensation of the polarization field by fixed defect and interface charges. In this study, the polarization-induced surface charge densities are assumed to be 50% of the values determined from theoretical calculation.

16.2.5 Interband Tunneling Model

Direct interband tunneling in a p-n junction between the valence and conduction bands in an external applied electric field is commonly referred to as a tunnel junction, or an Esaki junction, or a Zener tunneling diode under reverse bias. The probability of tunneling within the WKB approximation, following the derivation of band-to-band tunneling in the papers by Kane (1960), Duke (1969), and Moll (1964), is calculated as follows:

$$D = \exp(-2J) = P_0 \exp\left(\frac{-E_\perp}{E}\right), \tag{16.23}$$

where

$$J\left(E_{\parallel}\right) = \int_{x_2}^{x_1} \left| \left(\frac{2m^*}{\hbar^2}\right) \frac{\left[\left(E_g/2\right)^2 - \left(\varepsilon_c\right)^2\right]}{E_g} + E_{\perp} \right|^{1/2} dx, \tag{16.24}$$

$$P_0 = \exp\left[\frac{\pi(m^*)^{1/2}(E_g)^{3/2}}{2(2)^{1/2}qF\hbar}\right] = \exp\left(-\frac{E_g}{4\underline{E}}\right)$$
$$\underline{E} = \frac{(2)^{1/2}qF\hbar}{2\pi(m^*)^{1/2}(E_g)^{1/2}}, \tag{16.25}$$

$$m^* = \frac{2m_c m_v}{(m_c + m_v)}. \tag{16.26}$$

$E_{\parallel}$ and $E_{\perp}$ are the electron kinetic energies in the directions along and perpendicular to the tunneling direction, respectively. P_0 is the tunneling probability with a momentum of zero perpendicular to the x-direction. The electrons that can tunnel through the bandgap barrier are those electrons with perpendicular momentum near zero if $\underline{E}$ is small. $\underline{E}$ is a measure of significance of perpendicular momentum range and m^* is the effective tunneling mass.

16.3 III-Nitride Tunnel Junction

For the applications in Ga-polarity InGaN-based LEDs, with the use of InGaN as the barrier material between the heavily doped n- and p-type GaN layers, the vectors of the polarization and normal built-in fields have the same direction for an n-on-p tunnel diode, which helps the generation of the electric field required for tunneling. Furthermore, the reduced barrier height of InGaN also promotes the tunneling probability of carriers since the bandgap energy of InGaN is relatively low when compared to GaN. Figure 16.1 shows the schematic diagram of n-GaN/p-GaN and n-GaN/InGaN/p-GaN tunnel-junction structures. The illustrations for the directions of built-in field and polarization field are also indicated in Figure 16.1 for better reference. Here, let us begin with the n-GaN/InGaN/p-GaN tunnel-junction structure, which is composed of a heavily doped n-GaN layer, a 7-nm-thick InGaN barrier layer, and a heavily doped p-GaN layer (Krishnamoorthy et al., 2011), and calculate the tunneling current as a function of reverse bias voltage, as shown in Figure 16.2. In the simulation, the dopant concentrations of n- and p-type GaN layers are assumed to be identical, while the ionization energies for the computation of incomplete ionization of donors and acceptors are quite distinct, making the electron and hole concentrations also different. It is observed in Figure 16.2a that the tunneling current increases dramatically with the increase of indium composition. It is required that the Fermi levels are in the valence band and the conduction band

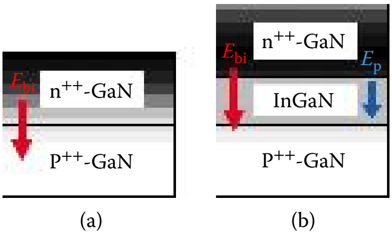


FIGURE 16.1 Schematic diagrams of (a) n-GaN/p-GaN and (b) n-GaN/InGaN/p-GaN tunnel-junction structures.

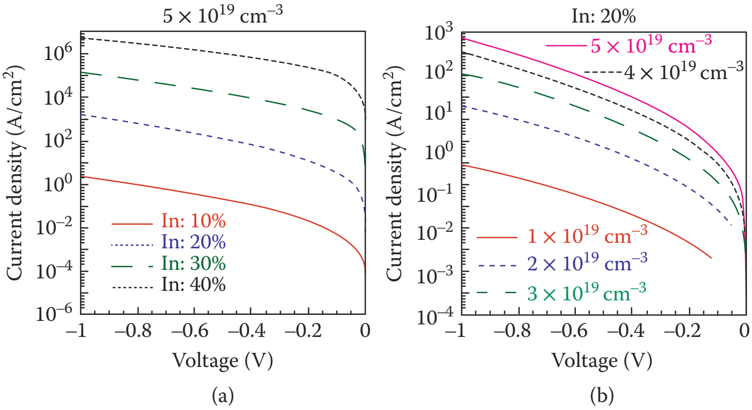


FIGURE 16.2 Electrical characteristics of the n-i-p tunnel diode in the reversed bias regime under various conditions of (a) indium composition and (b) acceptor and donor densities.

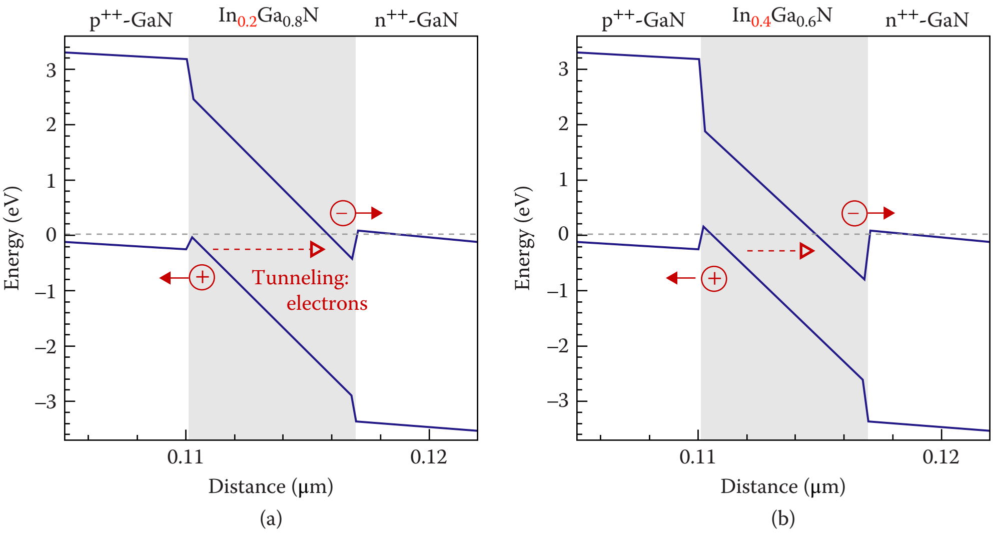


FIGURE 16.3 Energy band diagrams of (a) n-GaN/In$_{0.2}$Ga$_{0.8}$N/p-GaN and (b) n-GaN/In$_{0.4}$Ga$_{0.6}$N/p-GaN tunnel diodes at equilibrium.

on the p- and n-side, respectively, so that the carriers can tunnel between the bands as in a Zener diode. The improved tunneling characteristics with the increase of indium composition are thus ascribed to the promotion of the polarization field and the reduction of potential barrier height, as shown in Figure 16.3. However, it is noteworthy that when the composition of indium is high, the crystalline layer tends to relax and the film quality may deteriorate due to the large lattice mismatch and thermal mismatch between the InN and GaN. Under these circumstances, the strain-related polarization field reduces and the defect density increases, which is harmful for the applications in optoelectronics since the high defect density may act as the nonradiative recombination centers (Cherns et al., 2001). Moreover, for many applications in optoelectronics, it is usually required that the tunnel junction is with a reduced percentage of indium due to the problem of optical absorption. In the present study, the indium composition is fixed at 20% for subsequent simulations.

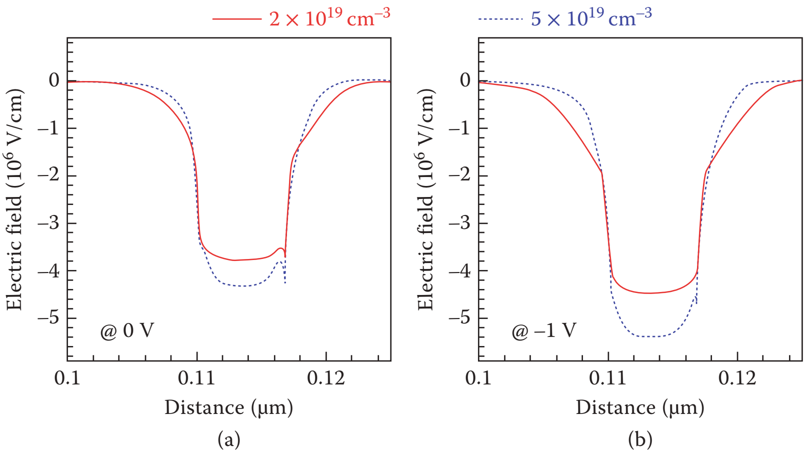


FIGURE 16.4 Electric fields of n-GaN/$In_{0.2}Ga_{0.8}N$/p-GaN tunnel diode with various acceptor and donor densities (a) at equilibrium and (b) at −1 V.

Another approach to improve the performance of tunnel junction is to increase the built-in field and reduce the depletion width by supporting more space charges through the increase of acceptor and donor densities of heavily doped p-type and n-type GaN layers. In Figure 16.4, it is found that the total electric field of the $In_{0.2}Ga_{0.8}N$ layer can be effectively enhanced when the acceptor and donor densities are increased from 2×10^{19} cm^{-3} to 5×10^{19} cm^{-3}. Consequently, the tunneling current increases with the increase of doping level, as shown in Figure 16.2b. Note that there is a limit for the above-mentioned improvement because the doping level of III-nitrides cannot be unlimitedly increased, especially for p-type doping. It is also noteworthy that the tunneling characteristics can be significantly promoted by increasing the Si doping level, while the tunneling current cannot be appreciably increased by increasing the Mg doping (Krishnamoorthy et al., 2013; Tsai et al., 2013). The findings provide another design approach that might be applicable for the realization of III-nitride tunnel junctions since heavy doping can be more easily achieved for the n-type materials than the p-type materials.

16.4 Tunnel-Junction LED

16.4.1 Single LED

Characterization of the blue single LED structure under study, which is used as a reference for the following investigations in stacking LEDs with tunnel junctions, is introduced first in this section. The LED structure is composed of a 3-μm-thick n-GaN layer (1×10^{19} cm^{-3}), four pairs of QWs with 3-nm-thick $In_{0.2}Ga_{0.8}N$ wells and 6-nm-thick GaN barriers, a 20-nm-thick p-$Al_{0.15}Ga_{0.85}N$ electron-blocking layer (EBL) (1×10^{19} cm^{-3}), and 100-nm-thick p-GaN layer (1×10^{19} cm^{-3}). The layer structure of the reference blue single LED is listed in Table 16.2. In the simulation, to focus on the topic explored and to eliminate the issue of current crowing, ideal ohmic contacts are assumed to cover the full top and bottom surfaces of the simulated vertical LED structure. The Shockley–Read–Hall (SRH) recombination lifetime and Auger recombination coefficient are set to be 50 ns and 2×10^{-30} cm^6/s, respectively (Yoshida et al., 2010; Piprek and Li, 2013). It is also assumed that the light extraction efficiency is 70%.

Figure 16.5 shows the simulated electrical and optical characteristics of the blue single LED. In the energy band diagram, as shown in Figure 16.5a, obvious band deformation caused by the polarization field is observed. In particular, the energy bands of QWs and barriers are severely sloped, forming triangular-shaped band profiles, which impact the transport, injection, and confinement of carriers in the MQW active region (Kim et al., 2007; Kuo et al., 2009). The polarization-induced sloped well results in spatial

TABLE 16.2 Layer Structure of the Reference Blue Single Light-Emitting Diode (LED) under Study

Parameter (unit)	*d*(nm)	N_{dop}(1/cm^3)
p-GaN	100	1×10^{19}
p-$Al_{0.15}Ga_{0.85}N$ (EBL)	20	1×10^{19}
i-GaN (barrier)	6	–
i-$In_{0.2}Ga_{0.8}N$ (well)	3	–
i-GaN (barrier)	6	–
i-$In_{0.2}Ga_{0.8}N$ (well)	3	–
i-GaN (barrier)	6	–
i-$In_{0.2}Ga_{0.8}N$ (well)	3	–
i-GaN (barrier)	6	–
i-$In_{0.2}Ga_{0.8}N$ (well)	3	–
i-GaN (barrier)	6	–
n-GaN	3000	1 × 1019

Note: *d*, layer thickness; N_{dop}, dopant density.

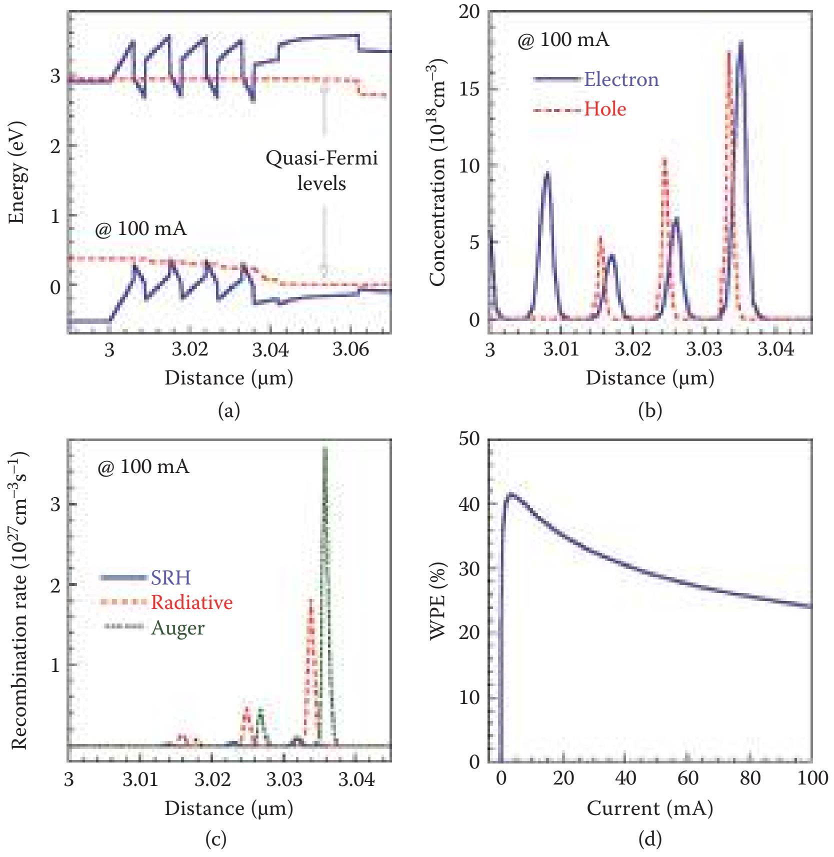


FIGURE 16.5 (a) Energy band diagram, (b) carrier concentrations, (c) recombination rates at 100 mA, and (d) wall-plug efficiency (WPE) of the reference blue single light-emitting diode (LED).

separation of electron and hole wave functions in the well, which is believed to be one of the key issues that deteriorate the radiative recombination. Moreover, the electron leakage current is negligibly small, which indicates that the capability of carrier confinement of the MQW active region is superior. As shown in Figure 16.5b, the distribution of carriers in the QWs is quite nonuniform. Most of the carriers accumulate in the QWs close to the p-side. Similar nonuniform carrier distribution in the InGaN LED has also been reported, which is attributed to the poor injection and transportation of holes (David et al., 2008). Under these circumstances, the illumination efficiency of the blue single LED is severely limited by the huge transition loss of Auger recombination. The droop of wall-plug efficiency (WPE), defined as $(\mathrm{WPE}_{\mathrm{max}} - \mathrm{WPE}_{100\mathrm{mA}})/\mathrm{WPE}_{\mathrm{max}}$ where $\mathrm{WPE}_{\mathrm{max}}$ is the maximum WPE and $\mathrm{WPE}_{100\mathrm{mA}}$ is the WPE at 100 mA, is as large as 42%, as shown in Figure 16.5c and d. Note that the WPE is referred to the energy conversion efficiency with which the LED converts the electrical power into optical power, as defined below:

$$\mathrm{WPE} = \frac{\text{Light output power}}{\text{Current} \times \text{Voltage}}. \tag{16.27}$$

The influence of Auger recombination on the efficiency of LEDs is typically analyzed based on the reduced rate equation. Assuming that the injection efficiency is 100%, the internal quantum efficiency (IQE), which is directly related to the light output power, can be expressed as

$$\mathrm{IQE} = \frac{B \cdot n^2}{A \cdot n + B \cdot n^2 + C \cdot n^3}, \tag{16.28}$$

where A, B, C, and n are the SRH coefficient, radiative coefficient, Auger coefficient, and carrier density, respectively. According to Equation 16.28, the Auger recombination dominates the total recombination rate at high carrier density because it scales with the cubic power of the free carrier density, which is believed to be one of the key factors contributing to the efficiency droop (Shen et al., 2007). If the carrier density in the active region can be reduced while the total carriers confined within the active region remains unchanged, the efficiency at high current may be improved due to the suppression of Auger recombination. It might be helpful to use more QWs in the active region because the carrier density per QW may be reduced under this situation. Figure 16.6 shows the simulated carrier concentrations and recombination rates of the single LED structure with 8 pairs and 12 pairs of QWs. It is observed that the carrier distribution of the LED structures with more QWs is still quite nonuniform and most of the carriers accumulate in the QWs close to the p-side. Therefore, the goal of reducing the carrier density in the QWs cannot be satisfactorily achieved. Under this circumstance, the structure with more QWs does not benefit from the suppressed Auger recombination and hence, similar radiative-Auger relations are observed for the 4-, 8-, and 12-QW LED structures as shown in Figure 16.6c and d. As a result, the efficiency characteristics do not exhibit much difference when the number of QWs changes.

16.4.2 Blue Tunnel-Junction LED

The monolithic structure with multiple LEDs stacked by tunnel junctions was introduced recently (Akyol et al., 2013; Piprek, 2014; Tsai et al., 2014). The stacking of multiple LED structures with tunnel junctions allows for the repeated use of electrons and holes for photon generation in each individual single LED. Ideally, if every individual LED in the stacking structure possesses identical carrier transport and recombination mechanism, the total photon number of the tunnel-junction LED should be approximately N times to that obtained from the single LED if the number of the unit LED in the stacking structure is N, provided that the tunnel junctions possess negligible loss. Therefore, an IQE of more than 100% is possible for the tunnel-junction LEDs. Note that, accompanying with the multiplication of output power, bias and input power multiply with the number of unit LED in the stacking structure simultaneously. Hence, the WPE of tunnel-junction LEDs remains below 100% (Piprek, 2014).

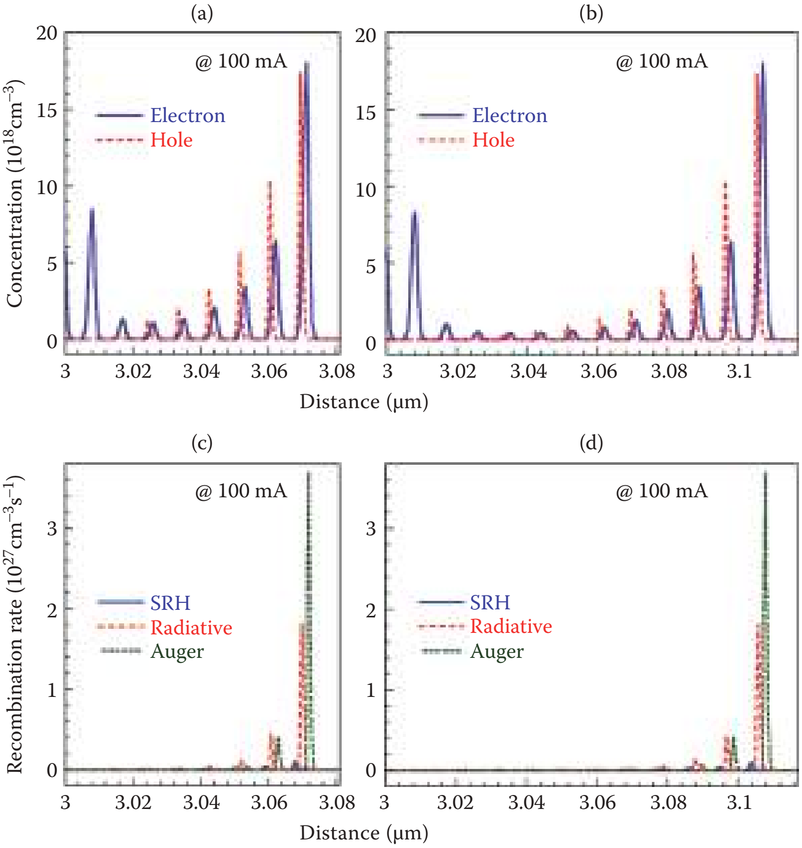
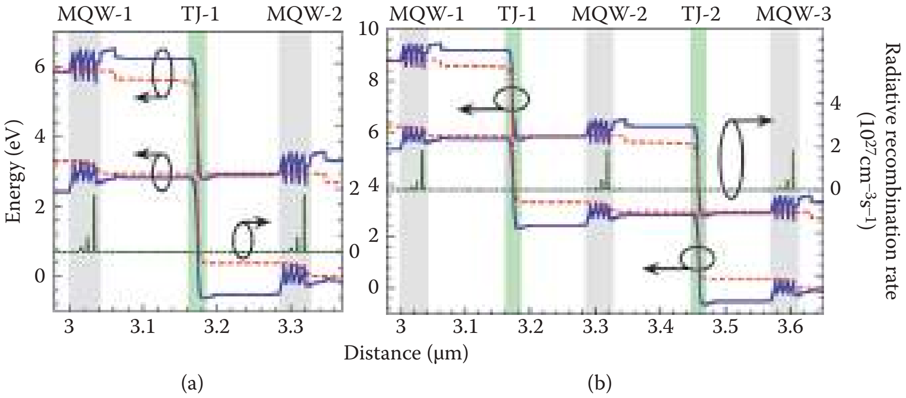


FIGURE 16.6 (a) and (b) Carrier concentrations and (c) and (d) recombination rates of the blue single light-emitting diode (LED) with 8 pairs and 12 pairs of QWs at 100 mA.

The simulated energy band diagram and radiative recombination rate at 100 mA for the blue tunnel-junction LEDs are depicted in Figure 16.7. Note that, in this figure, the n-type layer is located in the left-hand side of each unit LED. In the simulation, the n-GaN/$In_{0.2}Ga_{0.8}N$/p-GaN tunnel junction with a dopant concentration of 5×10^{19} cm^{-3} is utilized to ensure good tunneling characteristics, as shown in Figure 16.2c. When the tunnel-junction LED is turned on, electrons are injected and then transported in the conduction band of n-type layer of the first stacked LED. In the meantime, in the valance band of the p-type layer, electrons are transported and then transferred into the conduction band of the second stacked LED, which is regarded as holes moving in the opposite direction. The electrons and holes then recombine in the MQW active region of the first stacked LED to generate photons. The carrier transport and photon generation repeat in subsequent stacked LEDs. The emission profiles of all unit LEDs are almost identical, as shown in Figure 16.7.

Figure 16.8 shows the light output power as a function of current (L–I) and current as a function of voltage (I–V) characteristics of the blue single LED and tunnel-junction LEDs. It is observed that the light output power and operation voltage of the stacked LEDs are almost equivalent to the light output power and operation voltage of the blue single LED multiplied by the number of unit LED in the stacked LED structure. Under these circumstances, the WPE of the tunnel-junction LEDs would not vary too much comparing to that of the single LED. As mentioned previously, the approach of stacking LEDs with tunnel junctions is not intended to promote the quantum efficiency at high current density; instead, it finds a way to operate the LED devices at low current density that possesses relatively high IQE. Figure 16.9 shows the



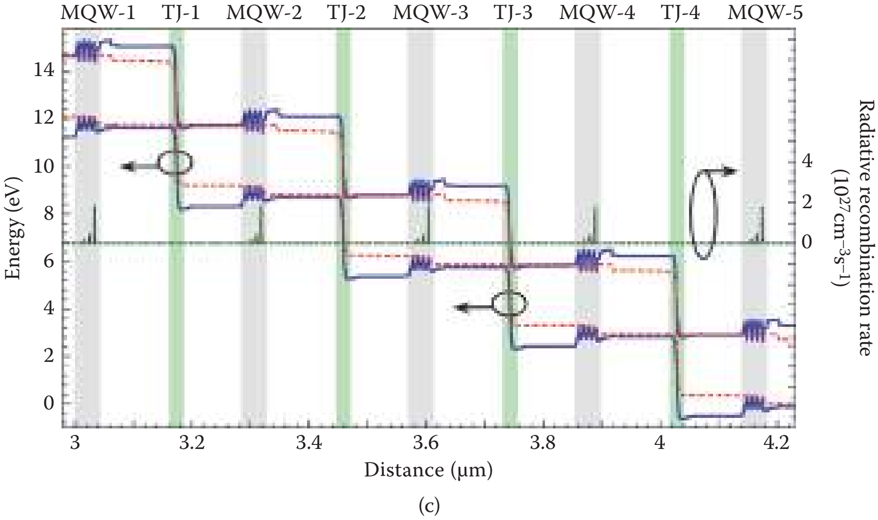


FIGURE 16.7 Energy band diagram and radiative recombination rate at 100 mA for the blue tunnel-junction light-emitting diodes (LEDs) with (a) two unit LEDs, (b) three unit LEDs, and (c) five unit LEDs.

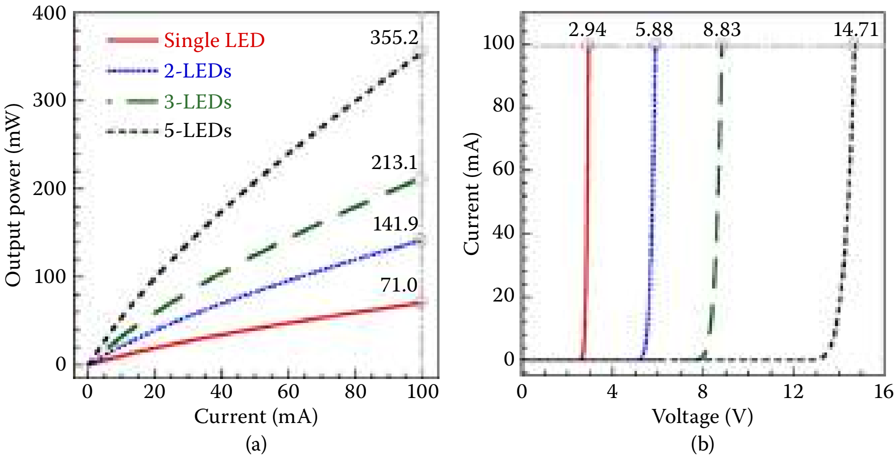


FIGURE 16.8 L–I and I–V characteristics of the blue single light-emitting diode (LED) and tunnel-junction LEDs.

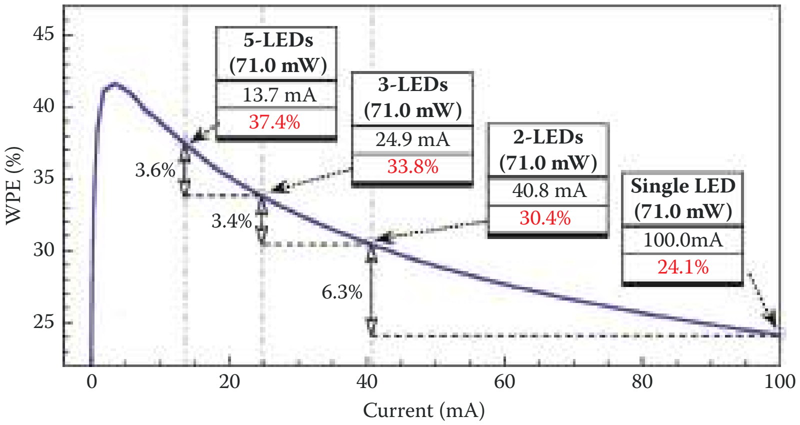


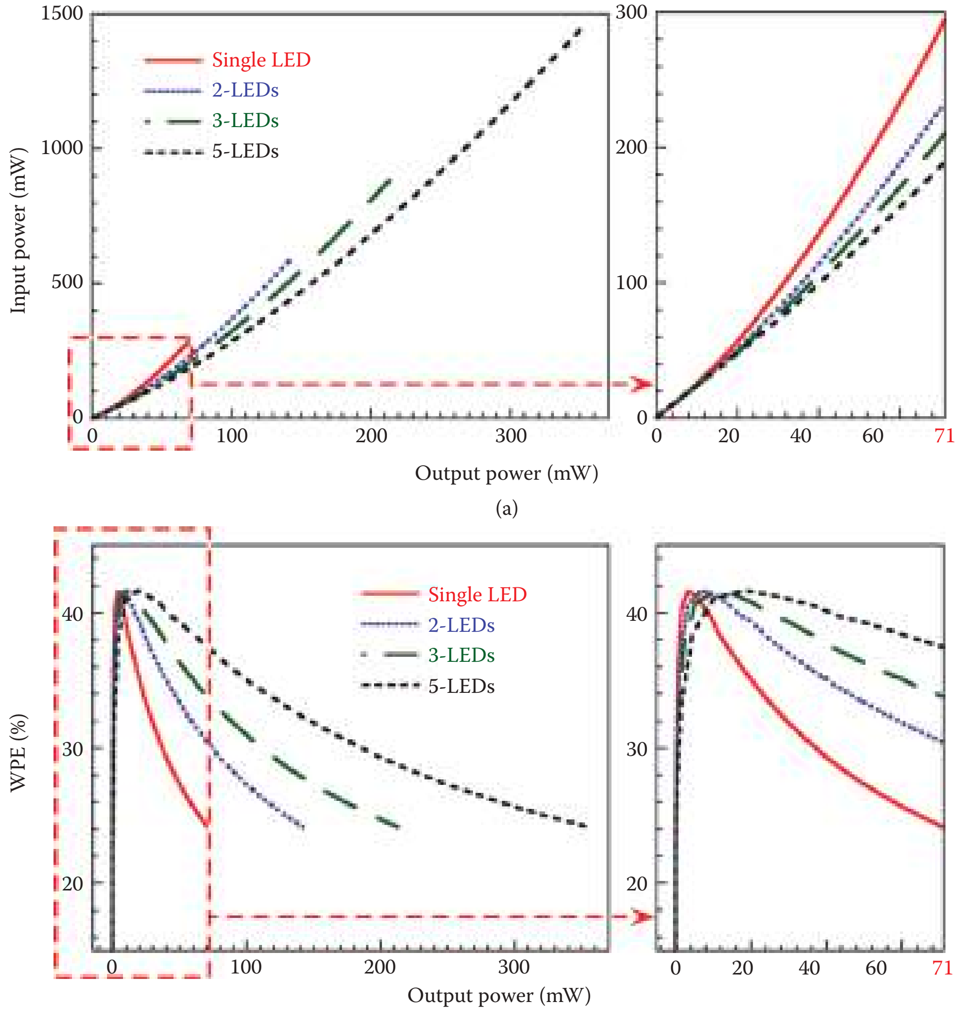
FIGURE 16.9 WPE of the blue light-emitting diode (LED) structures under study.

WPE of the blue LED structures under study. Only one curve is presented in this figure since all the WPE profiles are quite similar. For the single LED structure, the output power is 71 mW at 100 mA. For the tunnel-junction LEDs with two and three stacked unit LEDs, in order to obtain the same output power, the operation currents are 40.8 and 20.9 mA, and the WPE are 30.4% and 33.8%, respectively. As for the stacked structure of five unit LEDs, it can be operated at a relatively low current of 13.7 mA, which has a WPE of 37.4%. Consequently, if the tunnel-junction LED is stacked by more unit LEDs, the input power required to obtain a specific output power can be markedly reduced because of the relatively high WPE, as shown in Figure 16.10.

16.4.3 Green Tunnel-Junction LED

The concept of stacking identical LED structures with tunnel junctions is also investigated for green InGaN LEDs in this chapter to probe its effectiveness. The green single LED structure under study is identical to the blue LED except that the indium composition of InGaN wells is increased to 29%. Moreover, considering the enlarged lattice mismatch between the wells and barriers and the limited solubility of indium in GaN, the SRH recombination lifetime is reduced to 20 ns in the simulation. Figure 16.11 shows the energy band diagram, carrier concentrations, and recombination rates at 100 mA and WPE of the green single LED. The electrical and optical characteristics of the green single LED are quite similar to those of the blue single LED shown in Figure 16.5. However, since the indium composition of the InGaN wells in the green single LED is higher than that in the blue single LED, deeper QWs are observed in the band profile. Due to the deeper wells, the feature of nonuniform carrier distribution becomes much more severe, in which only the QW closest to the p-side contributes to interband transitions, as shown in Figure 16.11b and c. Under these circumstances, the droop of WPE in the green LED is more severe than that in the blue LED (48% versus 42%).

Since the holes almost accumulate in the last QW closest to the p-side, similar WPE is obtained no matter how many QWs are utilized in the green single LED under study, as shown in Figure 16.12a. Moreover, it was reported that the defect-related SRH recombination is an important interband transition loss that slashes the low current efficiency (Schubert et al., 2007). The WPEs of the green single LED structures with various SRH lifetimes are plotted in Figure 16.12b. The peak efficiency increases with the increase of SRH lifetime, which is similar to the results reported in the published literature (Schubert et al., 2007). The WPE droop deteriorates when the SRH lifetime increases, due mainly to the marked variation of peak efficiency. Since the high current efficiency remains almost unchanged, the high-power applications cannot benefit much from the improved film quality in the single LED. However, it is not the case for the tunnel-junction LED because the design concept for tunnel-junction LEDs is to operate the LEDs at low current



(a)

(b)

FIGURE 16.10 Required input power and corresponding wall-plug efficiency (WPE) of the blue single light-emitting diode (LED) and tunnel-junction LEDs as a function of the light output power.

density. Figure 16.13 shows the WPE of the green single LED and tunnel-junction LEDs when the SRH lifetimes are 20 and 50 ns. It is observed that the WPE of the tunnel-junction LED with 5 unit LEDs in the stacked structure is almost twice that of the single LED, especially when the SRH lifetime is high. The design strategy of tunnel-junction LED is thus applicable to the green LEDs. Since the efficiency of green LEDs is difficult to upgrade due to the nature of deeper InGaN wells, the idea of tunnel-junction LED may be of great benefit in improving the physical and optical performance.

16.5 Conclusion

In this chapter, several design approaches of the GaN-based polarization-assisted tunnel junctions are explored. With appropriate design via polarization engineering, low-resistivity and excellent tunneling characteristics can be achieved in the structures with practical indium composition and doping density. Note that the tunneling currents of the tunnel-junction structures obtained from experiments typically

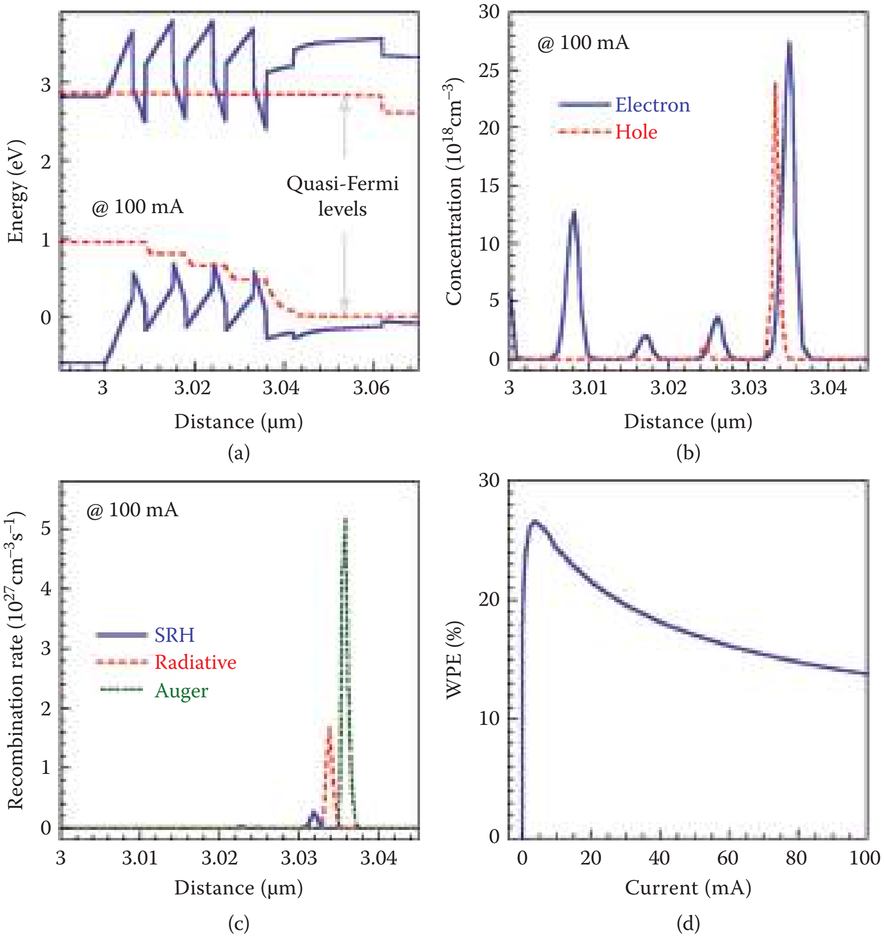


FIGURE 16.11 (a) Energy band diagram, (b) carrier concentrations, (c) recombination rates at 100 mA, and (d) wall-plug efficiency (WPE) of the green single light-emitting diode (LED).

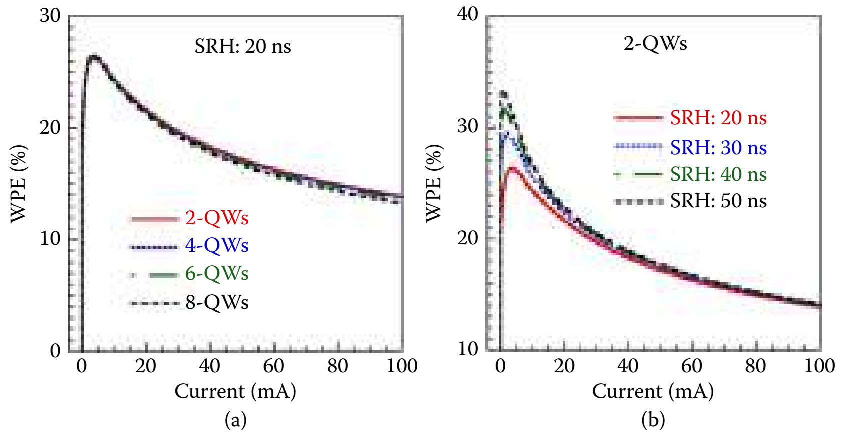


FIGURE 16.12 Wall-plug efficiency (WPE) of the green single light-emitting diode (LED) structures with (a) various pairs of QWs and (b) various Shockley–Read–Hall (SRH) lifetimes.

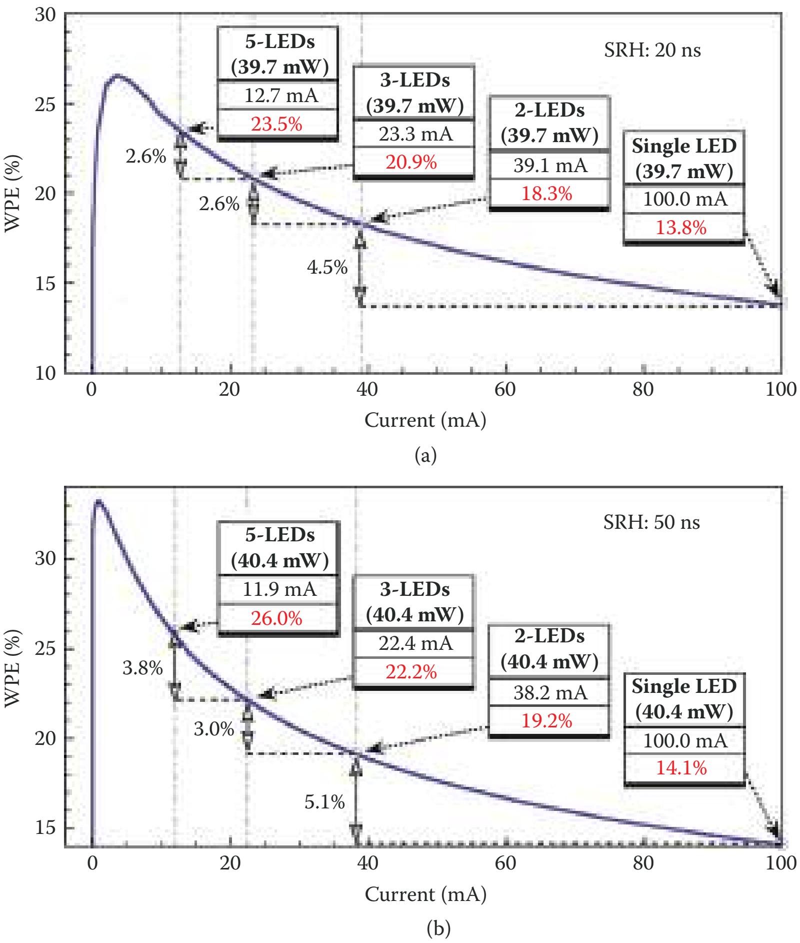


FIGURE 16.13 Wall-plug efficiency (WPE) of the green single light-emitting diode (LED) and tunnel-junction LEDs when the Shockley–Read–Hall (SRH) lifetimes are (a) 20 ns and (b) 50 ns.

exceed the values obtained from theoretical calculation. The additional tunneling current may be owing to the defect and impurity-related mid-gap electronic states (Schubert, 2010b), which are not considered in the present study. Hence, the utilization of tunnel junction in nitride-based applications is quite feasible, which enables the development of high-efficiency GaN-based devices, such as the multijunction solar cells, multijunction optoelectronic devices, and multijunction electronic devices. Specifically, in the present chapter, the application of low-resistivity tunnel junction in both the blue and green tunnel-junction LEDs is explored. Through the stacking of LEDs, it is shown that marked improvement in WPE can be achieved by operating the stacked LED devices at low current density. Furthermore, the tunnel-junction LED structure allows further design modification for better electrical and optical characteristics, such as the removal of the EBL and optimization in number of QWs. It is our belief that the method introduced in this chapter may be of great help in obtaining cost-effective and high-efficiency LED devices for solid-state lighting.

Acknowledgments

This work is supported by the Ministry of Science and Technology of Taiwan under grants NSC-99-2119-M-018-002-MY3 and NSC-102-2112-M-018-004-MY3.

References

Akyol F, Krishnamoorthy S, and Rajan S (2013) Tunneling-based carrier regeneration in cascaded GaN light emitting diodes to overcome efficiency droop. *Applied Physics Letters*. 103:081107-1—081107-4.

Cherns D, Henley SJ, and Ponce FA (2001) Edge and screw dislocations as nonradiative centers in InGaN/GaN quantum well luminescence. *Applied Physics Letters*. 78:2691–2693.

Chichibu SF et al. (2006) Origin of defect-insensitive emission probability in In-containing (Al,In,Ga)N alloy semiconductors. *Nature Materials*. 5:810–816.

Chuang SL and Chang CS (1996) k·p method for strained wurtzite semiconductors. *Physical Review B*. 54:2491–2504.

David A et al. (2008) Carrier distribution in (0001) InGaN/GaN multiple quantum well light-emitting diodes. *Applied Physics Letters*. 92: 053502-1–053502-3.

Delaney KT, Rinke P, and Van de Walle CG (2009) Auger recombination rates in nitrides from first principles. *Applied Physics Letters*. 94:191109-1–191109-3.

Duke CB (1969) *Tunneling in Solids*. New York, NY: Academic Press.

Fiorentini V, Bernardini F, and Ambacher O (2002) Evidence for nonlinear macroscopic polarization in III–V nitride alloy heterostructures. *Applied Physics Letters*. 80:1204–1206.

Gorczyca I, Suski T, Christensen NE, and Svane A (2011a) Band gap bowing in quaternary nitride semiconducting alloys. *Applied Physics Letters*. 98:241905-1–241905-3.

Gorczyca I, Suski T, Christensen NE, and Svane A (2011b) Size effects in band gap bowing in nitride semiconducting alloys. *Physical Review B*. 83:153301-1–153301-4.

Grundmann MJ and Mishra UK (2007) Multi-color light emitting diode using polarization-induced tunnel junctions. *Physica Status Solidi C*. 4:2830–2833.

Kane EO (1960) Zener tunneling in semiconductors. *Journal of Physics and Chemistry of Solids*. 12:181–188.

Khan N, Nepal N, Sedhain A, Lin JY, and Jiang HX (2007) Mg acceptor level in InN epilayers probed by photoluminescence. *Applied Physics Letters*. 91:012101-1–012101-3.

Kim MH et al. (2007) Origin of efficiency droop in GaN-based light-emitting diodes. *Applied Physics Letters*. 91:183507-1–183507-3.

Krishnamoorthy S, Akyol F, Park PS, and Rajan S (2013) Low resistance GaN/InGaN/GaN tunnel junctions. *Applied Physics Letters*. 102:113503-1–113503-5.

Krishnamoorthy S, Nath DN, Akyol F, Park PS, Esposto M, and Rajan S (2010) Polarization-engineered GaN/InGaN/GaN tunnel diodes. *Applied Physics Letters*. 97:203502-1–203502-3.

Krishnamoorthy S, Park PS, and Rajan S (2011) Demonstration of forward inter-band tunneling in GaN by polarization engineering. *Applied Physics Letters*. 99:233504-1–233504-3.

Kumakura K, Makimoto T, and Kobayashi N (2003) Mg-acceptor activation mechanism and transport characteristics in p-type InGaN grown by metal organic vapor phase epitaxy. *Journal of Applied Physics*. 93:3370–3375.

Kuo YK, Chang JY, Tsai MC, and Yen SH (2009) Advantages of blue InGaN multiple-quantum well light-emitting diodes with InGaN barriers. *Applied Physics Letters*. 95:011116-1–011116-3.

Laubsch A et al. (2009) On the origin of IQE-"droop" in InGaN LEDs. *Physica Status Solidi C*. 6:S913–S916.

Lin GB, Meyaard D, Cho J, Schubert EF, Shim H, and Sone C (2012) Analytic model for the efficiency droop in semiconductors with asymmetric carrier-transport properties based on drift-induced reduction of injection efficiency. *Applied Physics Letters*. 100:161106-1–161106-4.

Moll JL (1964) *Physics of Semiconductors*. New York, NY: McGraw-Hill.

Monemar B and Sernelius BE (2007) Defect related issues in the "current roll-off" in InGaN based light emitting diodes. *Applied Physics Letters*. 91:181103-1–181103-3.

Nam KB, Nakarmi ML, Li J, Lin JY, and Jiang HX (2003) Mg acceptor level in AlN probed by deep ultraviolet photoluminescence. *Applied Physics Letters*. 83:878–880.

Nepal N, Li J, Nakarmi ML, Lin JY, and Jiang HX (2005) Temperature and compositional dependence of the energy band gap of AlGaN alloys. *Applied Physics Letters*. 87:242104-1–242104-3.

Ozden I, Makarona E, Nurmikko AV, Takeuchi T, and Krames M (2001) A dual-wavelength indium gallium nitride quantum well light emitting diode. *Applied Physics Letters*. 79:2532–2534.

Piprek J, Li ZM, Farrell R, DenBaars SP, and Nakamura S (2007) Electronic properties of InGaN/GaN vertically-cavity lasers, In: *Nitride Semiconductor Devices: Principles and Simulation*, Piprek J (ed.), pp. 423–445. Weinheim: John Wiley & Sons.

Piprek J (2014) Origin of InGaN/GaN light-emitting diode efficiency improvements using tunnel-junction-cascaded active regions. *Applied Physics Letters*. 104:051118-1—051118-4.

Piprek J and Li SZM (2013) Origin of InGaN light-emitting diode efficiency improvements using chirped AlGaN multiquantum barriers. *Applied Physics Letters*. 102:023510-1–023510-4.

Rozhansky IV and Zakheim DA (2006) Analysis of the causes of the decrease in the electroluminescence efficiency of AlGaInN light-emitting-diode heterostructures at high pumping density. *Semiconductors*. 40:839–845.

Schubert MF (2010a) Interband tunnel junctions for wurtzite III-nitride semiconductors based on heterointerface polarization charges. *Physical Review B*. 81:035303-1–035303-8.

Schubert MF (2010b) Polarization-charge tunnel junctions for ultraviolet light emitters without p-type contact. *Applied Physics Letters*. 96:031102-1–031102-3.

Schubert MF et al. (2007) Effect of dislocation density on efficiency droop in GaInN/GaN light-emitting diodes. *Applied Physics Letters*. 91:231114-1–231114-3.

Shen YC et al (2007) Auger recombination in InGaN measured by photoluminescence. *Applied Physics Letters*. 91:141101-1–141101-3.

Shim JI et al. (2012) Efficiency droop in AlGaInP and GaInN light-emitting diodes. *Applied Physics Letters*. 100:111106-1–111106-4.

Simon J et al. (2009) Polarization-induced Zener tunnel junctions in wide-band-gap heterostructures. *Physical Review Letters*. 103:026801-1–026801-4.

Tiwari S (1992) *Compound Semiconductor Device Physics*. London: Academic Press.

Tsai MC, Leung B, Hsu TC, and Kuo YK (2013) Low resistivity GaN-based polarization-induced tunnel junctions. *Journal of Lightwave Technology*. 31:3575–3581.

Tsai MC, Leung B, Hsu TC, and Kuo YK (2014) Tandem structure for efficiency improvement in GaN based light-emitting diodes. *Journal of Lightwave Technology*. 32:1801–1806.

Vurgaftman I and Meyer JR (2003) Band parameters for nitrogen-containing semiconductors. *Journal of Applied Physics*. 94:3675–3696.

Vurgaftman I and Meyer JR (2007) Electron band structure parameters. In: *Nitride Semiconductor Devices: Principles and Simulation*, Piprek J (ed.), pp. 13–48. Weinheim: John Wiley & Sons.

Vurgaftman I, Meyer JR, and Ram-Mohan LR (2001) Band parameters for III–V compound semiconductors and their alloys. *Journal of Applied Physics*. 89:5815–5875.

Wu J (2009) When group-III nitrides go infrared: New properties and perspectives, *Journal of Applied Physics*. 106:011101-1–011101-27.

Wu J et al. (2003) Temperature dependence of the fundamental band gap of InN. *Journal of Applied Physics*. 94:4457–4460.

Yoshida H, Kuwabara M, Yamashita Y, Uchiyama K, and Kan H (2010) Radiative and nonradiative recombination in an ultraviolet GaN/AlGaN multiple-quantum-well laser diode. *Applied Physics Letters*. 96:211122-1–211122-3.

17

Quantum Disk Nanowire Light-Emitting Diodes

17.1 Introduction 541
17.2 Overview of Numerical Methods 542
Strain • Transport • Quantum Models
17.3 Strain Maps 543
17.4 Transport Properties 545
17.5 Quantum Calculations: EFA Models 545
17.6 Quantum Calculations: Empirical Tight-Binding Approach.... 547
17.7 Self-Consistent Calculations 549
Multiscale Simulation
17.8 Modeling of Surface States 552
17.9 Sensitivity to Geometrical and Material Parameters 553
17.10 Conclusions 556

Fabio Sacconi

17.1 Introduction

Due to their properties, nanowire (NW) based light-emitting diode (LED) structures offer very interesting advantages, such as defect-free material, due to small footprint on the substrate, which allows higher LED efficiency, and three-dimensional (3D) geometry features, such as lateral strain relaxation, which can also be exploited to increase efficiency (Ristic et al., 2005). Group III nitride semiconductors have attracted much attention for quite a long time especially for their light-emitting device applications. A new approach for reaching exceptionally high efficiencies of LEDs for the whole visible spectrum is therefore based on nanostructured InGaN emitters, where nanorods, which have been shown to be defect-free, serve as active light-emitting structures. Light-emitting nanorod arrays based on InGaN as a material for the active region have been investigated (Kim et al., 2004); in principle they could cover the whole visible spectrum from blue to red and be incorporated into a single device for phosphor-free white light emission. By controlling indium content in the InGaN quantum disk (QD) embedded in the NW, nanorod emitters are expected to exhibit efficient green, yellow, and red emission, allowing coverage of the full visible spectrum with GaN-based LEDs. This could overcome a major drawback of today's red–green–blue (RGB) light sources, that is, the lack of an efficient green emitter due to the efficiency gap existing between the blue InGaN and the red InGaAlP materials system (Piprek, 2010).

In this chapter, we discuss several simulation approaches for the numerical study of the transport and optoelectronic properties of GaN NW diode structures with an embedded InGaN QD. Nanostructure-based emerging electronic and optoelectronic devices, such as LEDs or photodetectors, need a fully

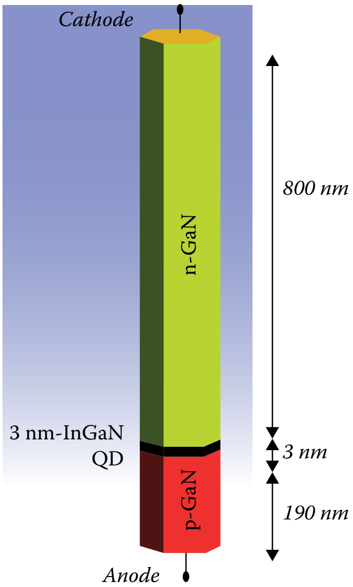


FIGURE 17.1 Nanowire (NW) quantum dot (QD) light-emitting diode (LED) model structure.

quantum mechanical or atomistic approaches to obtain a correct description of structural, electrical, and optical properties. Usually, the active region of a device, which needs such an advanced investigation, is small compared to the overall simulation domain. The computational cost of more accurate quantum mechanical models, however, makes their application to the whole device unfeasible. It is, therefore, necessary to implement a multiscale simulation approach, which couples the semiclassical models describing the bulk regions of the device to the quantum mechanical or atomistic models, acting only on the nanostructured regions of the device.

We see in the rest of the chapter how an integrated multiscale and multiphysics simulation environment (see Auf der Maur et al., 2011, 2013; Auf der Maur, 2015) may be capable of coupling different models on different scales, ranging from macroscopical to atomistic representations.

In the following, we first describe briefly the different physical models that need to be applied for an adequate description of these devices. Then we show the main results obtained from these numerical methods in the simulation of QD-based LEDs. If not stated otherwise, in the rest of the text we refer to the model NW LED structure with a single QD described in Figure 17.1. This model structure is a GaN NW usually around 1 μm high with a hexagonal base and a radius between 25 nm and several hundreds of nanometers. The embedded QD, which constitutes the LED active region, is a 3-nm-thick disk made of InGaN alloy, with In composition varying between 10% and 40%.

17.2 Overview of Numerical Methods

17.2.1 Strain

Strain has a critical influence on the behavior of heterostructures due to its effect on the band energies and the strain-induced piezoelectric polarization. The latter is particularly important in nitride-based devices. A straightforward approach for the calculation of strain in lattice mismatched heterostructures is the one based on the linear elasticity theory of solids. Usually, in this approach it is assumed that pseudomorphic interfaces are present between different materials (Povolotskyi and Di Carlo, 2006). One also assumes small deformations, such that the strain is a linear function of deformation and that Hooke's law, which linearly relates stress to strain, can be used. The strain and deformation field are found by minimizing the elastic energy of the system. As a result, one obtains the strain tensor in any point of the structure, the shape deformation, and the piezoelectric polarization, which to the first order depends linearly on strain. Self-consistent electromechanical simulations can be carried out by including the converse piezoelectric effect.

This continuum elastostatic model described here has several limitations.

In particular, it does not contain any information on internal strain, which can be very important, particularly in wurtzite crystals. A more accurate description of strain can be obtained by means of the valence force field (VFF) method, which offers the best compromise between accuracy and computational effort in the study of semiconductor nanostructures. One of the implementations is based on the Keating model (Keating, 1966), for diamond structures, as generalized by Camacho and Niquet (2010) to treat nonideal wurtzite structures. As VFF is computationally more intensive than the continuous media model, a possibly convenient implementation is the one that allows coupling of the two models in a multiscale simulation (see Auf der Maur et al., 2011, 2013).

17.2.2 Transport

Usually transport of electrons and holes in LED structures is treated in a semiclassical picture based on the drift–diffusion model (Sze, 1981). The particle fluxes are written in terms of the electrochemical potentials. The carrier statistics are given by Boltzmann or Fermi–Dirac statistics, assuming as usual local equilibrium. The conduction and valence band edges and effective masses may be obtained from bulk $\mathbf{k}\cdot\mathbf{p}$ calculations; in this way, the local corrections due to strain can be easily included.

To account for quantum effects, a correction to this purely classical approach is needed. Several methods can be applied, such as the density gradient quantum correction model (Ancona, 1990) up to the more computationally demanding nonequilibrium Green's function (NEGF) (Datta, 2000).

A self-consistent approach in many cases may be a good compromise. Some applications of self-consistent techniques will be described in the following.

17.2.3 Quantum Models

Quantum mechanical models are used for the calculation of eigenstates of confined particles in nanostructures, either based on the envelope function approximation (EFA) or on atomistic approaches. In the former, the Hamiltonian of the system is constructed in the framework of single-band and multiband $\mathbf{k}\cdot\mathbf{p}$ theory (Chuang, 1995; Chuang and Chang, 1996). The single particle wave functions are expanded in terms of bulk Bloch functions (usually taken at the zone center $\boldsymbol{k} = 0$ for direct band gap materials) of the constituent materials, leading to a system of equations for the envelope functions. The atomistic approaches are typically based on the empirical tight binding (ETB) method (Di Carlo, 2003). In this method, the electronic states are written as a linear combination of atomic orbitals (LCAO). In this case, an atomistic structure describing the heterostructure studied has to be generated according to the macroscopic device description and crystallographic orientation. The solution of the eigenvalue problems resulting from the EFA and ETB models provides the energy spectrum, the particle densities, and the probabilities of optical transitions. The particle densities are calculated by populating the electron and hole states according to the expectation value of the corresponding electrochemical potential. The particle densities may then be fed back to the Poisson/drift–diffusion model for a self-consistent Schrödinger–Poisson/drift–diffusion calculation.

17.3 Strain Maps

It is interesting to study the behavior of the strain tensor in an GaN NW embedding an InGaN QD. As previously discussed, strain induced by the lattice mismatch in a heterostructure can be calculated, for example, by means of an elasticity model. In the case of the QD NW LED, strain obtained from this numerical model is due to the lattice mismatch between the material of the NW or nanocolumn and the material that constitutes the QD. It has been shown (Hugues et al., 2013) that a GaN NW grown bottom-up becomes strain free after around 100 nm. Thus, we can expect that a typical GaN NW, with a height of 1 μm or more, is completely relaxed beyond the region surrounding the QD. This justifies the choice of the NW material (GaN in this case) as the substrate reference material for the strain calculations. This means that the components of the strain tensor as resulting from the minimization of the elastic energy in the system composed by the whole heterostructure will tend to vanish in the substrate material.

If we look at the calculated in-plane strain tensor component ε_{xx} in an NW, we clearly see a strong compressive strain inside the InGaN QD (Sacconi et al., 2010, 2012). In fact, the InN lattice constant is higher than that of the reference GaN lattice material; thus, a negative in-plane strain is determined in the pseudomorphic InGaN/GaN heterostructure. Besides, due to 3D effects, a tensile strain is present in the GaN barrier, close to the interface with the quantum well, while it vanishes in the bulk of the GaN NW. A peculiar feature of the NW structure is the strong reduction of the in-plane strain close to the column surface. The strain relaxation due to the surface boundary is visible in Figure 17.2 where ε_{xx} is shown, along the z-axis (growth direction), for two different radial positions: in the center of the QD ($r = 0$) and close to the surface boundary ($r = 25$). Strain is clearly higher in the center of the column, while it tends to vanish at the surface boundary. Piezoelectric polarization is obtained from the strain field and exhibits a similar behavior: polarization is higher at the center of the QD and tends to decrease toward the lateral surface of the NW. Piezoelectric polarization (together with spontaneous or pyroelectric polarization) has a direct effect on the band profiles and therefore on the quantum and optical properties.

It is interesting to analyze the dependence of the ε_{xx} strain tensor component in the QD on the column size. It is found (Sacconi et al., 2012) that, by increasing the NW size, the (tensile) strain in the barrier regions outside the QD tends to vanish, while the compressive strain slightly increases inside the QD. Moreover, it can clearly be seen that the effect of strain relaxation is more evident for columns with a radius lower than 50 nm and it tends to vanish for very large columns.

The dependence on QD thickness of the ε_{xx} strain tensor component in the NW is shown in Figure 17.3, for an $\text{In}_{0.1}$GaN QD in an NW with radius 25 nm. The value of ε_{xx} is shown along the z-axis, close to the surface boundary. It can be noted that by increasing QD thickness from 2 to 5 nm, the compressive strain slightly decreases inside the QD, while the tensile strain in the GaN surrounding the QD slightly increases. On the other hand, by comparing strain behavior in the center and close to the lateral surface (Figure 17.3)

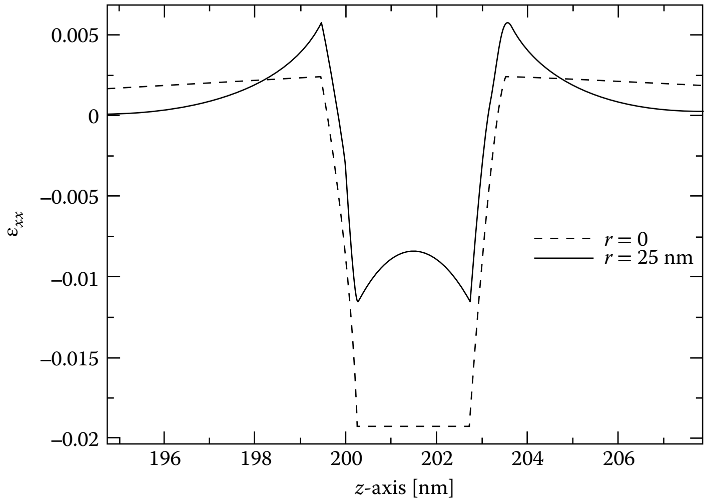


FIGURE 17.2 ε_{xx} strain tensor component along the z-axis, respectively, at the center of the column (dashed line) and close to the column surface (solid line), for an $\text{In}_{0.2}$GaN quantum dot (QD) with $r = 25$ nm.

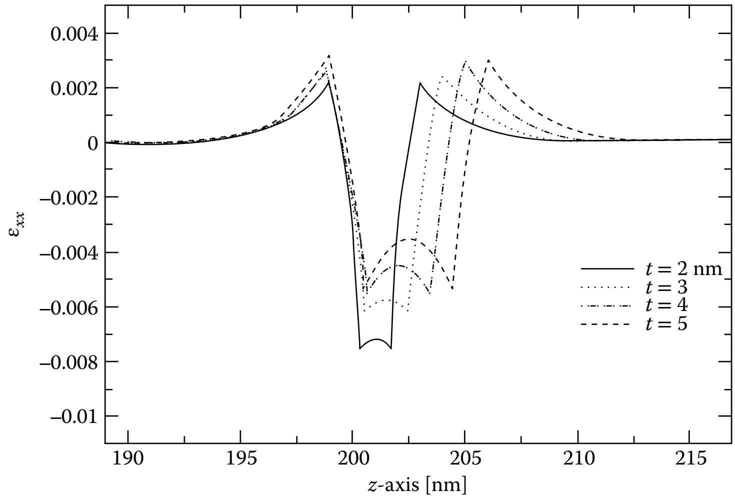


FIGURE 17.3 ε_{xx} strain tensor component along the z-axis, close to the lateral surface, for an $\mathrm{In}_{0.1}$GaN quantum dot (QD) thickness from 2 to 5 nm.

of the NW, the strain relaxation is clearly noticed, which is more evident for thicker QDs (50% reduction for 5-nm-thick QD). We see in the following how these peculiar strain features of the NW structure can be put in relation with the emission properties of the QD LED.

17.4 Transport Properties

Transport in NW QD LED may be studied in a first approach with a classical model. In this way, the IV characteristic of the forward-biased p-i-n LED diode is calculated by imposing ohmic boundary conditions to the contact regions at the two sides of the NW (anode and cathode) and by solving the Poisson/drift–diffusion model, taking care to include the effects of polarization (both spontaneous and piezoelectric), by means of an appropriate polarization vector P in the current equation. Figure 17.4 shows the conduction and valence band profiles and the electron and hole densities along the z-axis in the center of the NW, as obtained from these classical calculations. The dependence of the calculated IV characteristics on the In molar fraction in the InGaN QD is shown in Figure 17.5 for a column with radius $r = 25$ nm (Sacconi et al., 2010). As expected, the threshold voltage decreases from 3.2 to around 2.2 V with an increase of In concentration from $x = 0.1$ to $x = 0.4$, since the energy gap of the alloy material in the QD is decreasing. In addition, from the IV characteristics for a 20% In composition in the QD we obtain an increase of the output current by about two orders of magnitude for increasing values of the column size, from 25 to 200 nm. In the following, we see how a self-consistent calculation can be used to couple these classical results with a quantum model for the QD.

17.5 Quantum Calculations: EFA Models

For the calculations of the eigenvalues of a confined system, an approach based on EFA is usually a good trade-off between computational load and precision of the results. A solution based on the $\mathbf{k}\cdot\mathbf{p}$ model allows us to take into account the electronic structure details close to the valley minima, which is considered a reasonable approximation, provided that particle energies are not too high, as could be the case for hot electrons. The most complete model, the so-called 8×8 $\mathbf{k}\cdot\mathbf{p}$ model, takes into account the three valence bands: heavy-hole (HH), light-hole (LH), and crystal-field split-off (CH), together with the first conduction band (CB).

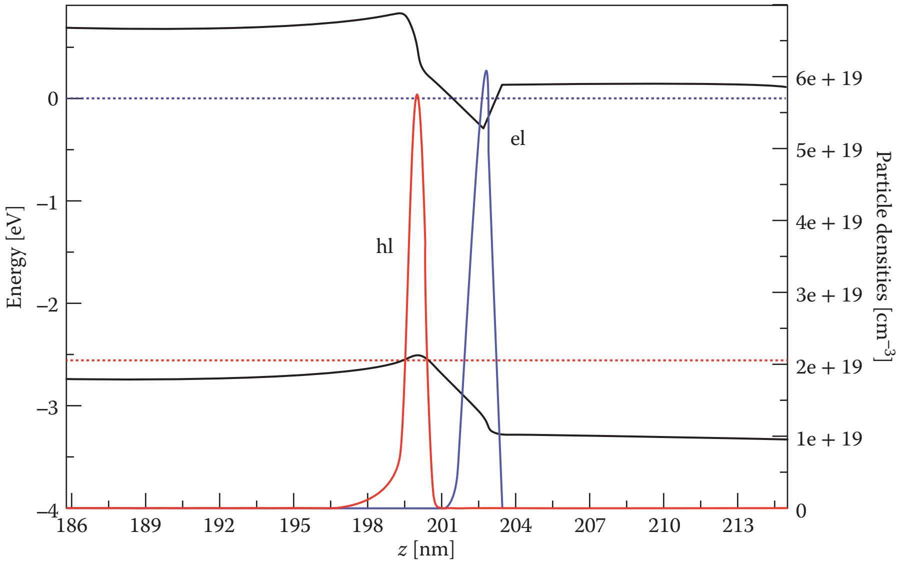


FIGURE 17.4 Band profiles and electron and hole densities from classical calculations.

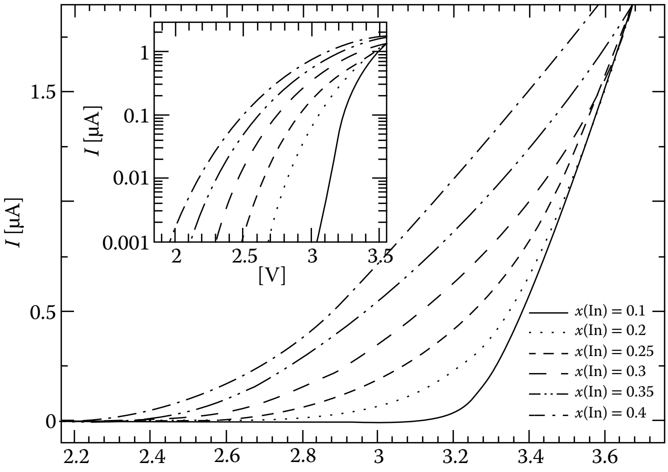


FIGURE 17.5 IV characteristics of the p-i-n diode with radius $r = 25$ nm for several values of In concentration in the InGaN quantum dot (QD).

However, when the energy gap of the material is wide enough, as is the case of GaN and its alloys with AlN and InN (for low InN contents), the interaction between conduction and valence bands is so weak that it can be neglected. Thus, one can apply separately a single band (effective mass) model for CB and a 6×6 **k·p** model for the three VBs, thus reducing the computational time needed without significant loss of accuracy in calculations.

To perform quantum calculations with the EFA model for an LED device, usually the drift–diffusion model is first solved, in order to apply a bias ramp to the diode until the desired operation conditions are obtained. Then a Schrödinger solver is applied for the solution of the eigenvalue problem, restricted

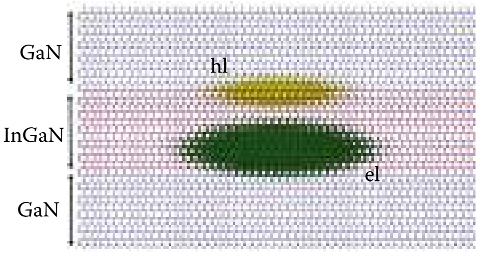


FIGURE 17.6 Conduction and valence band tight-binding states in a 14-nm-wide $In_{0.1}GaN$ quantum dot (QD).

to the active region of the device, comprising the QD and part of the GaN barrier regions. From the calculated conduction and valence ground states, the transition energies between the first electron and hole states can be obtained, for several values of In content (Sacconi et al., 2012). It is found that the transition energy decreases with higher In content mainly due to a lower energy gap in the QD. The transition energy dependence on geometrical and material parameters will be analyzed in more detail later in this chapter.

If we express the conduction and valence band ground states by means of the isosurface of probability density for each quantum state, it is evident that in both cases the quantum states are confined in the QD region (see also Figure 17.6). However, a spatial separation is clearly present between the electron and hole states due to the polarization fields that cause the so-called quantum confined stark effect (QCSE), contrary to what usually happens with GaAs-based structures.

In the following, we see how these results can be corrected by a self-consistent solution of Schrödinger and Poisson/drift–diffusion equations.

17.6 Quantum Calculations: Empirical Tight-Binding Approach

As we have seen before, a more accurate approach for the study of quantum properties of nanostructures is based on atomistic methods. It has been shown recently that this approach is able to provide a description of a realistic InGaN alloy features, contributing to an explanation of the decreasing emission efficiency on nitride-based quantum well (QW) LEDs (green droop) (Auf der Maur et al., 2016). We therefore present the results obtained when ETB calculations are performed to find confined electronic states in an InGaN/GaN QD NW.

In a first example, the simulation model is applied to an GaN NW with an $In_{0.1}GaN$ QD about 14 nm wide and with a thickness of 3 nm. First, an atomistic structure comprising around 77,000 atoms is generated, based on the finite element (FE) grid used for the discretization of continuous partial differential equation (PDE) models (Auf der Maur et al., 2011, 2013). This association with FE allows projection to the atomic positions the displacements obtained from strain calculation in the heterostructure with the PDE-based elasticity model.

Then eigenstates and eigenfunctions of the system are calculated on the strained atomistic system with an ETB model based on an $sp^3d^5s^*$ parameterization (Jancu et al., 2002).

The calculation is restricted to the atomistic structure, which describes the materials included in the LED active region consisting of the InGaN QD (see Figure 17.6) and an appropriate portion of the GaN NW surrounding the QD. As in the EFA **k·p** case, here a Poisson/drift–diffusion calculation also is performed earlier on the NW model (that is its FE grid). Then the potential profile obtained at a given bias, which includes polarization fields, is projected on the atomistic representation to be used for ETB calculations. This is generally accomplished with a shifting of the on-site energies of the ETB Hamiltonian matrix (Auf der Maur et al., 2011, 2013).

In Figure 17.6, we show the first hole and electron quantum states, calculated at an applied bias of V_d = 3.5 V. As in the previous EFA results, it is clearly seen that the polarization-induced electric field leads to a spatial separation of electron and hole states.

We see in the following how a multiscale simulation can be performed where both the atomistic (ETB) and the continuous (EFA) models are applied together for a better description of the QD electronic structure.

As a second example, we see an ETB calculation based on a random alloy approach.

Given a ternary alloy in the form $A_xB_{1-x}C$, such as $In_xGa_{1-x}N$, there are two ways to model it. The first is virtual crystal approximation (VCA), where we consider the alloy as a fictitious material whose properties are a weighted average of the properties of the binary components AC and BC (e.g. InN and GaN), in function of the molar fraction x, according to Vegard's law. The second is the random alloy approach, which consists of studying an ensemble of stochastic realizations of structures where we substitute the anion according to a probability given by the concentration of anion A in the alloy.

Let us consider then a QD NW with a diameter of 10.7 nm with a 2-nm-thick QD. The total number of atoms in the atomistic structure is around 51,000.

A set of atomistic structures of the GaN/InGaN/GaN QD region, including the InGaN QD and two 2-nm-thick GaN barrier regions, are created on the GaN substrate lattice. Random alloy is generated in the InGaN region by including In atoms with a probability given by the chosen In concentration. Then, as an initial guess, macroscopic strain is calculated and the displacements obtained are projected on the atom positions. VFF relaxation is applied on this structure until the atoms' equilibrium positions are found. Finally, ETB calculation is performed on the relaxed structure, taking into account the appropriate scaling of ETB parameters according to the bond length obtained by the relaxation procedure.

Electron and hole ground states have been calculated. Figure 17.7 shows the transition energy E_t for increasing In concentration in the InGaN QD. The average value of E_t for the set of calculations is plotted, together with the range between the minimum and maximum values. Random alloy results are compared with those obtained with ETB for the same model structure, but with a VCA approach and just a macroscopic strain. It can be seen that there is a significant discrepancy between the two models. In fact, they differ by about 50 meV for x(In) = 0.1 and the difference gets larger with increasing In content. These results show that VCA may not be adequate to describe nanostructures based on InGaN alloys due to the

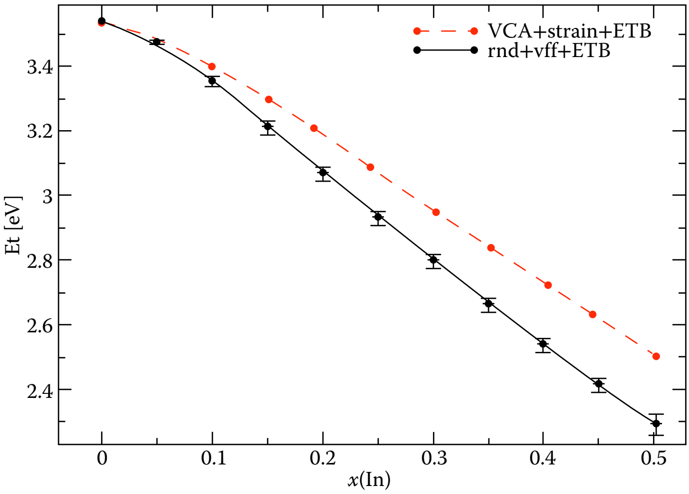


FIGURE 17.7 Dependence of transition energy on In concentration in the quantum dot (QD): Virtual crystal approximation and random alloy model.

great difference in the lattice properties of its component. The random alloy approach, with an atomistic treatment of strain with VFF, may be required to study these devices.

17.7 Self-Consistent Calculations

As discussed previously, a correction to a purely classical description of transport in the NW QD LEDs is needed to account for quantum effects. In this section, we describe a convenient procedure to take into account quantum charge densities of electrons and holes, obtained from an EFA model as described earlier, in the transport calculations of an NW QD LED. The solution to this issue described here consists of a self-consistent coupling of the Schrödinger equation in the EFA framework and the Poisson/drift–diffusion equations. This approach presents the advantage of a lower computational load with respect to more accurate methods for the calculation of quantum transport (e.g., NEGF; see Datta, 2000) while providing all the information about scattering features included in the drift–diffusion model. Thus, even if several approximations must be adopted, such as the assumption of quasiequilibrium of Fermi levels in the quantum regions, the results obtained may be considered reasonably accurate and have been validated by several experimental benchmarks. For more details on the method applied, see, e.g., Auf der Maur et al. (2011).

Let us see now an example of the application of a self-consistent model to an NW QD LED device. It is convenient to perform first purely classical drift–diffusion calculations to ramp the diode voltage up to a desired operating point (e.g., just above its threshold voltage V_{th}). In this way, the LED is biased at a point where the first quantized levels in the QD are populated; this guarantees that the calculated particle quantum density resulting from quantum EFA calculations is not vanishing and thus can be considered comparable with the classical density of the particle resulting from the Fermi level position in the QD. Then, the drift–diffusion model is coupled with a Schrödinger solver for the self-consistent solution of the eigenvalue problem restricted to the active region of the device. The quantum mechanical electron and hole densities are fed back into the Poisson/drift–diffusion equations for a self-consistent Schrödinger–Poisson/drift–diffusion calculation. A self-consistent loop may be implemented by using a simple predictor–corrector scheme that assumes that the quantum density n_{q} varies with the potentials as the classical density n_{cl} (for more details see Sacconi et al., 2012). As the electron and hole states in the system are calculated applying closed boundary conditions, the quantum densities near the interface between quantum mechanical and classical simulation domains suffer from artificial behavior. To obtain a continuous transition from the purely classical densities far away from the QD to the quantum mechanical densities, one may define an embracing region with an extension of a few nanometers where a linear mixing of classical and quantum density can be applied (Sacconi et al., 2012). Figure 17.8 shows the conduction and valence band profiles along the z-axis in the center of the NW, together with the particle densities obtained after the self-consistent cycle.

By comparing the self-consistent electron and hole charge densities of Figure 17.8 with those of Figure 17.4, obtained with purely classical calculations, the effect of quantum correction is evident. The charge density distribution is no longer peaked at the heterointerface between the QD and the GaN barrier material, as determined by the band bending in classical calculations. Instead, it is moved toward the center of the QD, accordingly with the behavior of the wave functions of the particle confined states.

The distribution of the electron quantum densities on a xy-plane orthogonal to the growth direction is shown in Figure 17.9. The planar quantum confinement toward the center of the QD is clearly visible for the electrons and it is found even more pronounced for holes, due to their larger effective mass. This effect is even more evident when we compare the classical and self-consistent electron density in the QD region along the y-axis on a slice on a xy-plane orthogonal to the growth direction (Sacconi et al., 2012). As for the classical results, for both conduction and valence band a band bending forms close to the lateral surface due to the strain relaxation. A lower strain determines a lower band-gap at the surface than in the center of the NW. This in turn determines a higher value of classical electron and hole densities close to the surface.

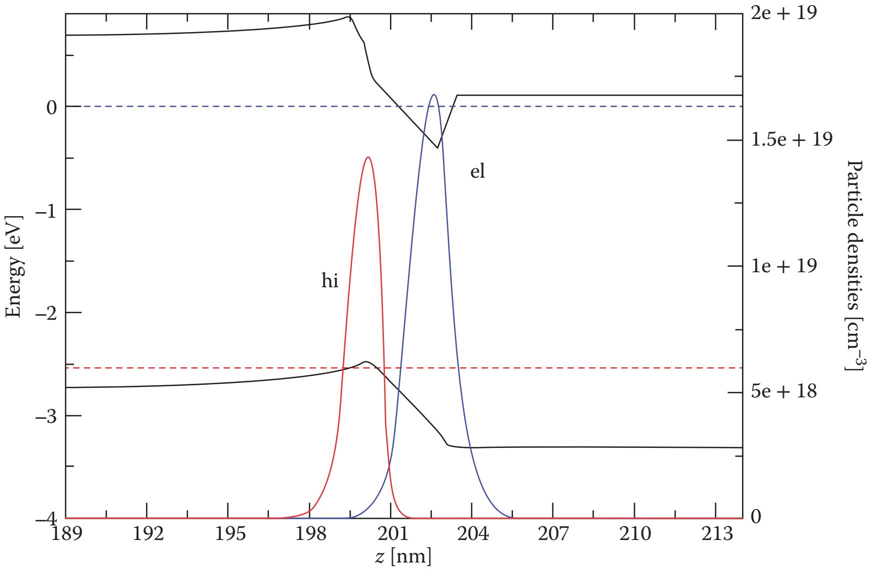


FIGURE 17.8 Self-consistent band profiles and densities along growth direction z in the quantum dot (QD).

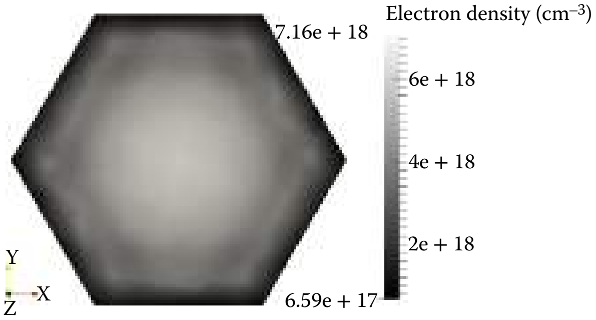


FIGURE 17.9 Electron quantum densities in the xy-plane in the middle of the quantum dot (QD).

On the other hand, the self-consistent densities show the effect of a lateral quantum confinement, so that the carriers are mainly confined to the center region of the QD, even if for electrons this behavior is slightly more complicated.

Due to the behavior of both hole and electron density, we can expect a very different qualitative result when comparing the radiative recombination rate $R = Bnp$ in the active region, that is, the QD, calculated with classical and quantum model. In fact, it is found that from the classical results the emission appears to originate from a region very close to the surface of the column; on the other hand, when quantum mechanical particle densities are used the radiative recombination is mainly concentrated at the center of the QD due to the spatial confinement of the carriers. In Figures 17.10 and 17.11, we show a comparison of the calculated current density through the NW QD, in both cases. Current flow lines in the self-consistent case (Figure 17.11) indicate that the confinement of the quantum states leads to a current crowding at the center of the NW. In the purely classical calculation (Figure 17.10), on the contrary, flow lines focus toward the lateral surfaces, where the classical densities are higher. This result can explain the difference in the IV characteristics obtained with the two approaches (Sacconi et al., 2012). In fact, it turns out that the self-consistent results show a lower threshold voltage and an integrated current value at the contacts of around

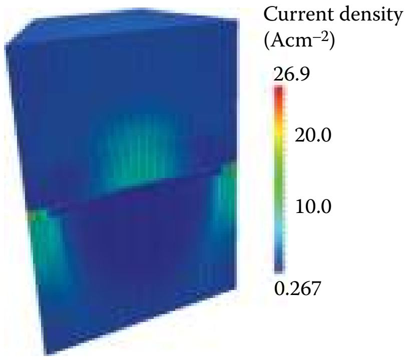


FIGURE 17.10 Classical calculations: Current flow is focused to quantum dot (QD) edges.

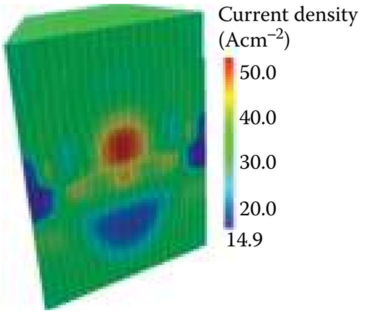


FIGURE 17.11 Self-consistent calculations: Current flow is focused to quantum dot (QD) center.

10 times higher with respect to the classical outcome. This discrepancy can be attributed to the increased recombination rate in the QD active region. In fact, on the one hand, the classical particle densities show higher peak values, but on the other hand, the average value of densities and thus of recombination is higher after quantum correction, which leads to a higher level of output current.

17.7.1 Multiscale Simulation

We now consider an example of a multiscale simulation applied to a NW QD LED (Auf der Maur et al., 2011). Here, a self-consistent classical/quantum calculation is again performed, but this time the information on quantum density of electrons and holes is obtained by an atomistic model. In this case, the device model is a p-i-n LED made of an AlGaN NW with an embedded GaN QD. The simulation is performed as follows.

First, the strain in the NW is calculated based on continuum elasticity theory. This result is then used as a guess for an atomistic VFF method applied for the relaxation of the atomistic structure generated according to the FEM model of the GaN device active region.

Then the electronic transport in the NW LED is simulated based on the self-consistent solution of the Schrödinger/Poisson/drift–diffusion equations using both the continuous media EFA model and the atomistic ETB approach.

To perform the transport calculation, first the drift–diffusion model is applied to get a bias where the diode is nearly in flat band condition. At this bias point the actual multiscale simulation is performed by concurrently solving the Poisson/drift–diffusion and the quantum mechanical models. For the latter both a $\mathbf{k}\cdot\mathbf{p}$ 6×6 model and an ETB model (using an $sp^3d^5s^*$parameterization) are used. Since the hole states result is very dense for this device structure, a large number of states are needed to obtain a reasonably converged quantum hole density. Moreover, the ETB calculation is computationally much more intensive than an EFA calculation. Therefore, one possible choice is to calculate only the first three electron states using ETB and the higher ones with EFA, while EFA alone is used for the hole states.

This approach is a good compromise between the accuracy of full-band ETB calculations of the particle quantum states and an affordable computational load. In fact, it turns out that the hole state energies calculated by EFA and ETB are very similar and thus we can rely on the less computationally demanding EFA model for them. On the other hand, the higher electron states are less populated and therefore they have less impact on the properties of the system. We limit then our ETB approach to the first more populated electron states. This can be considered a decomposition of the problem in energy space since EFA and ETB are used to calculate energetically different particle states.

The coupling between the different models constituting the multiscale system is obtained by a concurrent solution implemented by applying a self-consistent cycle, similar to that shown earlier.

This coupling can be viewed as a quantum correction of the local density of states (LDOS) in the classical expression for the particle density, where the local densities can be easily calculated based on the envelope functions. In addition to the purely continuous case, here we have also the particle density obtained by ETB, which needs to be projected onto the FE mesh that has been used for the discretization of the continuous media models. This projection may be obtained by using an exponentially decaying function centered on each atomic site.

The results of these self-consistent calculations show also in this case that the confinement of the quantum states leads to a current crowding toward the center of the NW (Auf der Maur et al., 2011). We can conclude that a multiscale approach is in principle able to describe with higher accuracy the confinement effects that are critical for the NW LED devices while keeping the simulation domain for the quantum mechanical models reasonably small.

17.8 Modeling of Surface States

Surface states are expected to be present on the surface of GaN NWs due to defects and dangling bonds (Van de Walle and Segev, 2007).

The effect of surface states can be taken into account in a numerical simulation through the implementation in the drift–diffusion model of an acceptor trap model so that the trap is negatively charged if not occupied by a hole. Following the results obtained in Calarco et al. (2005) we can place the trap at an energy value of 0.6 eV below the conduction band, assuming a trap charge density of 1×10^{12} cm^{-2}. If we apply this model to a GaN NW we obtain a conduction band profile that shows a pinning of the Fermi level at the surface at around 0.6 eV below the conduction band. As is shown in Calarco et al. (2005), when a doping is present the pinning of the Fermi level determines a depletion region that, in this case, covers the whole column size. In this way, we have fitted the surface trap model to the experimental value of the Fermi level pinning in a GaN NW. In the following, we assume that, as a first approximation, the same model holds for the InGaN/GaN QD NW structure.

To investigate the influence of surface states on the electronic properties of the QD NW LED, a Shockley–Read–Hall (SRH) surface recombination model has to be associated to the traps. In this way, an SRH model for nonradiative recombination takes into account the surface traps, with a given density and recombination time, and their contribution to the recombination part of the current (Sacconi et al., 2012). Now, if one performs classical drift–diffusion calculations, taking into account the SRH surface recombination component, results (Sacconi et al., 2012) show that the presence of the surface states induces a large recombination current in the diode, which provides an output current around 60 times larger than in the case where surface states are neglected.

As is reported in Sacconi et al. (2012), the flow lines focus close to the surface, where there is the peak of the classical particle densities, which, thus, largely increases the surface recombination contribution.

On the other hand, when self-consistent quantum densities are calculated, this effect almost vanishes, and the self-consistent current at the contacts is just 20% higher than in the case without surface recombination. In fact, it is found that the self-consistent current flow lines tend again to crowd inside the QD since contribution from surface states is now limited by the quantum confinement effect.

17.9 Sensitivity to Geometrical and Material Parameters

We discuss in the following the dependence of the transition energies and thus of the optical emission from the QD on the NW geometrical and material parameters.

The optical emission spectra from spontaneous recombination may be calculated in the following way (Chuang, 2009):

$$P(\hbar\omega) = \sum_{i,j}^{\infty} n \frac{1}{2\pi^2} \frac{\omega_{ij}^2 e^2}{m^2 c^3} \left| M_{i,j} e \right|^2 f_i(E_i) \left[1 - f_j(E_j) \right] \frac{\Gamma/2}{\left(\hbar\omega_{ij} - \hbar\omega \right)^2 + (\Gamma/2)^2} \mathrm{d}\Omega,$$

where f_i and f_j are the Fermi distributions and $M_{i,j}$ is the optical matrix element between the states i and j. The matrix elements may be obtained with different quantum models, such as ETB and EFA **k·p**, and provide the probability of optical transitions between each couple of states. The sum is extended to all the quantum states calculated for conduction and valence band. In this way, the optical transition probability is weighted with the occupation functions for each couple of particles, calculated according to the Fermi distribution at each state energy. The result is the total emitted power for transitions in a given energy range, given the particle population of the considered quantum states.

Usually, we are interested in the emission resulting from the spontaneous recombination of electrons in conduction band and holes in valence band. From the simulation point of view, this means that the device conduction band has to be populated with an optical or electrical pumping. In the following, the latter method has been performed, that is, a ramp bias has been applied until the LED is in conduction regime and the conduction and valence band are populated. At this point, a quantum model, such as **k·p**, is applied for the calculation of quantum states. Following that, the optical emission spectra are calculated based on the previous steps.

Figure 17.12 shows the optical spectra obtained for a GaN column with a radius of 25 nm and a 3-nm-thick InGaN QD (Sacconi et al., 2010). The results show a decrease of the peak emission energy with the increase of In concentration in the InGaN QD from 10% to 40%, indicating that an emission wavelength range between 400 and 680 nm can be covered in principle by these devices.

Generally speaking, in NW LEDs, similarly to QWs LED structures, emission energy decreases with higher In content due to a lower energy gap in the QD and a higher QCSE. The variation of the QCSE is caused by the increase of polarization fields with lattice mismatch between InGaN and GaN (piezoelectric polarization) and with In content (pyroelectric polarization). Moreover, it can be seen (Sacconi et al., 2012) that the emission power tends to decrease with higher In concentration due to the increase of spatial separation of the carriers caused by higher strain-induced QCSE. On the other hand, the calculated optical spectra for a fixed In composition and variable column size for the LED considered earlier show an increase in the value of peak emission energy of around 60 meV when the NW width is reduced from 400 to 50 nm. This result is a combination of several effects related to strain relaxation close to the NW boundaries and quantum confinement in the QD; we discuss this point in detail in the following paragraph.

As for the effect of InGaN QD thickness on the optical emission spectra, the behavior is again similar to what occurs in a QW LED. For example, the emission peak in a 50-nm-wide QD increases by 150 meV when the thickness of the QD decreases from 5 to 2 nm. This is mainly due to the higher quantum confinement in the thinner QD; moreover, for the same polarization field, the QCSE is lower for a smaller

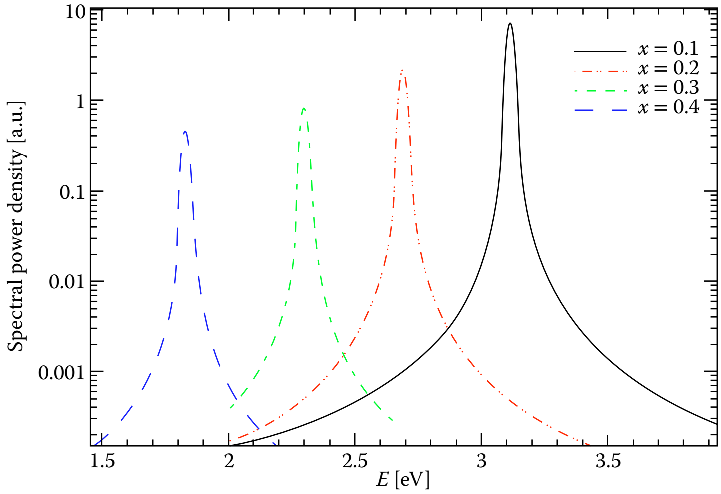


FIGURE 17.12 Optical emission spectra for a column with a radius of 25 nm and several In concentrations.

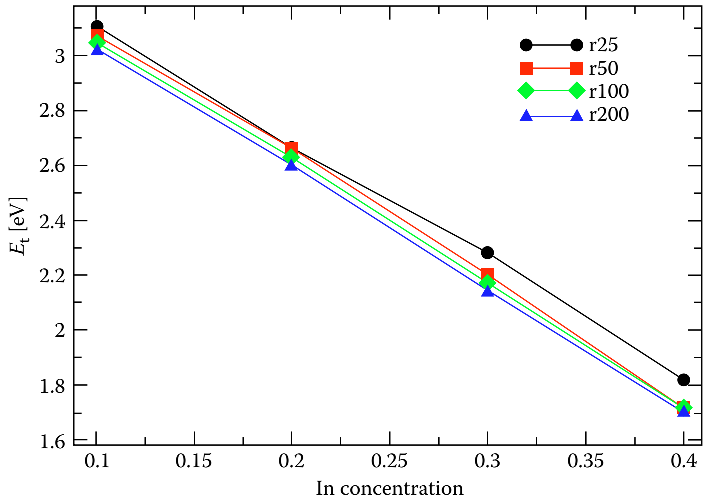


FIGURE 17.13 Dependence of emission energy on In concentration for several column widths.

thickness, and thus QCSE-induced red-shift is reduced. This also determines an increased emission power due to the reduced spatial separation of electron and hole quantum states.

Let us see now in more detail how the transition energies depend on the geometry parameters of the QD NW, in particular, how the lateral scaling of the NW affects the emission properties.

Figure 17.13 shows the dependence of emission energy on In concentration in an InGaN QD NW for several values of the column radius (Sacconi et al., 2012). It can be seen that, for all the column widths, the emission energy has a linear dependence on the In molar fraction. On the other hand, Figure 17.13 shows that the transition energy increases slightly with decreasing width, with a larger slope for radii lower than 50 nm. The effect gets larger with increased In concentration.

Calculations performed in a study by Sacconi et al. (2012) by applying a full self-consistent quantum/drift–diffusion coupling shows an emission energy increase of around 64 meV, from 2.440 to 2.504 eV, when the column radius decreases from 100 to 25 nm. These results show a trend in qualitative agreement with experiments (Ramesh et al., 2010), even if it must be noted that a comparison of calculations with experimental results is quite difficult because the results depend critically on the geometrical and structure details, as well as on the particle density distribution in the NW.

To point out the role of confinement and of strain distribution on this lateral size effect, the dependence of emission energy on NW geometry has been studied in Sacconi et al. (2012) for several structures, beginning with simulations for undoped structures at equilibrium. In Figure 17.14, the result of a 1D calculation for a QW structure, which can be assumed as a model of a NW with an infinite radius, is compared to the result for a 50-nm-wide NW. It is clear that the geometry of the NW increases the transition energy for a given value of In molar fraction. Moreover, calculations for a very wide column (10 μm) yield results very close to the thin film case. This result can be put in relation with the polarization-induced electric field obtained in the QD shown in Figure 17.15. It can be seen that the electric field gets lower in the NW QD

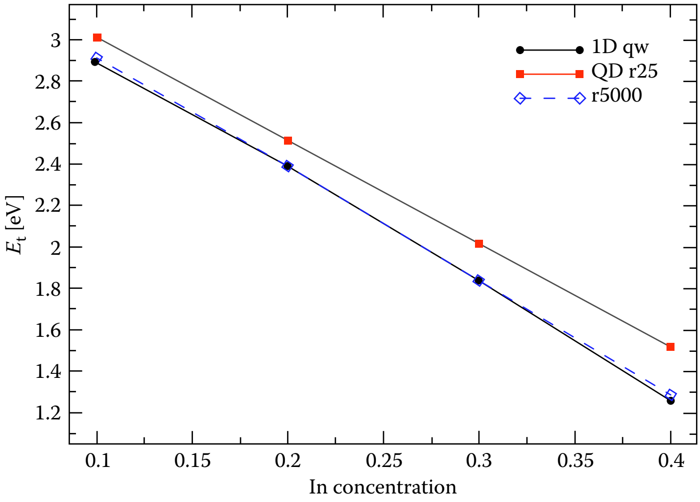


FIGURE 17.14 Dependence of emission energy on In content for undoped structures.

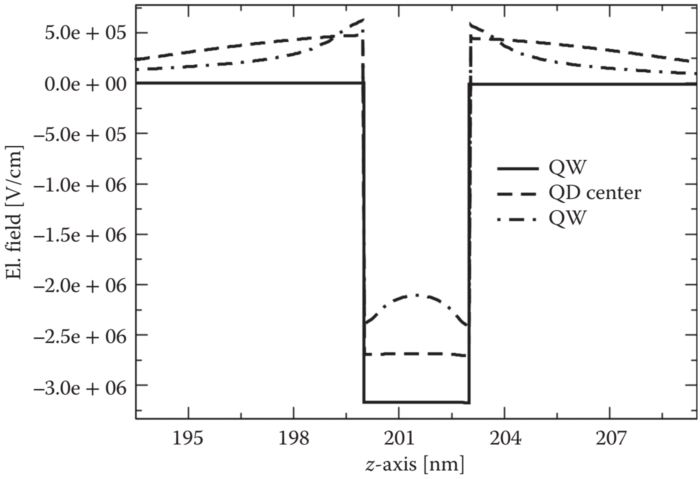


FIGURE 17.15 Electric field along the z-axis in an undoped nanocolumn.

with respect to the QW value due to the 3D strain effects, causing lower QCSE and therefore higher transition energy. Furthermore, the electric field in the QD is even more reduced close to the lateral surface due to the strain relaxation effect. The situation is anyway less clear in the doped structures, where screening effects become appreciable. Moreover, it is found that if polarization and strain are not taken into account, the lateral confinement has a significant effect only when the NW radius is reduced under 10 nm (Sacconi et al., 2012).

From these considerations, it seems reasonable to conclude that the main effect of the lateral scaling of the NW is the reduction of the strain inside the QD due to its relaxation at the NW lateral surfaces. In fact, the smaller the NW size, the more pronounced is this reduction of the strain. A lower strain implies less piezoelectric polarization and thus a lower value of QCSE, which causes a blue shift in emission energies. In addition, a smaller size induces a stronger quantum confinement, whose effect on transition energy would be, however, significant only at an extremely reduced size. Anyway, as we have seen earlier, an important role of quantum confinement is to keep particles far from the lateral surfaces, thus limiting the direct influence of NW surface effects, such as strain relaxation and surface states, on the behavior of current in the QD and thus on the LED emission properties.

We have seen the effects of scaling down the NW geometry; on the other hand, one can wonder what happens when the NW size is scaled up. It has been found experimentally (Kawakami et al., 2010) that for a 500-nm-wide NW, emission seems to occur at the strain relaxed region close to the lateral boundary. From simulations performed on an NW model with this size (Sacconi et al., 2012) it turns out that higher recombination rates are obtained for the regions close to the lateral surface, suggesting that for very wide NWs the confinement of quantum states on the NW lateral plane is not sufficient to focus the recombination at the center of the QD. Thus, emission takes place mainly in the lateral regions, where the strain relaxation induces a lower energy gap. However, it has been found from simulations that already for 200-nm-wide NWs the quantum confinement begins to affect the charge distribution, focusing the emission at the center of the QD.

17.10 Conclusions

After having discussed the several aspects of NW QD behavior, we can now underline the main beneficial features provided by NW geometry compared with planar LED structure.

First of all, the NW geometry can be grown almost without defects, yielding in principle high-emission efficiency. NWs also offer the possibility of covering a wide emission spectrum by means of band-gap engineering through variations of alloy concentration in the QD.

Another peculiar NW feature is the relaxation of strain induced in the QD, which occurs at the lateral surface of the NW. This is part of a 3D effect, which determines a lower strain in the QD active region, with respect to an analogous QW in a planar LED structure. This in turns causes less QCSE and higher overlap of quantum wave functions, increasing recombination rate, and internal quantum efficiency, with respect to the planar case.

A lower strain at lateral boundaries determines a lower energy gap with respect to the rest of the QD. This may affect the LED efficiency by reducing the recombination rate. However, as we have shown earlier, the quantum confinement on the lateral plane is sufficient, at least for NW size lower than 200 nm, to focus particle densities in a limited region at the center of the QD. The resulting current crowding in the QD and the increase of radiative recombination are then further advantages of NW geometry versus planar LED. It must be noted also that very often the embedded QD obtained by the NW growth is surrounded by the NW material (yielding a so-called InGaN inclusion in a GaN NW). In this case, the active region is even less affected by surface effects, such as surface states, while the confinement is increased and thereby also the current focusing and the radiative recombination.

For all of these features, one can expect NW QD structures to be very promising for the design of future high-efficiency LED devices.

References

M. G. Ancona, Density-gradient theory analysis of electron distributions in heterostructures, *Superlattices Microstruct.*, 7(2), 119–130 (1990).

M. Auf der Maur, Multiscale approaches for the simulation of InGaN/GaN LEDs, *J. Comput. Electron.*, 14(2), 398–408 (June 2015).

M. Auf der Maur et al., The multiscale paradigm in electronic device simulation, *IEEE Trans. Electron. Devices*, 58(5), 1425–1432 (May 2011).

M. Auf der Maur, A. Pecchia, G. Penazzi, F. Sacconi, and A. Di Carlo, Coupling atomistic and continuous media models for electronic device simulation, *J. Comput. Electron.*, 12(4), 553–562 (2013).

M. Auf der Maur, A. Pecchia, G. Penazzi, W. Rodrigues, and A. Di Carlo, Efficiency drop in green InGaN/GaN light emitting diodes: The role of random alloy fluctuations. *Phys. Rev. Lett.*, 116, 027401 (2016).

R. Calarco et al., Size-dependent photoconductivity in MBE-grown GaN-nanowires. *Nano Lett.*, 5, 981–984 (2005).

D. Camacho and Y. M. Niquet, Application of Keating's valence force field model to non-ideal wurtzite materials. *Phys. E,* 42, 1361–1364 (2010).

S. L. Chuang, *Physics of Optoelectronic Devices*, 1st ed. Wiley Series in Pure and Applied Optics. New York, NY: Wiley (1995).

S. L. Chuang, *Physics of Photonic Devices*, 2nd ed. New York, NY: Wiley (2009).

S. L. Chuang and C. Chang, k•p method for strained wurtzite semiconductors, *Phys. Rev. B,* 54, 2491–2504 (1996).

S. Datta, Nanoscale device simulation: The Green's Function Method, *Superlattices Microstruct.*, 28, 253–278 (2000).

A. Di Carlo, Microscopic theory of nanostructured semiconductor devices: Beyond the envelope-function approximation, *Semicond. Sci. Technol.,* 18, 1 (2003).

M. Hugues et al., Strain evolution in GaN nanowires: From free-surface objects to coalesced templates, *J. Appl. Phys.*, 114, 084307 (2013).

J.-M. Jancu et al., Transferable tight-binding parametrization for the group-III nitrides. *Appl. Phys. Lett.*, 81, 4838, (2002).

Y. Kawakami et al., Optical properties of InGaN/GaN nanopillars fabricated by postgrowth chemically assisted ion beam etching, *J. Appl. Phys.*, 107(2), 023522-1–023522-7 (January 2010).

P. N. Keating, Effect of invariance requirements on the elastic strain energy of crystals with application to the diamond structure. *Phys. Rev.,* 145(2), 26 (1966).

H. Kim et al., High-brightness light emitting diodes using dislocation-free indium gallium nitride/gallium nitride multiquantum-well nanorod arrays, *Nano Lett.*, 4(6), 1059 (2004).

J. Piprek, Efficiency drop in nitride-based light-emitting diodes, *Phys. Stat. Sol. A*, 207(10), 2217–2225 (October 2010).

M. Povolotskyi and A. Di Carlo, Elasticity theory of pseudomorphic heterostructures grown on substrates of arbitrary thickness, *J. Appl. Phys.*, 100, 063514 (2006).

V. Ramesh, A. Kikuchi, K. Kishino, M. Funato, and Y. Kawakami, Strain relaxation effect by nanotexturing InGaN/GaN multiple quantum well, *J. Appl. Phys.*, 107(11), 114303-1–114303-6 (June 2010).

J. Ristic et al., Carrier-confinement effects in nanocolumnar GaN/AlGaN quantum disks grown by molecular-beam epitaxy, *Phys. Rev. B*, 72, 085330 (2005).

F. Sacconi, M. Auf der Maur, and A. Di Carlo, Optoelectronic properties of nanocolumn InGaN/GaN LEDs, *IEEE Trans. Electron. Dev.*, 59(11), (November 2012).

F. Sacconi, G. Penazzi, A. Pecchia, M. Auf der Maur and A. Di Carlo, Optoelectronic and transport properties of nanocolumnar InGaN/GaN quantum disk LEDs, *Proc. SPIE 7597, Phys. Sim. Optoelectron. Dev.*, XVIII, 75970D (February 25, 2010); doi:10.1117/12.840533.

F. Sacconi, M. Auf der Maur, A. Pecchia, M. Lopez, and A. Di Carlo, Optoelectronic properties of nanocolumnar InGaN/GaN quantum disk LEDs, *Phys. Status Solidi C*, 9(5), 1315–1319 (2012); doi:10.1002/pssc.201100205.

S. M. Sze, *Physics of Semiconductor Devices*, 2nd ed. Hoboken, NJ: Wiley (1981).

C. Van de Walle and D. Segev, Microscopic origins of surface states on nitride surfaces, *J. Appl. Phys.*, 101, 081704 (2007).

18

Influence of Random InGaN Alloy Fluctuations on GaN-Based Light-Emitting Diodes

18.1 Introduction to Random Alloy Distribution 559
18.2 Methodology in Random Alloy Modeling 562
Method for Generating the Random Alloy Distribution • Strain Modeling • Calculation of Strain with 3D FEM Method • Calculation of Piezoelectric Polarization and the Implementation Method • Modeling of Carrier Transport under the Random Alloy Fluctuation with Drift–Diffusion Solver • Schrödinger Equation and Emission Rate
18.3 Unipolar Transport for Random Alloy System 569
Electron Transport in n-GaN/i-InGaN/n-GaN Structures
18.4 Modeling Result of MQW InGaN LED and the Comparison to the Traditional Model 573
Simulation Structure • The Influence of Indium Fluctuation on the Emission Spectrum • The Influence of Indium Fluctuation on the Carrier Transport and the Efficiency Droop
18.5 Carrier Transport in Green LEDs: Influence of Possible Imperfection in QWs 580
Introduction • Carrier Transport in Green LEDs Considering Indium Fluctuations Only • Imperfection of QWs in Green LEDs • 3D Examination of Green LEDs with Imperfect QW Model and Random Indium Fluctuations Model
18.6 Summary.. 584

Chen-Kuo Wu
Tsung-Jui Yang
and
Yuh-Renn Wu

18.1 Introduction to Random Alloy Distribution

Unlike the GaAs-based light-emitting diodes (LEDs), the carrier transport and recombination mechanism in nitride-based LEDs are relatively more complicated. As we know, nitride-based wurtzite structures have a strong spontaneous polarization field inside the material. In addition, when the InGaN quantum well (QW) is grown on the GaN buffer layer, it also suffers extra strain, which induces the piezoelectric polarization. Therefore, the polarization field difference at the interface will induce the quantum confined stark effect (QCSE) and increase the radiative lifetime. The lack of native substrate also makes the crystal quality of the GaN and InGaN layer relatively poor, where a typical 10^8 cm^{-2} dislocation density can be observed in the InGaN QW, which is expected to affect device performance significantly. The efficiency droop, where the light output power increases sublinearly with the drive current, has also been an issue in limiting LED efficiency. There are many studies focused on determining the origin of droop, such as electron overflow [1–3], Auger recombination [4–9], or defects [10–14]. The defect or phonon-assisted Auger recombination is also proposed to explain the larger Auger coefficient. Recently, the direct measurement of Auger electrons [7,15] and a clear correlation between the Auger current and droop seem to indicate that the Auger process might be the main cause of droop. In addition, as mentioned, the QCSE will increase the radiative lifetime and the contribution of nonradiative Auger recombination or carrier delocalization toward droop will become more significant.

Despite these issues, the nitride-based blue LED still has very high internal quantum efficiency (IQE) even under this high dislocation density. The 80% peak IQE of blue LEDs still can be easily achieved in today's commercial LEDs. The reason for this high quantum efficiency has been attributed to the carrier localization effects, where carriers are localized at a local potential minimum so that it will not diffuse into the dislocation center for nonradiative recombination. The spectrum broadening effect is also much stronger compared to GaAs-based LEDs. This is attributed to the random distributed localized state. The origin of carrier localization has been an issue whether it is from indium clustering effect, random alloy fluctuation, or even charged dislocation line-induced potential barrier, etc. More and more evidence shows that random alloy fluctuation should be the main reason for this effect. First, this localization and spectrum broadening is still observed in GaN-negative substrate, which could exclude the effect of a charged dislocation-induced potential barrier. In addition, the three-dimensional (3D) atom probe tomography (APT) data [16–24] show that the indium composition distribution is naturally disordered. Figure 18.1a shows the randomly fluctuated indium distribution in the lateral direction from the APT data [17]. The in-plane map reveals that there are some high indium composition locations corresponding to the red regions and relatively low indium composition sites scattered in the 2D map. Figure 18.1b also indicates that the average indium composition along the growth direction is not uniform and decreases from the middle of the QW to the interface of the InGaN/GaN. Therefore, the QW structure possesses an indium distribution closer to a Gaussian shape rather than the ideal "top-hat" function. A similar idea of the influence of Gaussian shape QW on the IQE has been discussed by Hader et al. [6]. However, their analyses were one dimensional (1D) and lateral fluctuations in indium composition were not considered.

Many simulation tasks [22,24] suggested that carrier localization, induced by these fluctuations, has a strong influence on the broadening of the light emission spectrum. As discussed in Refs. [6,22,23], due to

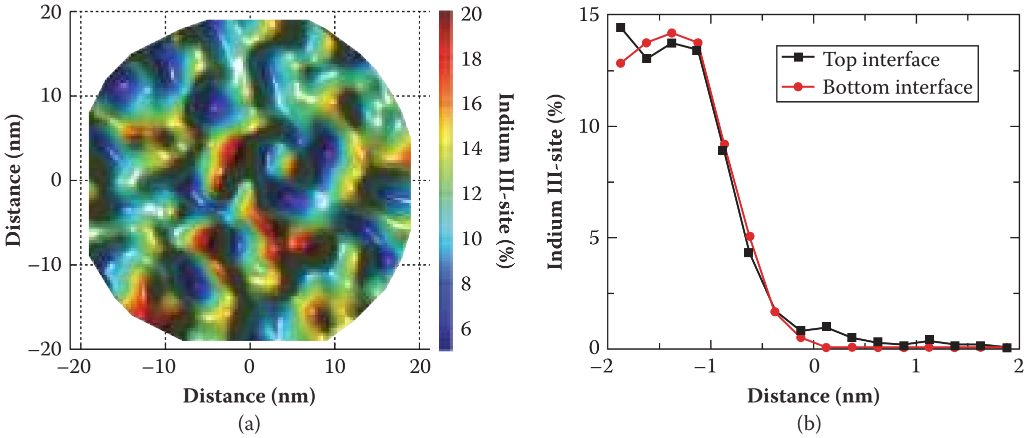


FIGURE 18.1 (a) The in-plane indium distribution of the quantum well (QW). (b) The average indium composition along the growth direction of one 3-nm QW. (From D. A. Browne et al., *Journal of Applied Physics*, 117, 185703, 2015.)

the indium fluctuation, the electrons and holes injected into InGaN QWs will localize in the high indium regions because of the deep localized potential. Additionally, the localized energy states originating from quantum confinement will vary due to fluctuations in indium composition and the QW width, thus broadening the emission spectrum [22,24]. In addition, the carrier localization will increase the local carrier density and the strong screening of the polarization field might be achieved at a smaller current density. The high local carrier density might increase the influence on Auger recombination or defect-assisted recombination [11] at a relatively smaller current density. In addition, the local polarization value induced by composition fluctuation will also influence polarization potential in the QW [25].

Besides these tasks that study the influence on radiative and nonradiative recombination due to the localization effect, there are a few other efforts that analyze the impact on carrier transport by considering random alloy potential fluctuation. In the past, many simulation studies on multiple QW LEDs have been performed under the assumption of a uniform composition and thus potential in the QWs. However, the simulations reveal a large deviation from the experimental data, especially in the InGaN MQWs cases. Due to the strong GaN barrier induced by the piezoelectric charges, a much higher applied bias is usually obtained in the simulation than in the experimental data. The deviation of simulation to experimental results becomes very significant, especially in MQW cases where many triangular shapes of GaN barriers exist in the system. Some studies reduced the theoretical polarization value to fit the experimental data. However, these assumptions might lead to more problems since they might improve electron–hole overlap and result in no droop or droop at a much larger bias, or need a much larger Auger coefficient to reproduce the experimentally observed droop behavior. Some used the tunneling effect to explain this phenomenon. However, the tunneling effect is significant only when a smaller polarization field was employed since the potential barrier induced by the 100% theoretical polarization filed could be very high. If the simulation only tries to fit the current versus IQE and neglects the fitting of $I-V$ curve, usually a higher voltage is obtained in simulation and a conclusion of overflow or overshot effects are often obtained. This is due to that the extra-voltage lowered down polarization barrier induced by polarization field and potential barrier in electron blocking layer and make carrier to overflow much easier. However, this extra voltage is not observed in experimental result in a good commercial LED. Therefore, the conclusion would be wrong since it is based on a non-existence factor.

As mentioned earlier, the influence of random alloy fluctuation on the transport is seldom considered due to the need of large amounts of computational power. Atomic simulation of carrier transport with the 3D random indium fluctuation, especially in MQWs at current computation power, is also impossible. Even modeling a small area of InGaN QW (10 nm × 10 nm × 3 nm) will require a few months calculation with a super computer. Therefore, some quasi-classical simulation work is needed to approach these issues. Our past work [23,24,26] indicated that the results obtained by including the indium fluctuation into the 2D and 3D Poisson, drift–diffusion, and Schrödinger equation simulation model could be closer to experimental data, without the necessity for the assumptions mentioned earlier. The randomly fluctuated alloy composition of ternary epilayers will result in irregular energy bandgap and piezoelectric potential distribution, which will make carrier percolation and confinement more complicated compared to the traditional 1D assumption.

Figure 18.2 shows a sketch of a carrier percolation in the random alloy system. The red regions refer to the high potential region and the blue regions refer to the low potential area. As the sketch shows, the potential distribution is fluctuated either in the QW region or in the electron-blocking layer (EBL) region, where carriers are much easier to percolate through smaller barrier sites and localize in the local low-energy areas. While the large number of localized carriers accumulate in QWs, the local polarization charges will be screened and the effect of Auger recombination will be enhanced at the same time. In addition, the inherent complex potential distribution might provide carriers a path to avoid flow into defect-related regions formed by threading dislocations (TDs) in GaN LEDs. Our past research [23,24] has shown that the calculated turn-on voltages and IQE performances can be better predicted without any parameter reductions by considering random alloy fluctuations. A more reasonable Auger coefficient is used in the model [15], and a broad emission spectrum is also observed due to different localized states produced by alloy fluctuations. Moreover, the emission spectrum shift can be modeled well in the paper [24]. In this chapter, we will introduce how we use the traditional 3D Poisson and drift–diffusion solver to address these issues.

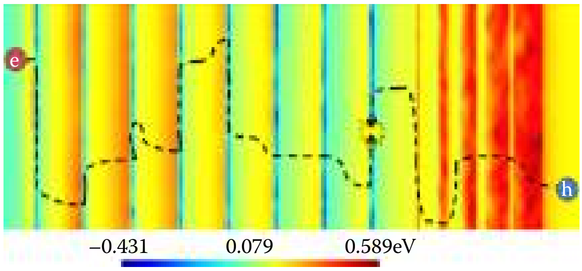


FIGURE 18.2 The sketch illustrates the carrier percolation in the random alloy system. The red regions refer to the high potential region, and the blue regions are low potential area.

18.2 Methodology in Random Alloy Modeling

To model the 3D structure properly, we need to use a full 3D model to examine the device performances of the different structures. The 3D finite element method (FEM) Poisson, drift–diffusion, and Schrödinger solver developed by our lab (named the 3D-DDCC) is mainly used for solving the 3D carrier transport issue of the semiconductor. This program can correlate the electrical characteristic with optical problems and give a self-consistent solution. Our lab has tested and verified the model [24,26–28] and the model has been further developed to apply in the 3D transport simulation. In addition, 3D drift-diffusion charge control (3D-DDCC) program is a very versatile software, in that we can easily use inserted functions in special cases to model the structures or the physics more precisely. In this chapter, the externally inserted functions have the decisive position in the simulation. The details of the functions will be introduced later.

Figure 18.3 shows the full diagram of the simulation flow chart in this modeling process. First, the Gmsh program is used to construct the 3D mesh structure [29]. Afterwards, we need to use the random number

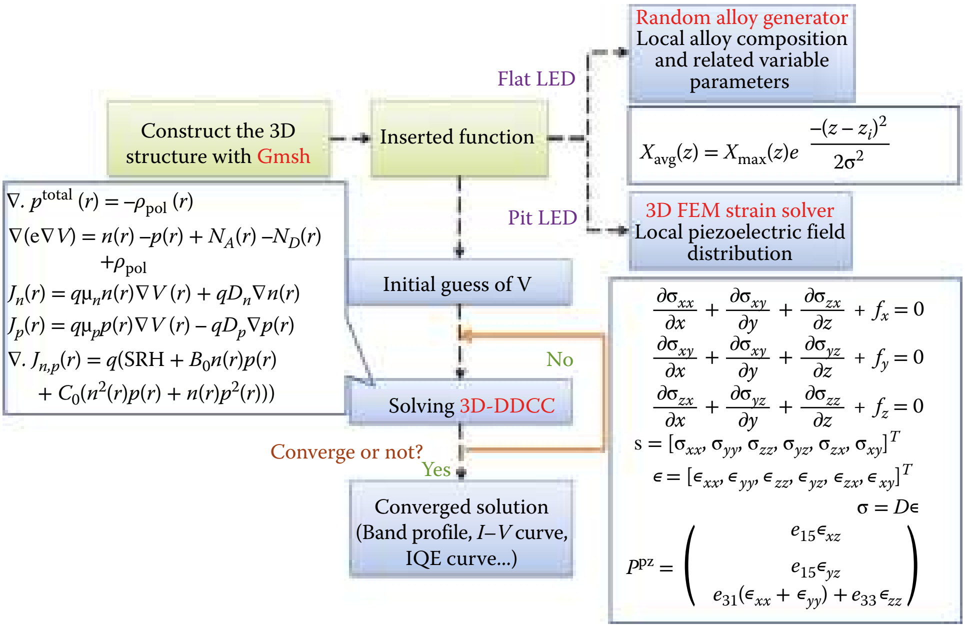

Random alloy generator
Local alloy composition and related variable parameters
Flat LED
Construct the 3D structure with Gmsh
Inserted function
$X_{avg}(z) = X_{max}(z)e^{\frac{-(z-z_i)^2}{2\sigma^2}}$
Pit LED
3D FEM strain solver
Local piezoelectric field distribution
$\nabla . p^{total}(r) = -\rho_{pol}(r)$
$\nabla(e\nabla V) = n(r) - p(r) + N_A(r) - N_D(r) + \rho_{pol}$
$J_n(r) = q\mu_n n(r)\nabla V(r) + qD_n\nabla n(r)$
$J_p(r) = q\mu_p p(r)\nabla V(r) - qD_p\nabla p(r)$
$\nabla . J_{n,p}(r) = q(\mathrm{SRH} + B_0 n(r)p(r) + C_0(n^2(r)p(r) + n(r)p^2(r)))$
Initial guess of V
No
Solving 3D-DDCC
Converge or not?
Yes
Converged solution (Band profile, I–V curve, IQE curve...)
$\frac{\partial\sigma_{xx}}{\partial x} + \frac{\partial\sigma_{xy}}{\partial y} + \frac{\partial\sigma_{zx}}{\partial z} + f_x = 0$
$\frac{\partial\sigma_{xy}}{\partial x} + \frac{\partial\sigma_{xy}}{\partial y} + \frac{\partial\sigma_{yz}}{\partial z} + f_y = 0$
$\frac{\partial\sigma_{zx}}{\partial x} + \frac{\partial\sigma_{yz}}{\partial y} + \frac{\partial\sigma_{zz}}{\partial z} + f_z = 0$
$s = [\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{yz}, \sigma_{zx}, \sigma_{xy}]^T$
$\epsilon = [\epsilon_{xx}, \epsilon_{yy}, \epsilon_{zz}, \epsilon_{yz}, \epsilon_{zx}, \epsilon_{xy}]^T$
$\sigma = D\epsilon$
$P^{pz} = \begin{pmatrix} e_{15}\epsilon_{xz} \\ e_{15}\epsilon_{yz} \\ e_{31}(\epsilon_{xx} + \epsilon_{yy}) + e_{33}\epsilon_{zz} \end{pmatrix}$

FIGURE 18.3 The full diagram of the simulation flow chart in this modeling process.

generator to generate the random alloy composition which is utilized in the examination of carrier transport in the random alloy system. After obtaining the indium map, we assign the indium composition in each node of the mesh element and all physical parameters change with the indium composition. To obtain a proper consideration of the piezoelectric polarization under the random alloy conditions, the 3D FEM-based strain solver can be used to account for the piezoelectric polarization and calculate the polarization charge. The 3D-DDCC developed by our lab is then used to solve the transport equations with the appropriate input parameters generated by the external modules until we get the converged solutions. Finally, we can obtain the physical profiles such as the potential distributions, current–voltage (I–V) curve, and IQE curve.

18.2.1 Method for Generating the Random Alloy Distribution

As shown in Figure 18.4a, first the In(Al) and Ga atoms are randomly assigned by the random number generator according to the average indium composition and are aligned to the cation lattice site. The lattice site size is decided by the atom density. In addition, according to the APT data, the average alloy composition of each lateral plane along the z-direction, $\chi_{\mathrm{avg}}(z)$, is like a Gaussian shape distribution as shown in Figure 18.5a and b, especially when the QW is very thin [6,23,24,30]. Note that the adenosine triphosphate (ATP) resolution in the depth is about one monolayer and lateral resolution is around 2 nm. Therefore, we can obtain the average indium composition of the ith QW by the following equation:

$$\chi_{\mathrm{avg}}(z) = \chi_{\max} e^{\frac{-(z-z_i)^2}{2\sigma_s^2}}, \tag{18.1}$$

where $\chi_{\max}$ is the peak average composition of the epilayer, z_i is at the middle of the ith QW, and σ_s is the half width of the Gaussian broadening coefficient. The value of σ_s is around the half width of the epilayer in the z-direction. When atoms are assigned according to Equation 18.1, we need to choose a volume size to get the average local alloy composition. The volume size we use to calculate the local composition is around 2 nm × 2 nm × 0.667 nm, which is close to the APT resolution. Then the alloy composition is weighted at this volume region with Gaussian shape weighting criteria.

As shown in Figure 18.6, the local alloy composition extracted on volumes as mentioned before is following a binomial distribution, which is similar to Refs. [16,18–22].

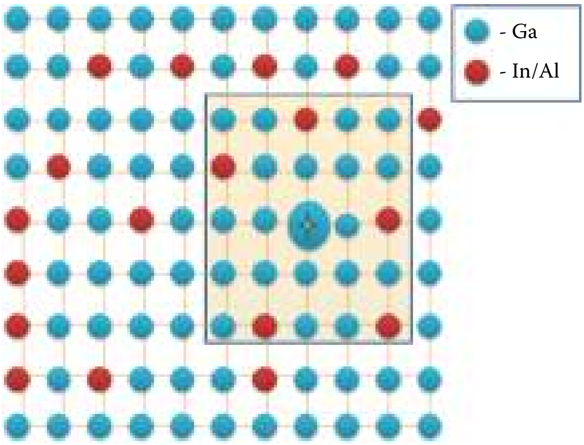


FIGURE 18.4 The concept of the alloy generator. The In, Al, or Ga atom at each lattice site (all with N atoms) is assigned by the random number. Then we choose a proper volume to average the local alloy composition. (From C.-K. Wu et al., *Journal of Computational Electronics*, 14, 416–424, 2015.)

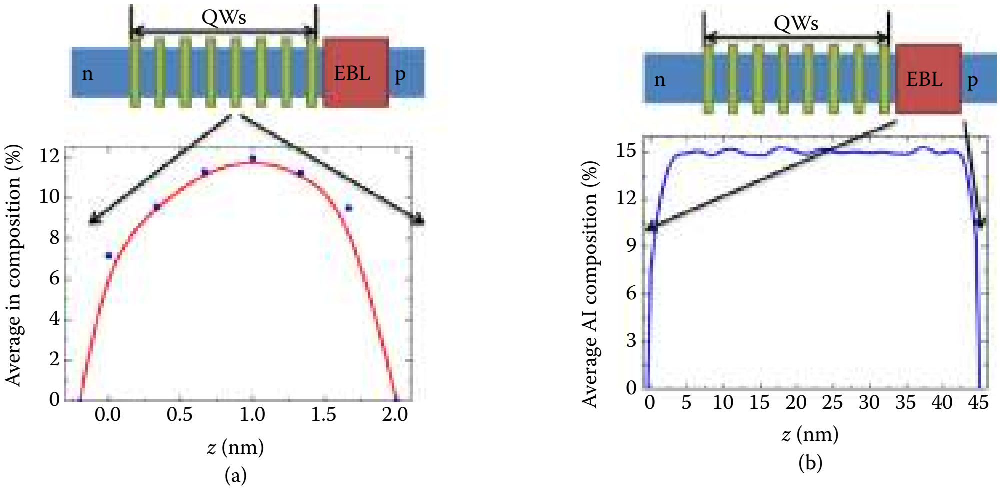


FIGURE 18.5 (a) The average alloy composition distributes along the 2-nm quantum well. (b) The average alloy composition distributes along the 45-nm electron blocking layer (EBL). (From C.-K. Wu et al., *Journal of Computational Electronics*, 14, 416–424, 2015.)

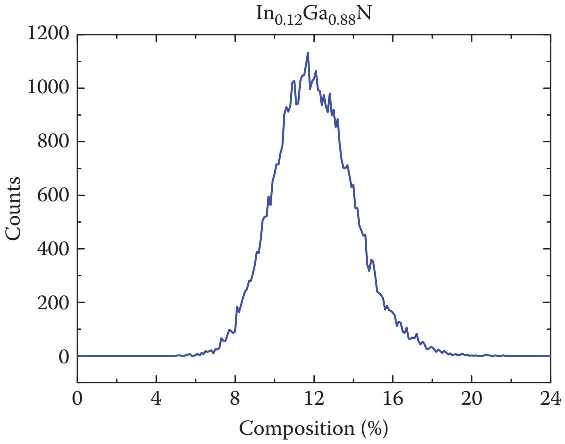


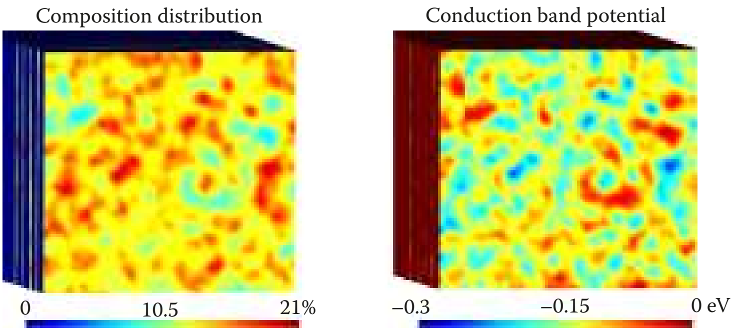
FIGURE 18.6 The counting distribution of certain alloy compositions of the $In_{0.12}Ga_{0.88}N$ quantum well (QW) [26].

After assigning the alloy landscape as shown in Figure 18.7, we combine the mesh file of the structures with the alloy maps and produce the local input parameters simultaneously. The material parameters of each mesh node are assigned according to the local alloy composition map $x(r)$. All parameters (e.g., bandgap [31], effective mass [32,33]) in each node are calculated locally with the indium composition map $x(r)$ by our in-house 3D simulation solver.

18.2.2 Strain Modeling

18.2.3 Calculation of Strain with 3D FEM Method

To understand how the polarization is induced in the random alloy fluctuation condition, we need to do a 3D finite element strain analysis. We need to solve the displacement field by the FEM after considering the incorporation of incoherent lattice constants.



(a) (b)

FIGURE 18.7 (a) One random alloy distribution of the n-i-n InGaN quantum wells (QW). (b) The section view of the fluctuated conduction band potential, which is corresponds to the random alloy distribution, in the n-i-n InGaN quantum well case at 0 V bias. (From C.-K. Wu et al., *Journal of Computational Electronics*, 14, 416–424, 2015.)

18.2.3.1 Equilibrium Equations

When the element is located in the x,y,z coordinates and some regions are constrained, the element deforms under the balanced force. We can describe the deformation of a point x (= $[x, y, z]^T$) in three components of its displacement:

$$u = [u, v, w]^T. \tag{18.2}$$

The force per unit volume in the vector form can be expressed as:

$$f = [f_x, f_y, f_z]^T. \tag{18.3}$$

In Figure 18.8, we can find the stresses and the body force distributed on the elemental volume $\mathrm{d}V$. Here, we describe the stress by six components as

$$\boldsymbol{\sigma} = [\sigma_{xx}, \sigma_{yy}, \sigma_{zz}, \sigma_{yz}, \sigma_{xz}, \sigma_{xy}]^T, \tag{18.4}$$

where σ_x, σ_y, σ_z are the normal stresses and σ_{yz}, σ_{xz}, σ_{xy} are the shear stresses. Under the equilibrium conditions, the total forces along each direction are zero in each elemental volume. In Figure 18.8, we can pay attention to the x-axis and multiply the stresses to the corresponding areas and list the following equation:

$$\frac{\partial \sigma_{xx}}{\partial x}\mathrm{d}x(dydz) + \frac{\partial \sigma_{xy}}{\partial y}\mathrm{d}y(dxdz) + \frac{\partial \sigma_{zx}}{\partial z}\mathrm{d}z(\mathrm{d}x\mathrm{d}y) + f_x \mathrm{d}V = 0. \tag{18.5}$$

Knowing the $\mathrm{d}V = \mathrm{d}x\mathrm{d}y\mathrm{d}z$, Equation 18.5 can be simplified as the following:

$$\frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} + \frac{\partial \sigma_{zx}}{\partial z} + f_x = 0. \tag{18.6}$$

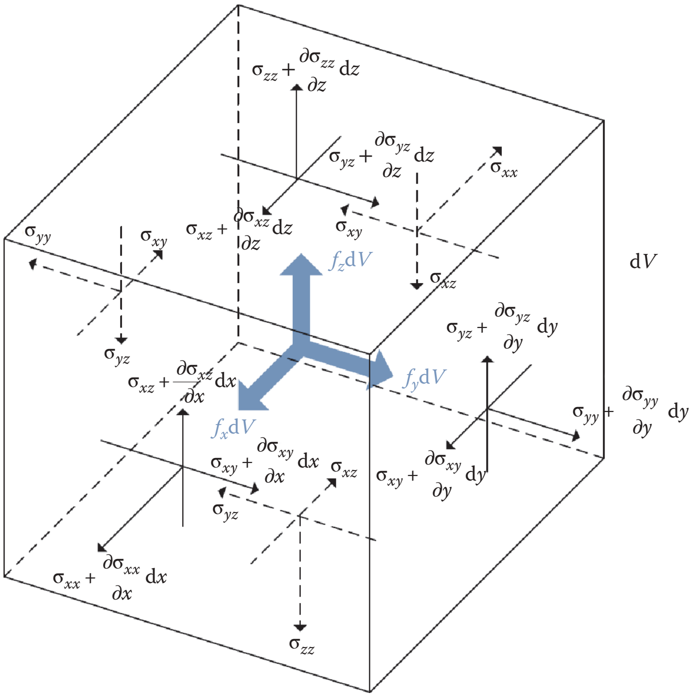

$\sigma_{zz} + \frac{\partial \sigma_{zz}}{\partial z} dz$
$\sigma_{yz} + \frac{\partial \sigma_{yz}}{\partial z} dz$
σ_{xx}
σ_{yy}
σ_{xy}
$\sigma_{xz} + \frac{\partial \sigma_{xz}}{\partial z} dz$
σ_{xy}
$f_z dV$
σ_{xz}
dV
$\sigma_{yz} + \frac{\partial \sigma_{yz}}{\partial y} dy$
σ_{yz}
$\sigma_{xz} + \frac{\partial \sigma_{xz}}{\partial x} dx$
$f_y dV$
$\sigma_{yy} + \frac{\partial \sigma_{yy}}{\partial y} dy$
$f_x dV$
$\sigma_{xy} + \frac{\partial \sigma_{xy}}{\partial x} dx$
σ_{xz}
$\sigma_{xy} + \frac{\partial \sigma_{xy}}{\partial y} dy$
σ_{yz}
$\sigma_{xx} + \frac{\partial \sigma_{xx}}{\partial x} dx$
σ_{zz}

FIGURE 18.8 Equilibrium condition of an elemental volume. (From T. R. Chandrupatla and A. D. Belegundu. *Introduction to Finite Elements in Engineering*. Pearson Education International, 2002.)

After considering the total force along x-, y-, and z-directions, we can list the total equilibrium equations as follows:

$$\frac{\partial \sigma_{xx}}{\partial x} + \frac{\partial \sigma_{xy}}{\partial y} + \frac{\partial \sigma_{zx}}{\partial z} + f_x = 0 \tag{18.7}$$

$$\frac{\partial \sigma_{xy}}{\partial x} + \frac{\partial \sigma_{yy}}{\partial y} + \frac{\partial \sigma_{yz}}{\partial z} + f_y = 0 \tag{18.8}$$

$$\frac{\partial \sigma_{zx}}{\partial x} + \frac{\partial \sigma_{yz}}{\partial y} + \frac{\partial \sigma_{zz}}{\partial z} + f_z = 0. \tag{18.9}$$

18.2.3.2 Stress–Strain–Displacement Relations

In Equations 18.2 and 18.4, we know the form of the stresses and displacement. The strains have the corresponding form as following:

$$\boldsymbol{\epsilon} = [\epsilon_{xx}, \epsilon_{yy}, \epsilon_{zz}, \epsilon_{yz}, \epsilon_{xz}, \epsilon_{xy}]^T, \tag{18.10}$$

where ϵ_{xx}, ϵ_{yy}, and ϵ_{zz} are normal strains and ϵ_{yz}, ϵ_{xy}, and ϵ_{xz} are shear strains. We can also write the strain–displacement relation as

$$\boldsymbol{\epsilon} = \left[\frac{\partial u}{\partial x}, \frac{\partial v}{\partial y}, \frac{\partial w}{\partial z}, \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}, \frac{\partial u}{\partial z} + \frac{\partial w}{\partial x}, \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right]^T. \tag{18.11}$$

These strain–displacement relations hold under small deformations.

In our research, we assume that the linearly electric stress–strain relations come from the generalized Hooke's law.

$$\boldsymbol{\sigma} = \boldsymbol{D}\boldsymbol{\epsilon}. \tag{18.12}$$

D is the material matrix. In our research, we focused on the wurtzite structure, AlN, GaN, and InN, and ***D*** is given by

$$\boldsymbol{D} = \begin{pmatrix} c_{11} & c_{12} & c_{13} & 0 & 0 & 0 \\ c_{12} & c_{11} & c_{13} & 0 & 0 & 0 \\ c_{13} & c_{13} & c_{33} & 0 & 0 & 0 \\ 0 & 0 & 0 & c_{44} & 0 & 0 \\ 0 & 0 & 0 & 0 & c_{44} & 0 \\ 0 & 0 & 0 & 0 & 0 & c_{66} \end{pmatrix}. \tag{18.13}$$

18.2.4 Calculation of Piezoelectric Polarization and the Implementation Method

Due to the random alloy fluctuation, each element has its own spontaneous polarization $P^{\mathrm{sp}}(r)$ and piezoelectric polarization $P^{\mathrm{pz}}(r)$ depending on the local alloy composition. After calculating the strain energy, we can further analyze the piezoelectric polarization field distribution using the following equation:

$$P^{\mathrm{pz}} = \begin{pmatrix} 0 & 0 & 0 & 0 & e_{15} & 0 \\ 0 & 0 & 0 & e_{15} & 0 & 0 \\ e_{31} & e_{31} & e_{33} & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \epsilon_{xx} \\ \epsilon_{yy} \\ \epsilon_{zz} \\ \epsilon_{yz} \\ \epsilon_{xz} \\ \epsilon_{xy} \end{pmatrix} = \begin{pmatrix} e_{15}\epsilon_{xz} \\ e_{15}\epsilon_{yz} \\ e_{31}(\epsilon_{xx} + \epsilon_{yy}) + e_{33}\epsilon_{zz} \end{pmatrix}. \tag{18.14}$$

The e_{15}, e_{31}, and e_{33} are the piezoelectric coefficients, which are listed in Table 18.1.

After obtaining the piezoelectric polarization, the total polarization of each element can be calculated by

$$P^{\mathrm{total}}(r) = P^{\mathrm{sp}}(r) + P^{\mathrm{pz}}(r), \tag{18.15}$$

where P^{sp} is the spontaneous polarization of InGaN alloy where the value can be found in Ref. [35]. After calculating the polarization $P^{\mathrm{total}}(r)$, the polarization charge, $\rho^{\mathrm{pol}}(r)$, induced at each element can be calculated by

$$\nabla \cdot P^{\mathrm{total}}(r) = -\rho_{\mathrm{pol}}(r). \tag{18.16}$$

The induced fixed polarization charges at different positions will be put into the 3D Poisson and drift–diffusion solver to obtain the potential inside the devices.

TABLE 18.1 Piezoelectric Coefficients

	$e_{33}(\mathrm{cm}^{-2})$	$e_{31}(\mathrm{cm}^{-2})$	$e_{15}(\mathrm{cm}^{-2})$
GaN	0.73	−0.49	−0.4
InN	0.73	−0.49	−0.4

18.2.5 Modeling of Carrier Transport under the Random Alloy Fluctuation with Drift–Diffusion Solver

To study the 3D carrier transport simulation under the random alloy fluctuation, we need to apply our 3D FEM-based Poisson and drift–diffusion solver. Although the 3D Poisson and drift–diffusion solver cannot describe the tunneling effect, which can estimate the device performance more accurately, it has been proved that it is a suitable model to describe the carrier transport behavior with proper physical parameters, and a solver that yields an acceptable calculation time and required computation memory [24,26–28]. The solver is based on solving the following equations:

$$\nabla \cdot P^{\text{total}}(r) = -\rho_{\text{pol}}(r) \tag{18.17}$$

$$\nabla(\epsilon \nabla V(r)) = n(r) - p(r) + N_A(r) - N_D(r) + \rho_{\text{pol}}(r), \tag{18.18}$$

$$J_n(r) = q\mu_n n(r)\nabla V(r) + qD_n \nabla n(r), \tag{18.19}$$

$$J_p(r) = q\mu_p p(r)\nabla V(r) - qD_p \nabla p(r), \tag{18.20}$$

$$\nabla \cdot J_{n,p}(r) = q(\text{SRH} + B_0 n(r)p(r) + C_0(n^2(r)p(r) + n(r)p^2(r))), \tag{18.21}$$

$$\text{SRH} = \frac{n(r)p(r) - n_{\text{i}}^2}{\tau_n\left(p(r) + n_{\text{i}}e^{\frac{E_{\text{i}}-E_t}{k_{\text{B}}T}}\right) + \tau_p\left(n(r) + n_{\text{i}}e^{\frac{E_{\text{t}}-E_{\text{i}}}{k_{\text{B}}T}}\right)}, \tag{18.22}$$

where V is the potential, and ϵ is the static dielectric constant. N_A^- and N_D^+ are the doping density. n and p are the free carrier concentration of the electron and hole. $\rho_{\text{pol}}(r)$ is the local analytic polarization charge which varies with local indium composition or the calculated results by the 3D FEM elastic strain solver. q is 1.6×10^{-19}C. $\mu_{n,p}$ are electron and hole mobility. $D_{n,p}$ are the coefficients of diffusion, and $J_{n,p}(r)$ are the electron and hole current, respectively. The Shockley–Read–Hall (SRH) is the defect-assisted nonradiative recombination, where τ_n and τ_p are the nonradiative carrier lifetime depending on the crystal quality. E_{t} is the trap energy level located at the midgap. E_{i} and n_{i} are the intrinsic energy level and intrinsic carrier density, respectively. k_{B} is the Boltzmann constant and T is the temperature (300 K here). B_0 is the radiative recombination coefficient. C_0 is the Auger recombination coefficient. We solved those equations until we obtained a converged solution.

18.2.6 Schrödinger Equation and Emission Rate

To analyze the emission properties of the LED, we need to solve the time-independent Schrödinger equation:

$$H\Psi\left(x,y,z\right) = E\Psi\left(x,y,z\right), \tag{18.23}$$

where H is the Hamiltonian operator, which is a second-order differential operator. The time-independent Schrödinger equation can be also expressed as

$$\left[\frac{-\hbar}{2m^*}\nabla^2 + E_{\text{c,v}}\left(x,y,z\right)\right]\Psi\left(x,y,z\right) = E\Psi\left(x,y,z\right). \tag{18.24}$$

$E_{\text{c,v}}$ is the conduction and valence band potential calculated by the 3D-DDCC. Ψ is the carrier wavefunction. m^* is the effective mass and $\hbar$ is the Planck constant divided by 2π. Since the Schrödinger equation is an eigenvalue problem, we obtain the eigenenergy E corresponding to the eigenwave function.

The formula we used to calculate the spontaneous emission rate [36] can be expressed as

$$R_{\mathrm{sp}}(\hbar\omega) = \frac{e^2 n_{\mathrm{r}} \hbar\omega}{m_0^2 \varepsilon_0 c^3 \hbar^2} \frac{1}{V} \sum_{i,j} |\mathbf{a} \cdot \mathbf{p_{i,j}}|^2 \times \frac{1}{\sigma_{\mathrm{b}} \sqrt{2\pi}} \exp\left[\frac{-\left(E_{i,j} - \hbar\omega\right)^2}{2\sigma_{\mathrm{b}}^2}\right]$$
$$\times f_{\mathrm{e}}\left(E_i^e\right) f_{\mathrm{h}}\left(E_j^h\right) \left(\mathrm{cm}^{-3}\mathrm{s}^{-1}\mathrm{eV}^{-1}\right), \qquad (18.25)$$

where n_{r} is the refractive index and c is the light velocity. The $|\mathbf{a} \cdot \mathbf{p_{i,j}}|$ is the momentum matrix element, which includes the overlapping of the electron and hole wave function. E_i^e and E_j^h are electron and hole eigenenergy states, respectively. $E_{i,j}$ is the effective bandgap and is equal to $E_i^e - E_j^h$. f_{e} and f_{h} are the Fermi–Dirac distribution of electrons and holes, respectively. The equations are shown below:

$$f_{\mathrm{e}}\left(E_i^e\right) = \frac{1}{1 + \exp\left(\left(E_i - E_{fn}\right)/k_{\mathrm{B}}T\right)} \qquad (18.26)$$

$$f_{\mathrm{h}}\left(E_j^h\right) = \frac{1}{1 + \exp\left(\left(E_{fp} - E_j\right)/k_{\mathrm{B}}T\right)}. \qquad (18.27)$$

The $|\mathbf{a} \cdot \mathbf{p_{i,j}}|$ is the momentum matrix element term where electron and hole overlapping is taken into account in both the localization by fluctuations. The effect of QCSE is already accounted for in solving the Poisson equation. A Gaussian broadening σ_{b} was used in the modeling. The value of σ_{b} used in modeling was 10 meV, which is slightly smaller than $k_{\mathrm{B}}T$ to limit the spectrum broadening due to σ_{b}. Therefore, the calculated emission spectrum broadening will be mainly determined by the different energy levels of localized states.

18.3 Unipolar Transport for Random Alloy System

In this section, the percolation transport study of the random alloy system will be presented. To understand how the piezoelectric field affects the transport, we investigate the pure electron transport in the n-GaN/i-InGaN/n-GaN QW structures and the simulation results will also be compared to the experimental work from Browne et al. [17,37]. The 3D numerical model considering random alloy fluctuations will be applied.

18.3.1 Electron Transport in n-GaN/i-InGaN/n-GaN Structures

The existence of piezoelectric polarization and spontaneous polarization has been confirmed by many studies [31,38,39]. As shown in Figure 18.9, if the InGaN QW is grown in between the GaN, it will induce a huge electric field (>1 MV/cm) in the QW and cause the potential bending as shown in Figure 18.9b. If the quantum barrier (QB) thickness is 10 nm, the induced potential barrier peak could be more than 1 eV without electron screening. Theoretically, under this large electric field, we need to apply a strong bias to overcome this barrier for carrier to go through the multiple quantum wells (MQWs). For the single QW case, the influence of the polarization-induced barrier is not significant, especially as it will be screened by carriers from the n-GaN cap layer. However, under the MQW condition, it could cause a huge resistance for carrier transport. For the n-i-n structure, there is no influence of radiative or nonraditive recombination. Therefore, it is a good platform to test how alloy fluctuation and the polarization-induced potential barrier affect the carrier transport. In this section, in order to understand the carrier transport mechanism across the InGaN QWs, we show that we must study how carriers overcome the piezoelectric barriers before we start to investigate the LED structures.

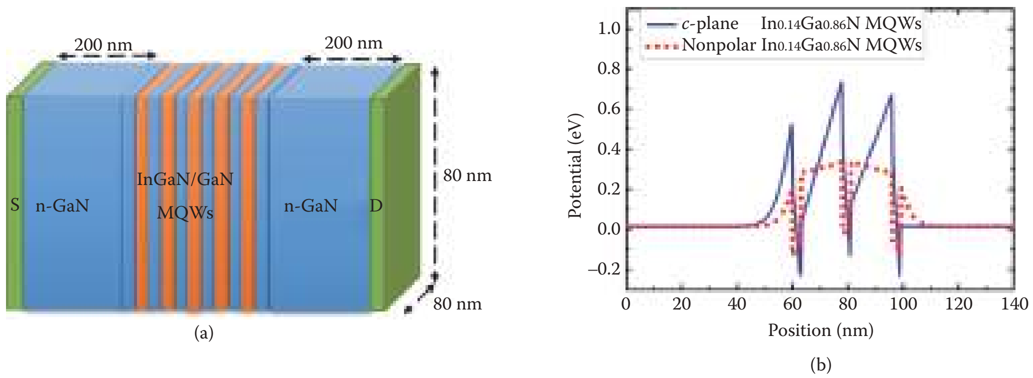


FIGURE 18.9 (a) The modeled device paradigm of n-GaN/i-InGaN/n-GaN quantum well structures for modeling the pure electron transport. (b) The potential distribution of $In_{0.14}Ga_{0.86}N$ *c*-plane and nonpolar piezoelectric barriers.

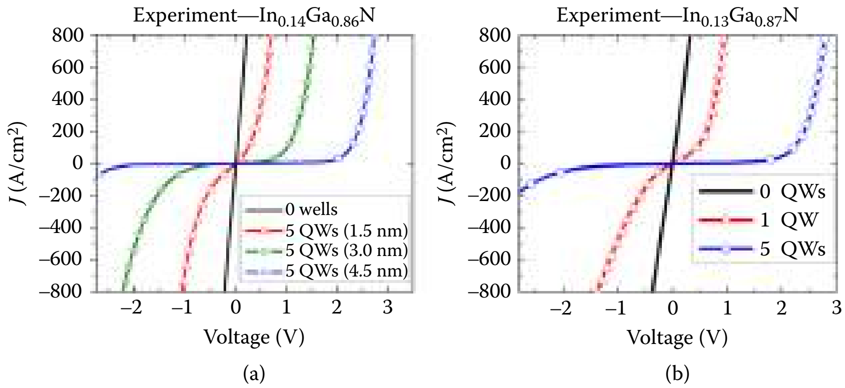


FIGURE 18.10 (a) The experimental *I–V* curves of *c*-plane $In_{0.14}Ga_{0.86}N$ with different thicknesses of quantum wells (QWs). (b) The experimental *I–V* curves of *c*-plane $In_{0.13}Ga_{0.87}N$ with different numbers of QWs. (From D. A. Browne et al., *Journal of Applied Physics*, 117, 185703, 2015; D. A. Browne et al., Investigation of electron transport through InGaN quantum well structures. In *14th Electronic Materials Conference*, Santa Barbara, CA, June 25–27, 2014.)

The traditional *c*-plane MQWs have a larger degree of piezoelectric barriers than nonpolar (*m*-plane) MQWs, as shown in Figure 18.9b. That means the performance of different traditional *c*-plane MQW structures has a higher correlation with their level of piezoelectric barriers. Browne et. al. [17,37] have found a clear rectifying *I–V* curve in *c*-plane MQWs which is related to the QW thicknesses and numbers, where the equivalent piezoelectric barriers will increase with increasing the QW thicknesses and numbers. Figure 18.10 shows that the higher piezoelectric barriers in the devices will increase the driving voltages at the same current density. However, if the QW model is used to study the vertical transport by considering either a full theoretical polarization value or a zero polarization value (nonpolar MQWs), there is a large deviation of predicted driving voltages between experimental work at the same current density, as shown in Figure 18.11.

Besides, it should not be Auger-assisted hot carriers overflow since no Auger recombination could occur in an n-i-n structure and an investigation of the n-i-n structure can be regarded as one that focuses on the carrier transport without any recombination mechanism. The percolation transport through the random alloy system might be a possible reason. A similar effect is also observed experimentally in n-GaN/

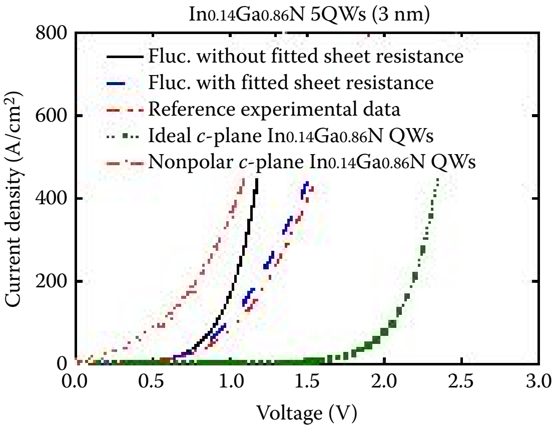


FIGURE 18.11 The *I–V* curves of five experimental and simulated 3-nm quantum wells (QWs) in the n-i-n system. (From D. A. Browne et al., *Journal of Applied Physics*, 117, 185703, 2015; D. A. Browne et al., Investigation of electron transport through InGaN quantum well structures. In *14th Electronic Materials Conference*, Santa Barbara, CA, June 25–27, 2014.)

TABLE 18.2 Detailed Parameters of Each Epilayer for n-GaN/i-InGaN/n-GaN Structures

Epilayer	n-GaN	i-InGaN/GaN
μ_e (cm^2/Vs)	200	600
Doping (1/cm^3)	5.0×10^{18}	1.0×10^{17}
E_a (meV)	25	–

i-AlGaN/n-GaN EBL cases [40]. To verify this, we apply the 3D program by considering the random alloy fluctuation in the InGaN QW.

The structures modeled consiste of two n-type doping sides (5×10^{18} cm^{-3}) that sandwich an intrinsic MQW region. The thicknesses of the QWs are 1.5, 3, and 4.5 nm. The barriers in all the cases are 10 nm. The numbers of QWs are 1 and 5. The configuration of the n-i-n structures is shown in Figure 18.9a. In order to present the nanoscale random alloy fluctuations, a small mesh element size is needed. Therefore, we need to limit our chip to 80 nm × 80 nm in the lateral direction due to computational limitations. The random alloy distribution is based on the rule mentioned in the previous section. The average indium composition is 14% according to Refs. [17,37]. The fluctuation range is around 8%–21%, which follows the binomial distribution. We have simulation results of the ideal $In_{0.14}Ga_{0.86}N$ QW cases for comparison. The physical parameters are listed in Table 18.2.

One of the alloy distributions is shown in Figure 18.7a, and the corresponding fluctuated conduction band potential at 0 V is shown in Figure 18.7b, where a local composition site will induce a reversed trend of potential distribution. Moreover, the lower indium regions will induce relatively low polarization barriers and the electrons could percolate through the relatively lower barrier, as plotted in Figure 18.12.

Hence, the fluctuated case performs smaller turn-on voltages than the uniform cases, as shown in Figure 18.11. Compared with the experimental data [17,37], the turned-on voltage of the *I–V* curve with indium fluctuations is close to the experimental results. Since the experimental result has additional sheet resistance not modeled by our vertical transport study due to the size limitation, with the fitted sheet resistance, the calculated results can further match the experimental results. Figure 18.13 shows a comparison of

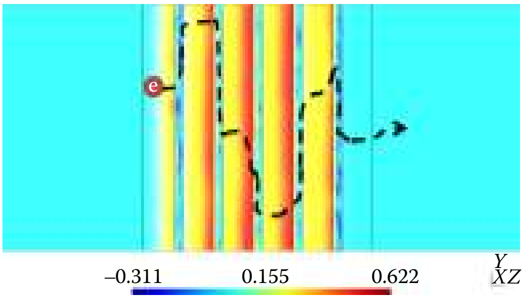


FIGURE 18.12 The side view of conduction band potential in the n-i-n InGaN quantum wells and the scheme of the carrier transport along the fluctuated barriers.

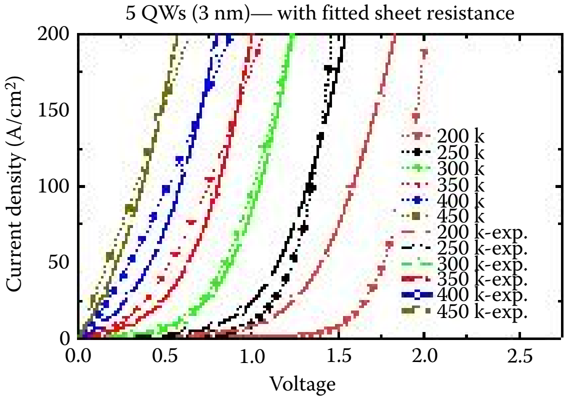


FIGURE 18.13 The *I–V* curves of the experimental and calculated work at various temperatures. (From D. A. Browne et al., Investigation of electron transport through InGaN quantum well structures. In *14th Electronic Materials Conference*, Santa Barbara, CA, June 25–27, 2014.)

the experimental work with the calculated outcomes at different temperatures. At low temperature, the calculations do not fit well. The result reveals that a small portion of tunneling current might dominate at the low temperature rather than the thermionic emission current in the random alloy system because the tunneling current should have a weaker temperature dependence. Consequently, the deviation between predicted current behavior and experimental results might be due to the lack of tunneling process in our transport model.

On the other hand, Figure 18.14 shows the increasing turn-on voltage with increasing the thickness or number of QWs. As we know, for the same polarization electric field, a thicker QW will cause a much larger potential band bending, making it harder for carriers to go across the junction. The turn-on voltage of a single QW is very low because it has only one barrier to be overcome by applying the bias voltage. Even though the thickness of the QW increases to 4.5 nm, the turn-on voltage is still less than 0.5 V. When the number of QWs increases to five, the positive turn-on voltages for 1.5, 3, and 4.5 nm at 20 A/cm^2 current density are 0.10, 0.68, and 1.73 V, respectively.

In addition, there is an asymmetric *I–V* behavior induced by the polarization field, which needs a larger negative voltage to reduce the barriers at reverse bias range. However, since the thermionic current is lower at the reverse bias region, the tunneling current might play a more important role, which requires a program that can handle tunneling transport to model this problem in the future.

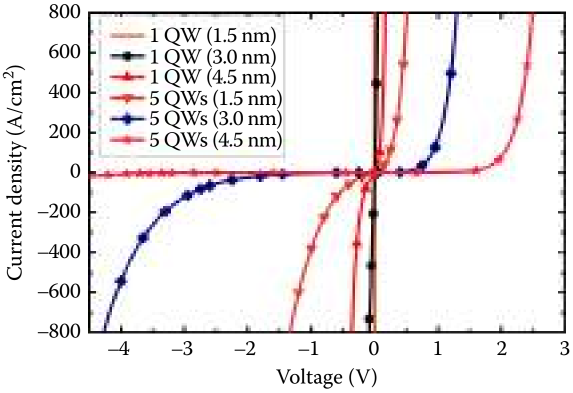


FIGURE 18.14 The *I*–*V* curves of various thicknesses and numbers of the fluctuated quantum wells (QWs).

18.4 Modeling Result of MQW InGaN LED and the Comparison to the Traditional Model

From the unipolar transport, we find that the polarization-induced barrier does play an important role in limiting the carrier transport behavior. However, without using the random alloy fluctuation, it is hard to approach the experimental *I*–*V* behavior. In this section, we focus on how indium fluctuation affects the carrier transport in InGaN LEDs.

As we know, InGaN QW LED has become a popular technology in solid-state lighting. Due to the strong lattice mismatch between InN and GaN layers, the self-formed random indium fluctuation has played an important role in influencing LED electrical and optical properties. In this section, we first study the influence of nanoscale indium fluctuation on the emission spectrum and then the relation of carrier transport and radiative efficiency in LEDs. The randomly generated indium fluctuation in the QW in this simulation will be used again in the study with 3D modeling.

18.4.1 Simulation Structure

As we mentioned previously, we divide our study into two parts: the influence of indium fluctuation on (1) emission spectrum and (2) carrier transport. In these two parts, we use different simulation structures. In part (1), since the 3D eigenvalue solver requires huge computer memory and calculating time, the simulation structure is relatively small. The simulation domain for a double QW LED was 30 nm × 30 nm, as shown in Figure 18.15a. The detail parameters setting can be found in Table 18.3. On the other hand, the simulation structure is an 80 nm × 80 nm 6-pair QW LED and included a 40-nm AlGaN EBL in part (2), as shown in Figure 18.15b. The detailed setting can also be found in Table 18.4. We considered a MQW with a 100-nm p-doped GaN layer and a 200-nm n-doped GaN layer, and the QW and barrier width are 3 and 10 nm, respectively. We focus pm a 450-nm MQW blue LED ($In_{0.17}Ga_{0.83}N$). In the uniform QW case, the indium composition is uniformly chosen to be 17%. In our case, as we mentioned previously, we assigned the indium fluctuation in each QW randomly. The maximum local indium composition, which is determined by the random number generator, was around 18%–19%.

A constant nonradiative lifetime was assumed to be 5.0×10^{-8} s, and the radiative recombination coefficient B_0 was assumed to be 2.0×10^{-11} cm^3/s [5,15,41]. Note that recombination rate is decided by B_0 $n(r)\cdot p(r)$ where the electron–hole overlap term is in the term of $n(r)\cdot p(r)$. In all cases, a 100% theoretical polarization value was applied [38]. The EBL thickness was 40 nm with 15% Al content AlGaN. For studying the carrier transport, we solved the 3D Poisson and drift–diffusion solver with the classical particle model. A larger area can be used in the modeling because we do not need to solve for the Schrödinger equation for the eigenstates for transport properties, which needs a much longer computing

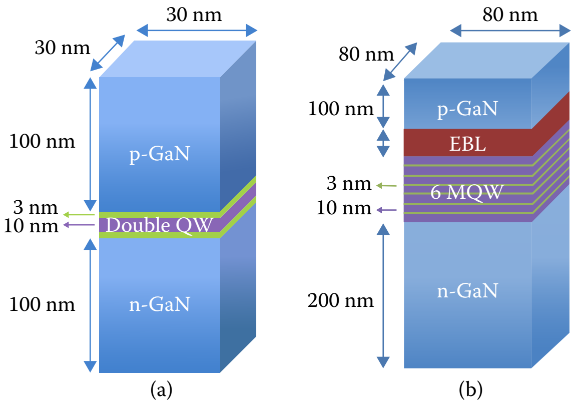


FIGURE 18.15 (a) The simulation structure in analyzing the emission spectrum. (b) The simulation structure in analyzing the carrier transport.

TABLE 18.3 The Material Parameter Settings of the Structure in Analyzing the Emission Spectrum

Unit	Thickness (nm)	Doping (1/cm³)	E_a (meV)	Impurity (1/cm³)	e Mobility (cm²/V · s)	h Mobility (cm²/V · s)
p-GaN	100	2.0×10^{19}	170	0.0	200	5
InGaN	3	0.0	0.0	1.0×10^{17}	600	10
GaN	10	0.0	0.0	1.0×10^{17}	200	10
InGaN	3	0.0	0.0	1.0×10^{17}	600	10
n-GaN	100	5.0×10^{18}	25	0.0	200	10

Note: The unit of thickness is *nm*. The unit of doping and impurity is 1/cm³. The unit of activation energy (E_a) is meV. The unit of electron and hole mobility is cm²/V·s.

time. We modeled 10 different fluctuation maps and took the averaged *I*–*V* and IQEs. Due to memory limitation, we cannot model the whole area of LED. Therefore, we only modeled the limited area in the p-i-n region, as shown in Figure 18.15b. Therefore, the sheet resistance in our simulation was not included due to the memory limitation. In general, the sheet resistance will cause voltage drops in the p-GaN and n-GaN layers so the calculated voltage across the active QW region is smaller than the externally applied bias in experiments. In addition, the real device structures also suffer current crowding effects, which will lead to a more serious droop effect than the result we obtain here.

18.4.2 The Influence of Indium Fluctuation on the Emission Spectrum

In this section, we discuss the influence of indium fluctuation on the emission spectrum. First, as we mentioned previously, the simulation structure we used in this section is a double QW LED and the simulation domain was 30 nm × 30 nm due to computer memory limitations required for 3D simulation, especially for the eigenvalue solver. To avoid the result being limited by the restricted area, we ran a set of different random cases and took the average results. First, the 3D Poisson and drift–diffusion solver developed by our lab was used to obtain a converged band potential at a fixed current density of 20 A/cm^2. Second, we solved the 3D Schrödinger equation with the calculated potential and obtained the confined eigenstate $E_{i,j}$.

Figure 18.16 shows the effects of randomly generated fluctuations in the QW for five different random maps. A total of 22 different random cases were run and averaged (dashed line in Figure 18.16). As we can see, the emission spectra broadened when the indium fluctuations were included in the simulation. The main reason is that the indium fluctuations form indium-rich regions in the QW (Figure 18.7a and b) with

TABLE 18.4 The Material Parameter Settings of the Structure in Analyzing the Carrier Transport

Unit	Thickness (nm)	Doping (1/cm^3)	E_a (meV)	Impurity (1/cm^3)	e Mobility (cm^2/V · s)	h Mobility (cm^2/V · s)
p-GaN	100	2.0×10^{19}	170	0.0	200	5
EBL	40	2.0×10^{19}	200	0.0	200	5
p-GaN	10	1.0×10^{18}	170	0.0	200	5
InGaN	3	0.0	0.0	1.0×10^{17}	600	10
GaN	10	0.0	0.0	1.0×10^{17}	200	10
InGaN	3	0.0	0.0	1.0×10^{17}	600	10
GaN	10	0.0	0.0	1.0×10^{17}	200	10
InGaN	3	0.0	0.0	1.0×10^{17}	600	10
GaN	10	0.0	0.0	1.0×10^{17}	200	10
InGaN	3	0.0	0.0	1.0×10^{17}	600	10
GaN	10	0.0	0.0	1.0×10^{17}	200	10
InGaN	3	0.0	0.0	1.0×10^{17}	600	10
GaN	10	0.0	0.0	1.0×10^{17}	200	10
InGaN	3	0.0	0.0	1.0×10^{17}	600	10
n-GaN	10	5.0×10^{18}	25	0.0	200	10
n-GaN	200	5.0×10^{18}	25	0.0	200	10

Note: The unit of thickness is *nm*. The unit of doping and impurity is 1/cm^3. The unit of activation energy (E_a) is meV. The unit of electron and hole mobility is (cm^2/V·s).

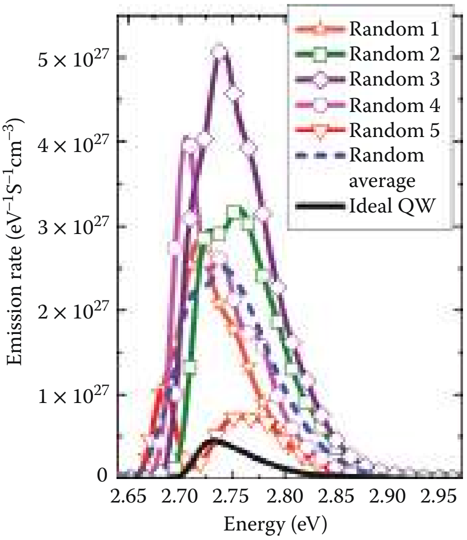


FIGURE 18.16 The calculated emission spectra when including the indium fluctuation. The dashed line is the average result of 22 random cases.

different bandgaps and confined energy levels. However, local peaks are observed in some cases (the random 2 and random 3 cases in Figure 18.16). This might be the result of the small sampling volume used in the simulation. In real devices with much larger area, these different local peaks merge into the overall spectral broadening. Figure 18.17b shows the calculated full width half maximum (FWHM) of the emission spectrum is around 80–120 meV at room temperature, which is close to the experimental results.

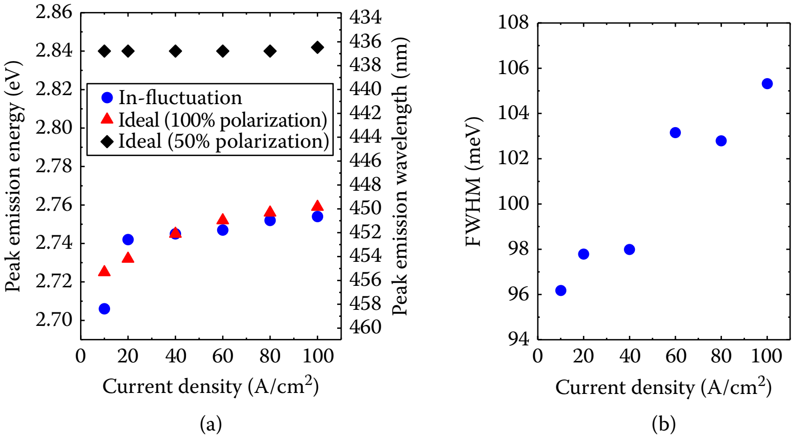


FIGURE 18.17 (a) The position of emission peak shift at different current density. (b) The full width half maximum (FWHM) of the emission spectrum. (From T.-J. Yang et al., *Journal of Applied Physics*, 116, 113104, 2014.)

Homogeneous broadening mechanisms usually will not lead to such large linewidths. If we compare the calculated intensity to the uniform QW case, the emission is stronger due to three factors occurring in the high In content regions where carriers localize in high indium region (deep potential): (1) there is less electron–hole separation; (2) there is more screening of electric field at a given current; and (3) there is an increased QW occupancy factor due to carrier localization. Figure 18.17a shows the blue shift of the spectrum when increasing the current density. A 40 meV (~ 7 nm) blue shift is observed when the current density increases from 10 to 100 A/cm^2, which is close to most experimental observations. Comparing with simulations of ideal QWs, assuming 50% polarization only leads to a very small blue shift due to a very reduced QCSE and the emission peak shifts to much shorter wavelength, while assuming a 100% polarization gives a reasonable shift, while, however, leading to very large forward bias (see the discussion in Section 18.4.3.1).

18.4.3 The Influence of Indium Fluctuation on the Carrier Transport and the Efficiency Droop

To model the vertical carrier transport in the 3D indium fluctuation cases, we used LEDs with six InGaN QWs with GaN barriers and included an AlGaN EBL, as shown in Figure 18.15b. We considered an MQW with a 100-nm p-doped GaN layer and a 200-nm n-doped GaN layer. The QW and barrier widths are 3 and 10 nm, respectively. We focus on 450-nm MQW blue LEDs ($In_{0.17}Ga_{0.83}N$). In the ideal QW case, the indium composition was uniformly 17%. In our case, we assigned the indium fluctuation in each QW randomly as described above. The maximum indium composition, which is determined by the random number generator, was around 18%–19%.

18.4.3.1 The Simulation Result of *c*-Plane LED

The simulation result is shown in Figure 18.18a. This is the calculated conduction band potential landscape. The indium-rich regions correspond to regions with a lower potential. The carriers will localize in the relatively lower potential regions because carriers inherently tend to stay in low-energy regions, as shown in Figure 18.18b. Consequently, the radiative recombination rate increases in the indium-rich region because of the carrier screening of the polarization fields, as shown in Figure 18.18c.

18.4.3.2 The *I*–*V* Curve

Next, we focused on the comparison of transport between the QW including indium fluctuation and uniform QW. Figure 18.19a shows the *I*–*V* curves. The forward voltage V_f used here is for a current density of

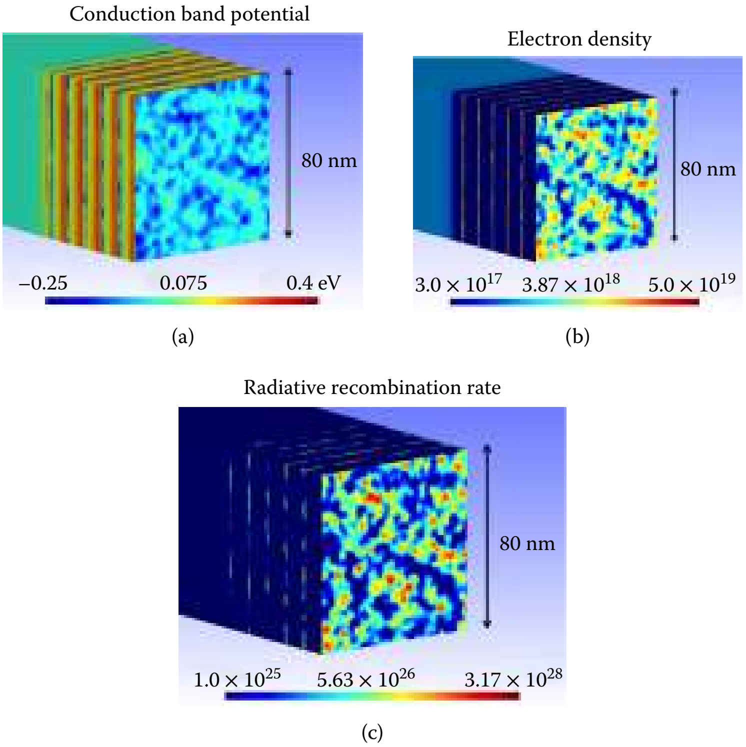

Conduction band potential
80 nm
−0.25
0.075
0.4 eV
(a)
Electron density
80 nm
3.0 × 10^17
3.87 × 10^18
5.0 × 10^19
(b)
Radiative recombination rate
80 nm
1.0 × 10^25
5.63 × 10^26
3.17 × 10^28
(c)

FIGURE 18.18 (a) The conduction band potential at 3.1 V. (b) The electron density at 3.1 V. (c) The radiative recombination rate at 3.1 V. (From T.-J. Yang et al., *Journal of Applied Physics*, 116, 113104, 2014.)

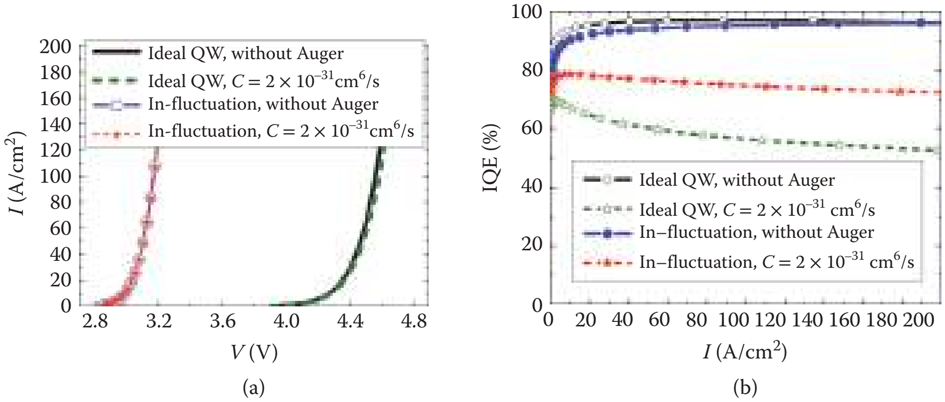
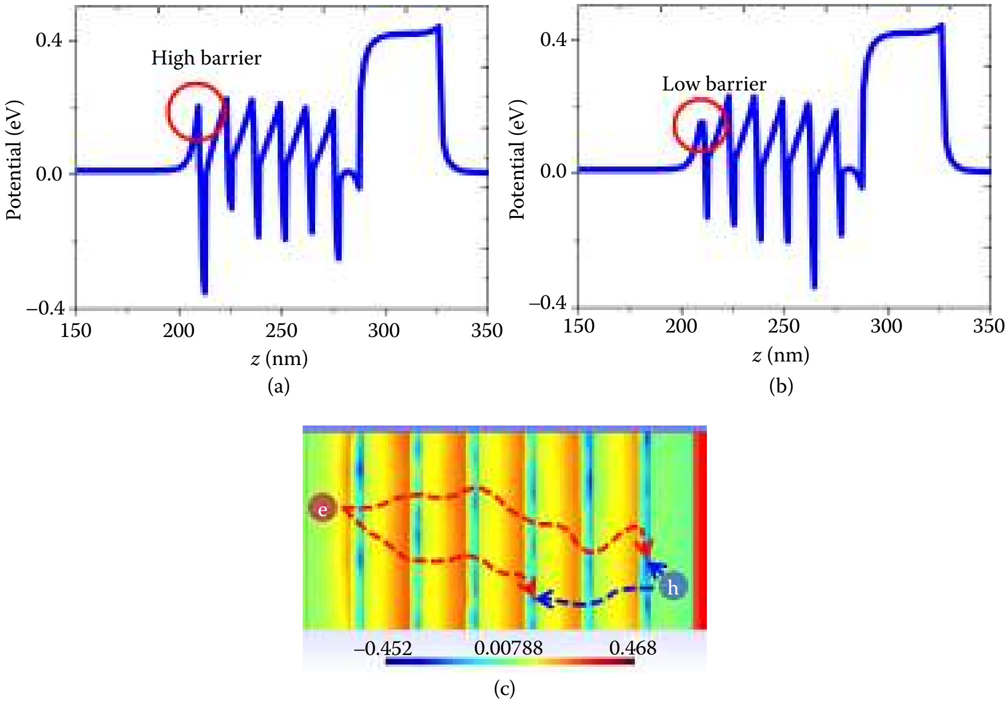


FIGURE 18.19 (a) The influence of indium fluctuation on the I–V curves. (b) The internal quantum efficiency (IQE) curves. (From T.-J. Yang et al., *Journal of Applied Physics*, 116, 113104, 2014.)

20 A/cm^2. For commercial blue light LEDs, V_f is around 2.8–3.0 V. For the ideal QW simulation (with uniform indium composition and 100% theoretical polarization value), V_f is almost 4.4 V, which is far larger than experimental results. This is also observed in most commercial simulation software with the same parameters set, which was also discussed in our twodimensional (2D) modeling results [23]. Some suggest including tunneling in the model to get a smaller V_f. But our calculation shows that using Wentzel–Kramers–Brillouin (WKB) tunneling between ideal QWs and 100% polarization cannot push V_f



(a)

(b)

(c)

FIGURE 18.20 (a) The high barrier in higher indium region in the first quantum well (QW). (b) The low barrier in low indium region in the first QW. (c) The illustration of carrier transport in the device.

lower than 3.7 V. On the other hand, when including indium fluctuations, we calculated a V_f of 3.05 V even without considering tunneling.

The reason why the V_f shifts to a reasonable value is that the indium composition will directly affect the bandgap potential in the QW due to different strength of the polarization field that in turn affects the bandgap potential in the QW. As we can see for the first QW in Figure 18.20a and b, the potential at the high indium region is lower than the potential at the lower indium region. However, due to the strong lattice mismatch, the piezo polarization charge will induce a triangular shape barrier between the GaN barrier and InGaN QW. The piezo polarization-inducing barrier height also depends on the indium composition. As shown in picture, the barrier at the high indium region is high due to the large polarization difference. On the other hand, the barrier at the low indium side will be reduced. For carriers, they are much easier to flow through the low barrier (low indium side) and localize in the low potential region (high indium side). This is the main reason why the *I*–*V* curve shifts to a value that matches experimental data more closely. Because the carriers will inherently find a percolation path and transport in the device, the indium fluctuation will strongly influence the carrier transport. Figure 18.20c illustrates the carrier transport in the device. Nevertheless, the V_f calculated is still slightly higher than the experimental result. Recent studies show that QW LEDs without *V*-pit or GaN substrate LED with very low dislocation density have a V_f around 3.2–3.4 eV, which is close to our prediction because the influence of V-pit has not yet been included in our simulation. In addition, the indium fluctuations of the piezoelectric field–induced GaN barrier might enhance the tunneling since the induced barrier height will be smaller at the lower indium composition site. This might further lower the V_f.

18.4.3.3 The Droop Behavior of IQE

Turning to the dependence of IQE on current, there are several reasons proposed for origins of droop, such as overflow [1–3], Auger recombination [4–9], and defects [10–14]. For the idealized uniform QW without considering the Auger effect and with EBL, the efficiency peak occurs at 80 A/cm^2 and at a very large bias (~ 4.5V). (Figure 18.19b) (considering only vertical transport and neglecting the influence of sheet resistance). Although the drift process becomes dominant at this high bias, the electrons are blocked by the EBL, and the overflow is weak. Even including the Auger recombination, the droop effect only occurs at 4.0 V, which is very large compared to commercial data sheets. Note that this large bias condition across the junction should not occur in the real device application. If we observe the most recent experimental results, we can find that the droop occurs at very low current density (~10 A/cm^2) and the applied bias is typically lower than 3.1 V. In the past, most researchers focused on the current density and disregarded the voltage except when discussing the wall plug efficiency. However, here the voltage plays a key role in understanding physical processes in LEDs.

When Auger recombination is excluded from the simulation, there is no droop even when the current is ~200 A/cm^2 (Figure 18.19b). This is due to the blocking action of EBL. With In fluctuations included, a small droop occurs when the current density is above 400 A/cm^2 and the applied voltage above 3.5 V is larger than V_{BI} and ΔE_c, which is unphysical as the resistive voltage drop should only occur in n and p-layer surrounding the depletion region. The biased voltage larger than built-in voltage V_{BI} including the ΔE_c should not be exist in the depletion region. Such voltages are very different from observed ones since the experimental droop typically occurs at 2.8–3.0 V from commercial data sheets (Figure 18.24).

At 3.0 V, the bias is smaller than the built-in voltage, V_{BI} (3.3 V in our case), and the carrier transport is mainly dominated by the diffusion process, as we can see in Figure 18.22a. The potential in the p-region

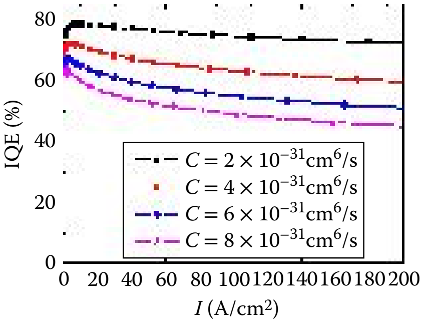


FIGURE 18.21 The internal quantum efficiency (IQE) curves with different Auger coefficients.

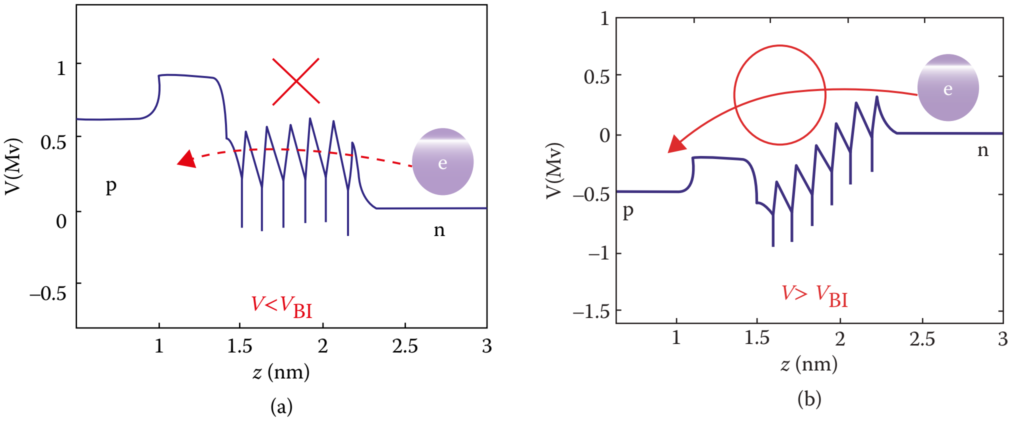


FIGURE 18.22 (a) When $V < V_{BI}$, it is mainly dominated by the diffusion process. It is harder for the carrier to overflow or overshoot the active region. (b) When $V > V_{BI}$, the carrier start to overflow or overshoot the device. However, the droop of a commercial grade light-emitting diode (LED) occurs at $V << V_{BI}$.

including the extra barrier height from the AlGaN EBL is much higher than the n-region and forms a large barrier to the carriers. Therefore, it is harder for carriers to spill over the barrier especially with the existence of the AlGaN barrier. Even with the polar optical phonon absorption process to gain the electron energy (~90 meV), it is still hard to overcome the 0.3–0.4 eV barrier. On the other hand, when the voltage exceeds the V_{BI}, as we can see in Figure 18.22b, the band structure reverses. In this situation, the carrier transport turns from the diffusion process to the drift process. The carriers start to overflow the device. Therefore, the simulation only sees the droop caused by overflow at a voltage higher than the built-in voltage (>3.3 V), if the Auger recombination is not considered. Therefore, the droop caused by the overflow only happens when the bias is larger than the built-in voltage or even larger bias when the AlGaN EBL is added.

Since the actual droop occurs at a very low bias in the experiments, we need to consider other factors that might cause droop effect at low bias. When the Auger recombination is included, the result shows that the droop occurs at 3.0 V, which matches closely to the experimental data. Consequently, we conclude that the droop effect is mainly dominated by the Auger recombination rate or other reasons such as carrier density dependent nonradiative recombination by defects.

As we can see in Figure 18.19a and b, when Auger recombination is included, droop occurs in both cases of ideal QW or QW incorporating In fluctuations. However, in the former case the bias voltage is again an unphysical 4.0 V, while in the latter case it is 3.0 V, which matches experimental data. Consequently, we conclude that the droop effect is mainly dominated by Auger recombination, enhanced by the effect of indium fluctuations. In addition, it should be pointed that due to the memory limitation, we calculated the vertical carrier transport within a limited area. In a real device, current crowding [42] issues can make local current density much higher than expected, making the droop effect worse, as discussed in earlier studies [43,44]. Figure 18.21 shows the IQE curves with the different Auger recombination coefficients, C. With larger values of C, the IQE curves show earlier droop onset, lower IQE peak, and more severe droop effects.

18.4.3.4 The Comparison of IQE between the Ideal QW and In-Fluctuation QW

In Figure 18.19b, the indium fluctuation case shows a good IQE performance with the higher IQE peak value and lower droop effect. Due to the carrier localization, the QCSE will be screened by the localized carrier charge. As a result, the better electron–hole overlapping will lead to a higher efficiency peak. On the other hand, in the ideal QW case, the IQE curve reaches the peak value when the applied bias is over 4.0 V, which is much larger than V_{BI} (3.3V). Therefore, the carrier leakage starts to affect the droop effect so the IQE curve shows a severe droop effect in the ideal QW case.

In conclusion, we prove that by considering the indium fluctuation in MQW LEDs, the electrical and optical properties are much closer to the commercial blue light LEDs. The droop behavior might be dominated by the Auger recombination at the lower current density since the voltage is much smaller than V_{BI}, and the ratio of the overflow mechanism might increase only when the bias is close to or larger than V_{BI} and the current density gets larger.

18.5 Carrier Transport in Green LEDs: Influence of Possible Imperfection in QWs

18.5.1 Introduction

Simulations with a proper model can provide some suitable evidence or clues to improve the optimization of devices. Before the 3D random alloy fluctuation model was proposed, scientists had focused on the current behavior of blue LEDs, because the I–V behavior could not be modeled well with 1D transport computer-aided design (CAD) software without considering the random alloy fluctuation. To explore the carrier transport and recombination mechanism in green LEDs, we conduct a preliminary examination of the electric property of green p-n LEDs.

18.5.2 Carrier Transport in Green LEDs Considering Indium Fluctuations Only

Initially, we constructed a green emitting p-n LEDs with five pairs of uniform 3-nm/10-nm $In_{0.27}Ga_{0.73}N$/GaN MQWs (517-nm green emission) and a fluctuated case for comparison. Both the cases are composed of 295-nm n-doped layer with 5×10^{18} cm^{-3} doping density and a 100-nm p-type layer with 2×10^{19} cm^{-3} doping density. Figure 18.23a shows the epilayers of all the simulated structure here. The chip size for the 3D case is 250 nm × 40 nm. The average composition for the 3D fluctuated case is 21%, as shown in Figure 18.23b, where the maximum composition of the fluctuated case is around 27%. Table 18.5 lists the input parameters of all the cases in this section. Figure 18.24 shows the calculated I–V curves of green LEDs. As expected, the 1D vertical transport model with a 100% theoretical polarization value failed to describe the experimental observation because of strong polarization fields. Even when the ideal polarization is reduced to 50%, the calculated result still shows a large difference with the extracted experimental values. However, the simulated green emitting LED with random alloy fluctuations also performs a larger driving voltage compared to the experimental result. Obviously, carriers still could not percolate through the fluctuated piezoelectric barriers in the simulation because the average compositions of QWs are so high that the piezoelectric fields will result in high potential barriers. The assumption of average 21% fluctuated QWs with 27% maximum composition might be far removed from the real devices. Our past study also shows that a much larger period of composition fluctuation will further reduce the turn-on voltage [23]. Therefore, it is worth taking further investigation.

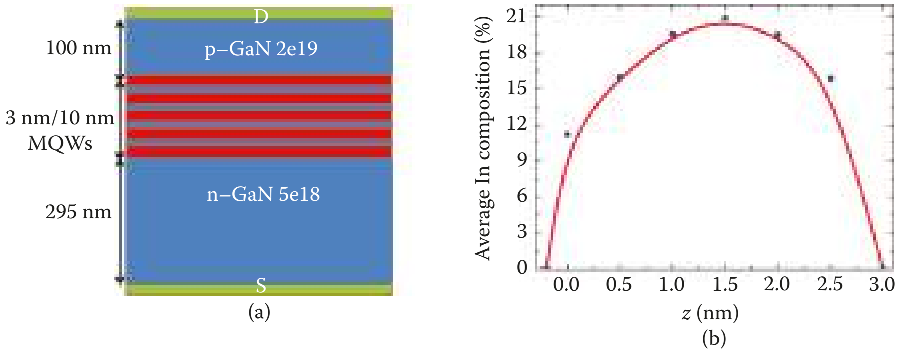


FIGURE 18.23 (a) The structure of the simulated green light-emitting diode (LED). (b) The average composition of the quantum well along the growth direction in the 3D fluctuated case.

TABLE 18.5 Detailed Parameters of Each Epilayer for Green Emission LEDs

Epilayer	n-GaN	i-InGaN/GaN	p-GaN
μ_h (cm^2/Vs)	10	10	5.0
μ_e (cm^2/Vs)	200	600	200
Doping (1/cm^3)	5.0×10^{18}	1.0×10^{17}	2.0×10^{19}
E_a (meV)	25	–	170
τ_n (s^{-1})	5.0×10^{-8}	5.0×10^{-8}	5.0×10^{-8}
τ_p (s^{-1})	5.0×10^{-8}	5.0×10^{-8}	5.0×10^{-8}
B_0(cm^3/s)	2.0×10^{-11}	2.0×10^{-11}	2.0×10^{-11}
C_0(cm^6/s)	2.0×10^{-31}	2.0×10^{-31}	2.0×10^{-31}

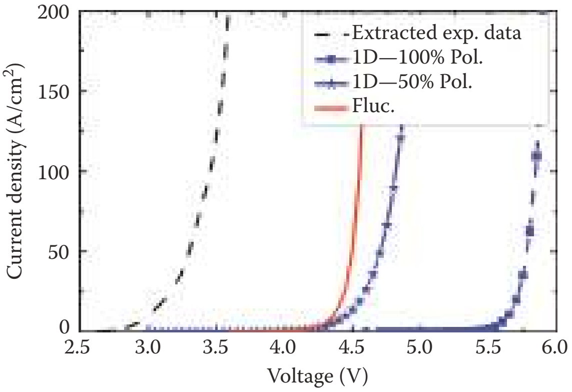


FIGURE 18.24 The *I–V* curves of the several models and the extracted experimental data. (From C.-H. Lu et al., *Journal of Alloys and Compounds*, 555, 250–254, 2013.)

18.5.3 Imperfection of QWs in Green LEDs

Some studies [47,48] have shown that the QWs are not perfect. The studies in Refs. [47,48] further indicated that large-scale well-width fluctuation exists in the commercial green LED samples. Inconsistent growth temperature between GaN layer and $In_xGa_{1-x}N$ might be the main reason for the interrupted void regions among the QWs. Especially when the indium composition increases, the difference of the required growth temperature with GaN will be larger. To further understand how the imperfect QWs affect the electrical property of green p-n LEDs, we apply the 3D model in examining the imperfection of QW in the green LED device.

18.5.4 3D Examination of Green LEDs with Imperfect QW Model and Random Indium Fluctuations Model

To examine the carrier transport in the void structures, we construct a 3D structure with random alloy fluctuations that is much closer to the real physics. The structure is with different void densities. The distance of void is 30 and 40 nm. The input parameters are listed in Table 18.5. Due to the huge computation time in the 3D calculation, the calculation area is limited to 250 nm × 40 nm in the lateral direction. All the epilayers are the same as mentioned previously. The side view and section view composition distributions are shown in Figure 18.25. Compared to the normal fluctuated QWs, the map with imperfect QWs will have some void regions in the QW region. The void regions are also not aligned with one another according to experimental observations [47,48]. Figure 18.26a shows that the 3D model with different void lengths can better model the V_f of the experimental data. The conduction band potential at 3.05 V of 3D fluctuation with 40-nm l_{void} shows that the piezoelectric barriers are smaller in the void region, which is shown in Figure 18.26b. In addition the piezoelectric fields near the normal fluctuated QW region have the insufficient impact on the void region. Hence, the carriers would prefer to percolate through the void regions into the active area, which might be closer to the real devices.

Figure 18.27 shows the calculated vector profile of both electron and hole currents. The results indicate that the void regions dominate the current flow path for both electron and hole currents. While carriers percolate into the void regions, other carriers might prefer to flow into the low potential areas locatied at the stripe QWs. As a result, the smaller V_fs in imperfect QW cases are attributed to an alternative percolation path provided by the void regions.

To conclude this section, we applied the 3D model to show that the void regions in the green emission LEDs indeed affect the carrier injection. The imperfect regions provide an additional injection path for the

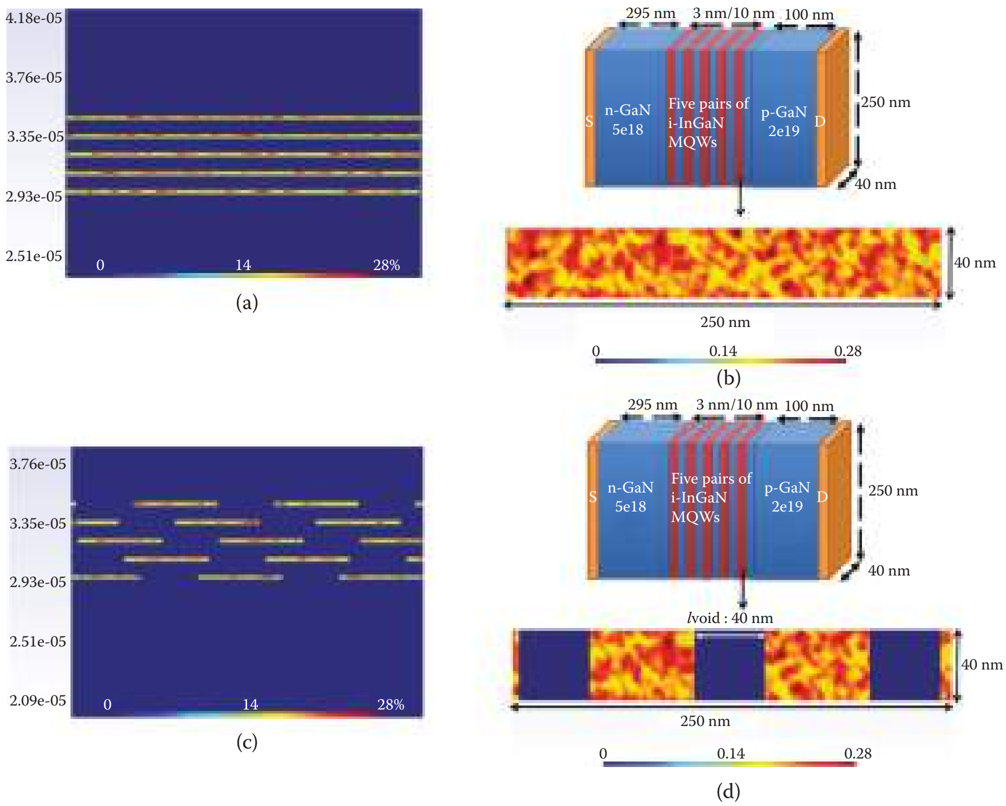

4.18e-05
3.76e-05
3.35e-05
2.93e-05
2.51e-05
0
14
28%
(a)
295 nm
3 nm/10 nm
100 nm
S
n-GaN
5e18
Five pairs of
i-InGaN
MQWs
p-GaN
2e19
D
250 nm
40 nm
40 nm
250 nm
0
0.14
0.28
(b)
3.76e-05
3.35e-05
2.93e-05
2.51e-05
2.09e-05
0
14
28%
(c)
295 nm
3 nm/10 nm
100 nm
S
n-GaN
5e18
Five pairs of
i-InGaN
MQWs
p-GaN
2e19
D
250 nm
40 nm
lvoid : 40 nm
40 nm
250 nm
0
0.14
0.28
(d)

FIGURE 18.25 (a) The side view of the composition map with normal fluctuated QWs. (b) The section view of the composition map with normal fluctuated QWs. (c) The side view of the composition map with imperfect fluctuated QWs. (d) The section view of the composition map with imperfect fluctuated QWs.

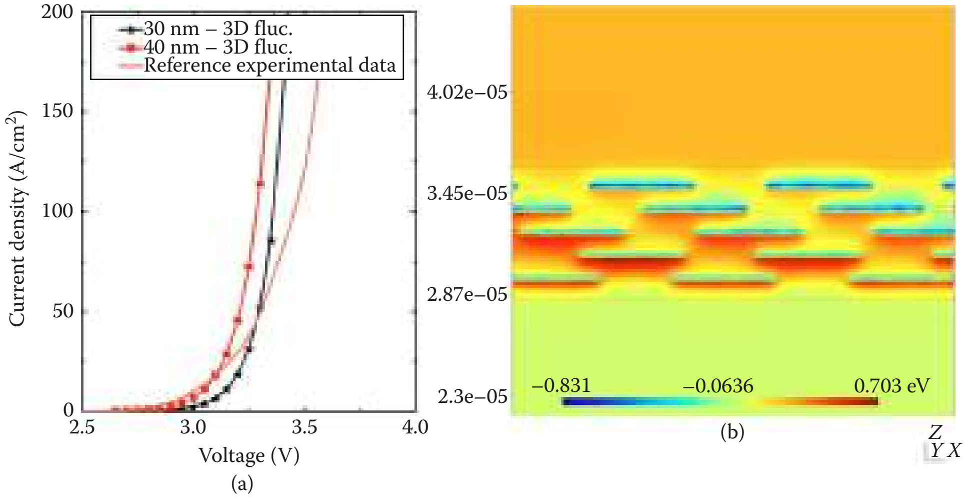
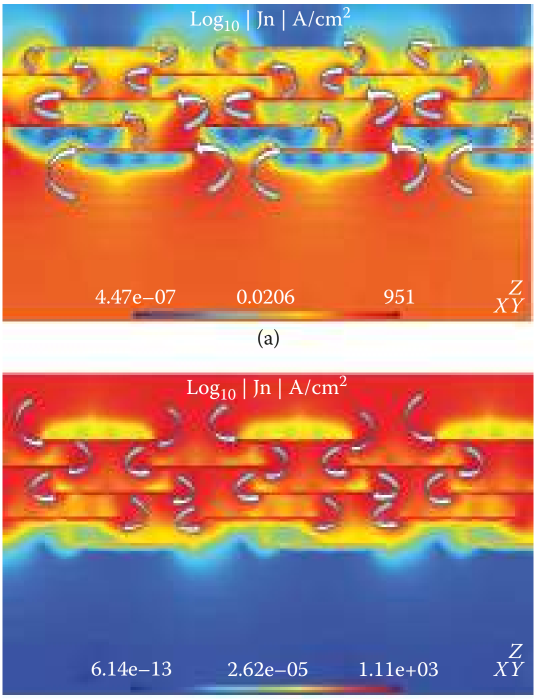


FIGURE 18.26 (a) The I–V curves of the different widths of the void region taking into consideration alloy fluctuations [45]. (b) The conduction band potential of the 40 nm case, where the location of the void regions is consistent, is at 3.05 V, which is near the turn-on point at 20 A/cm^2.



(a)

(b)

FIGURE 18.27 (a) The calculated vector profile of electron current. (b) The calculated vector profile of hole current.

carrier percolation. Although more reports of microscope observations in green emitting QWs are needed to prove that the structure of green LEDs are commonly imperfect, this preliminary simulated survey can still provide some hints for future modeling of the physical properties of green LEDs.

18.6 Summary

For the sake of simplicity, the simulation of the InGaN/GaN multiple QW LED usually assumes the uniform QW. However typical simulation results, such as the extremely high forward voltage (V_f), were far from the experimental data. These extra voltages will make the interpretation of carrier transport inaccurate. Because the extra voltage applied in the junction does not appear in the real device and will change the current transport mechanism from diffusion process into drift process. Therefore, it is important to make voltage fitting at accurate as possible. The simulation results show that by including the random indium fluctuation distribution in the QW, we can get a more reasonable fitting in the I–V curve.

For the transport issue with the random alloy fluctuation in the unipolar system, the simulation results of n-i-n GaN/InGaN/GaN MQW structures show that the fluctuated piezoelectric barrier induced by random alloy fluctuations will provide some current flow paths for the percolation transport, and the fluctuation model can give a reliable explanation for how the carriers perform in real devices. In modeling the emission spectrum, we found that the indium fluctuation will lead to a broader emission spectrum. The main reason is that the indium fluctuation forms several In-rich clusters, which look like quantum dots. The carriers localize in the QD-like regions and lead to different levels of quantum confinement. The effective bandgaps at each region vary from the different indium composition and the confined size of the indium clusters, which are the two important roles to affecting the discrete energy of quantum confinement. In addition, the emission strength will also be enhanced due to the better electron–hole localization effects.

We also analyzed the influence of indium fluctuations on the carrier transport. The simulation result of the *I*–*V* curve shifts to a value that is close to the experimental data. It indicates that the indium fluctuation does strongly affect the transport in the device. The main reason is that the low indium composition region will reduce the large potential barrier induced by the QCSE. For the uniform QW structure, the large V_f mainly comes from the strong band bending, which makes carriers hard to travel across the quantum barrier. For the indium fluctuation case, the reduction of the band bending at low indium regions provides a leaky way for carriers to flow through. The tunneling effect is not included in the indium fluctuation model but this will be a future project and may push the *I*–*V* curve to an even lower value. Furthermore, the IQE peaks occur at about 3.0 V, which also matches up to the experimental observation. Based on these simulation results, we tried to explain the cause of the droop effect. At a low bias condition, the droop effect is mainly dominated by the Auger recombination since the voltage is much lower than V_{BI}. When the bias gets higher and reaches the built-in voltage, the ratio of the droop influenced by carrier leakage will increase since the carrier transport turns from diffusion process into drift process at the reverse band condition.

Finally, we analyzed the green LED. We found that green emitting LEDs are difficult to use the traditional model to describe their physical property. Even for simulations that consider nanoscale alloy disorders, the calculated result differs from the experimental result. According to some reports, the green LEDs might exist in large-scale fluctuations, such as imperfect QWs in devices. Hence, we applied 3D models to examine the performance of the LEDs with imperfect QWs. The results show that the void region in the QWs provide an alternative path for carrier percolation. Thus, the IV of our simulation cases may approach the experimental observation more closely. In the future, we can model the green emission LEDs more elaborately based on the imperfect QW model.

References

1. H.-Y. Ryu, D.-S. Shin, and J.-I. Shim. Analysis of efficiency droop in nitride light-emitting diodes by the reduced effective volume of InGaN active material. *Applied Physics Letters*, 100(13):131109, 2012.
2. X. Ni, X. Li, J. Lee, S. Liu, V. Avrutin, A. Matulionis, U. Ozgur, and H. Morkoc. Pivotal role of ballistic and quasi-ballistic electrons on led efficiency. *Superlattices and Microstructures*, 48:133–153, 2010.
3. F. Akyol, D. N. Nath, S. Krishnamoorthy, P. S. Park, and S. Rajan. Suppression of electron overflow and efficiency droop in n-polar GaN green light emitting diodes. *Applied Physics Letters*, 100(11):111118, 2012.
4. F. Bertazzi, M. Goano, and E. Bellotti. A numerical study of Auger recombination in bulk InGaN. *Applied Physics Letters*, 97(23):231118, 2010.
5. K. T. Delaney, P. Rinke, and C. G. V. de Walle. Auger recombination rates in nitrides from first principles. *Applied Physics Letters*, 94(19):191109, 2009.
6. J. Hader, J. V. Moloney, B. Pasenow, S. W. Koch, M. Sabathil, N. Linder, and S. Lutgen. On the importance of radiative and Auger losses in GaN-based quantum wells. *Applied Physics Letters*, 92(26):261103, 2008.
7. J. Iveland, L. Martinelli, J. Peretti, J. S. Speck, and C. Weisbuch. Direct measurement of Auger electrons emitted from a semiconductor light-emitting diode under electrical injection: Identification of the dominant mechanism for efficiency droop. *Physical Review Letters*, 110:177406, April 2013.
8. H.-Y. Ryu, H.-S. Kim, and J.-I. Shim. Rate equation analysis of efficiency droop in InGaN light-emitting diodes. *Applied Physics Letters*, 95(8):081114, 2009.
9. Y. C. Shen, G. O. Mueller, S. Watanabe, N. F. Gardner, A. Munkholm, and M. R. Krames. Auger recombination in InGaN measured by photoluminescence. *Applied Physics Letters*, 91(14):141101, 2007.

10. J. Hader, J. V. Moloney, and S. W. Koch. Density-activated defect recombination as a possible explanation for the efficiency droop in GaN-based diodes. *Applied Physics Letters*, 96(22):221106, 2010.
11. J. Hader, J. V. Moloney, and S. W. Koch. Temperature-dependence of the internal efficiency droop in GaN-based diodes. *Applied Physics Letters*, 99(18):181127, 2011.
12. N. Okada, H. Kashihara, K. Sugimoto, Y. Yamada, and K. Tadatomo. Controlling potential barrier height by changing V-shaped pit size and the effect on optical and electrical properties for InGaN/GaN based light-emitting diodes. *Journal of Applied Physics*, 117(2):025708, 2015.
13. J. Smalc-Koziorowska, E. Grzanka, R. Czernecki, D. Schiavon, and M. Leszczyski. Elimination of trench defects and V-pits from InGaN/GaN structures. *Applied Physics Letters*, 106(10):101905, 2015.
14. H. Wang, X. Wang, Q. Tan, and X. Zeng. V-defects formation and optical properties of InGaN/-GaN multiple quantum well LED grown on patterned sapphire substrate. *Materials Science in Semiconductor Processing*, 29:112–116, 2015.
15. E. Kioupakis, P. Rinke, K. T. Delaney, and C. G. Van de Walle. Indirect Auger re-combination as a cause of efficiency droop in nitride light-emitting diodes. *Applied Physics Letters*, 98(16):161107, 2011.
16. S. E. Bennett, D. W. Saxey, M. J. Kappers, J. S. Barnard, C. J. Humphreys, G. D. Smith, and R. A. Oliver. Atom probe tomography assessment of the impact of electron beam exposure on InxGa1$-x$N/GaN quantum wells. *Applied Physics Letters*, 99(2):021906, 2011.
17. D. A. Browne, B. Mazumder, Y.-R. Wu, and J. S. Speck. Electron transport in unipolar InGaN/GaN multiple quantum well structures grown by NH$_3$ molecular beam epitaxy. *Journal of Applied Physics*, 117(13):185703, 2015.
18. M. J. Galtrey, R. A. Oliver, M. J. Kappers, C. J. Humphreys, P. H. Clifton, D. Larson, D. W. Saxey, and A. Cerezo. Three-dimensional atom probe analysis of green and blue-emitting InxGa1$-x$N/GaN multiple quantum well structures. *Journal of Applied Physics*, 104(1):013524, 2008.
19. M. J. Galtrey, R. A. Oliver, M. J. Kappers, C. J. Humphreys, D. J. Stokes, P. H. Clifton, and A. Cerezo. Three-dimensional atom probe studies of an InxGa1$-x$N/GaN multiple quantum well structure: Assessment of possible indium clustering. *Applied Physics Letters*, 90(6):061903, 2007.
20. J. R. Riley, T. Detchprohm, C. Wetzel, and L. J. Lauhon. On the reliable analysis of indium mole fraction within InxGa1$-x$N quantum wells using atom probe tomography. *Applied Physics Letters*, 104(15):152102, 2014.
21. R. Shivaraman, Y. Kawaguchi, S. Tanaka, S. DenBaars, S. Nakamura, and J. Speck. Comparative analysis of 2021 and 2021 semipolar GaN light emitting diodes using atom probe tomography. *Applied Physics Letters*, 102(25):251104, 2013.
22. D. Watson-Parris, M. J. Godfrey, P. Dawson, R. A. Oliver, M. J. Galtrey, M. J. Kappers, and C. J. Humphreys. Carrier localization mechanisms in InxGa1$-x$N/GaN quantum wells. *Physical Review B*, 83:115321, March 2011.
23. Y.-R. Wu, R. Shivaraman, K.-C. Wang, and J. S. Speck. Analyzing the physical properties of InGaN multiple quantum well light emitting diodes from nano scale structure. *Applied Physics Letters*, 101(8):083505, 2012.
24. T.-J. Yang, R. Shivaraman, J. S. Speck, and Y.-R. Wu. The influence of random indium alloy fluctuations in indium gallium nitride quantum wells on the device behavior. *Journal of Applied Physics*, 116(11):113104, 2014.
25. M. A. Caro, S. Schulz, and E. P. O'Reilly. Theory of local electric polarization and its relation to internal strain: Impact on polarization potential and electronic properties of group-III nitrides. *Physical Review B*, 88:214103, December 2013.
26. C.-K. Wu, C.-K. Li, and Y.-R. Wu. Percolation transport study in nitride based LED by considering the random alloy fluctuation. *Journal of Computational Electronics*, 14(2):416–424, 2015.

27. C.-K. Li, H.-C. Yang, T.-C. Hsu, Y.-J. Shen, A.-S. Liu, and Y.-R. Wu. Three dimensional numerical study on the efficiency of a core-shell InGaN/GaN multiple quantum well nanowire light-emitting diodes. *Journal of Applied Physics*, 113(18):183104, 2013.
28. C.-K. Li, P.-C. Yeh, J.-W. Yu, L.-H. Peng, and Y.-R. Wu. Scaling performance of Ga_2O_3/GaN nanowire field effect transistor. *Journal of Applied Physics*, 114(16):163706, 2013.
29. C. Geuzaine and J.-F. Remacle. Gmsh: A 3-D finite element mesh generator with built-in pre- and post-processing facilities. *International Journal for Numerical Methods in Engineering*, 79:1309–1331, 2009.
30. M. Sabathil, A. Laubsch, and N. Linder. Self-consistent modeling of resonant PL in InGaN SQW LED-structure. *SPIE Proceedings*, 6486:64860V–648609, 2007.
31. I. Vurgaftman, J. R. Meyer, and R. L. R. Mohan. Band parameters for III–V compound semiconductors and their alloys. *Journal of Applied Physics*, 89(11):5815–5875, 2001.
32. S. L. Chuang and C. S. Chang. k p method for strained wurtzite semiconductors. *Physical Review B*, 54:2491–2504, July 1996.
33. J. Wu. When group-III nitrides go infrared: New properties and perspectives. *Journal of Applied Physics*, 106(1):011101, 2009.
34. T. R. Chandrupatla., et al. *Introduction to Finite Elements in Engineering*. Vol. 2. Upper Saddle River, NJ: Prentice Hall, 2002.
35. A. Romanov, T. Baker, S. Nakamura, and J. Speck. Strain-induced polarization in wurtzite III-nitride semipolar layers. *Journal of Applied Physics*, 100(2):023522, 2006.
36. Y.-R. Wu, Y.-Y. Lin, H.-H. Huang, and J. Singh. Electronic and optical properties of InGaN quantum dot based light emitters for solid state lighting. *Journal of Applied Physics*, 105(1):013117, 2009.
37. D. A. Browne, B. Mazumder, Y.-R. Wu, and J. S. Speck. Investigation of electron transport through InGaN quantum well structures. In *14th Electronic Materials Conference,* Santa Barbara, CA, June 25–27, 2014.
38. O. Ambacher, J. Majewski, C. Miskys, A. Link, M. Hermann, M. Eickhoff, M. Stuzmann, F. Bernardini, V. Fiorentini, V. Tilak, B. Schaff, and L. F. Eastman. Pyroelectric properties of Al(In)GaN/GaN hetero- and quantum well structures. *Journal of Physics: Condensed Matter*, 14:3399–3434, 2002.
39. O. Ambacher, J. Smart, J. R. Shealy, N. G. Weimann, K. Chu, M. Murohy, W. J. Schaff, L. F. Eastman, R. Dimitrov, L. Wittmer, M. Stutzmann, W. Rieger, and J. Hilsenbeck. Two-dimensional electron gases induced by spontaneous and piezo-electric polarization charges in N- and Ga-face AlGaN/GaN heterostructures. *Journal of Applied Physics*, 85(6):3222–3233, March 1999.
40. D. N. Nath, Z. C. Yang, C.-Y. Lee, P. S. Park, Y.-R. Wu, and S. Rajan. Unipolar vertical transport in GaN/AlGaN/GaN heterostructures. *Applied Physics Letters*, 103(2):022102, 2013.
41. J. Piprek. Efficiency droop in nitride-based light-emitting diodes. *Physica Status Solidi (A)*, 207(10):2217–2225, 2010.
42. H.-Y. Ryu and J.-I. Shim. Effect of current spreading on the efficiency droop of InGaN light-emitting diodes. *Optics Express*, 19(4):2886–2894, February 2011.
43. M. Calciati, M. Goano, F. Bertazzi, M. Vallone, X. Zhou, G. Ghione, M. Meneghini, G. Meneghesso, E. Zanoni, E. Bellotti, G. Verzellesi, D. Zhu, and C. Humphreys. Correlating electroluminescence characterization and physics-based models of InGaN/GaN LEDs: Pitfalls and open issues. *AIP Advances*, 4(6):067118, 2014.
44. C. Kang Li, M. Rosmeulen, E. Simoen, and Y.-R. Wu. Study on the optimization for current spreading effect of lateral GaN/InGaN LEDs. *IEEE Transactions on Electron Devices*, 61(2):511–517, February 2014.
45. C.-H. Lu, Y.-C. Li, Y.-H. Chen, S.-C. Tsai, Y.-L. Lai, Y.-L. Li, and C.-P. Liu. Out-put power enhancement of InGaN/GaN based green light-emitting diodes with high-density ultra-small In-rich quantum dots. *Journal of Alloys and Compounds*, 555:250–254, 2013.

46. W. Lv, L. Wang, L. Wang, Y. Xing, D. Yang, Z. Hao, and Y. Luo. InGaN quantum dot green light-emitting diodes with negligible blue shift of electroluminescence peak wavelength. *Applied Physics Express*, 7(2):025203, 2014.
47. N. Van der Laak, R. Oliver, M. Kappers, and C. Humphreys. Characterization of InGaN quantum wells with gross fluctuations in width. *Journal of Applied Physics*, 102(1):013513, 2007.
48. N. K. van der Laak, R. A. Oliver, M. J. Kappers, and C. J. Humphreys. Role of gross well-width fluctuations in bright, green-emitting single InGaN/GaN quantum well structures. *Applied Physics Letters*, 90(12):121911–121911, 2007.

19

Superluminescent Light-Emitting Diodes

19.1 Introduction 589
19.2 Design and Modeling of SLED Devices 591
SLED Geometry and Epitaxial Design • Modeling and Simulation Approaches • *L–I* Characteristics of Reflecting SLEDs • Experimental and Numerical Verification of the Analytical Model
19.3 SLED Designs with Specified Targets 600
Broad-Bandwidth SLED Designs • Coherence Function and Coherence Length • High-Power SLED Designs
19.4 Summary 605

Nicolai Matuschek
and
Marcus Duelk

19.1 Introduction

More than 30 years ago, the concept of a superluminescent light-emitting diode (SLED) was proposed for the first time (Kaminow and Marcuse 1983). It can be briefly described as an edge-emitting semiconductor light source that operates in the so-called superluminescence regime, also known as the amplified spontaneous emission (ASE) regime. This means that the gain medium is pumped to a level beyond transparency but below the threshold for starting lasing activity. As a consequence of this particular operating range, SLEDs combine some aspects of the electro-optical performance from standard light-emitting diodes (LEDs) and laser diodes (LDs). They are able to produce high-output powers with high brightness similar to high-power LDs. On the other hand, they show a broadband emission spectrum similar to LEDs. With respect to their coherence properties, this translates into high spatial coherence similar to LDs and low temporal coherence like LEDs. A more detailed comparison is given in Rossetti et al. (2012).

Due to these intermediate characteristics, SLEDs are preferred light sources for many applications, including fiber-optic gyroscopes (FOGs) (Burns et al. 1983), fiber-optic current sensors (FOCSs) (Bohnert et al. 2002), optical coherence tomography (OCT) (Schmitt 1999; Drexler and Fujimoto 2008), structural health monitoring with optical fiber sensors (Wild and Hinckley 2009), speckle-free illumination (Rossetti et al. 2012), metrology systems (Dufour et al. 2005), or optical test equipment for fiber-optic networks (Senior 2009). In order to meet the different requirements of these applications, broadband SLEDs at various emission wavelengths with different spectral bandwidths and shapes are used. SLEDs can be designed for wavelengths ranging from 390 nm up to 2700 nm. They can be realized in GaN (390–570 nm), GaAs (570–1150 nm), InP (1150–2000 nm), or GaSb (2000–2700 nm). Figure 19.1 shows a selection of typical ASE spectra obtained for various commercially available SLED modules operating at center wavelengths from around 400 to 1550 nm (Exalos 2017).

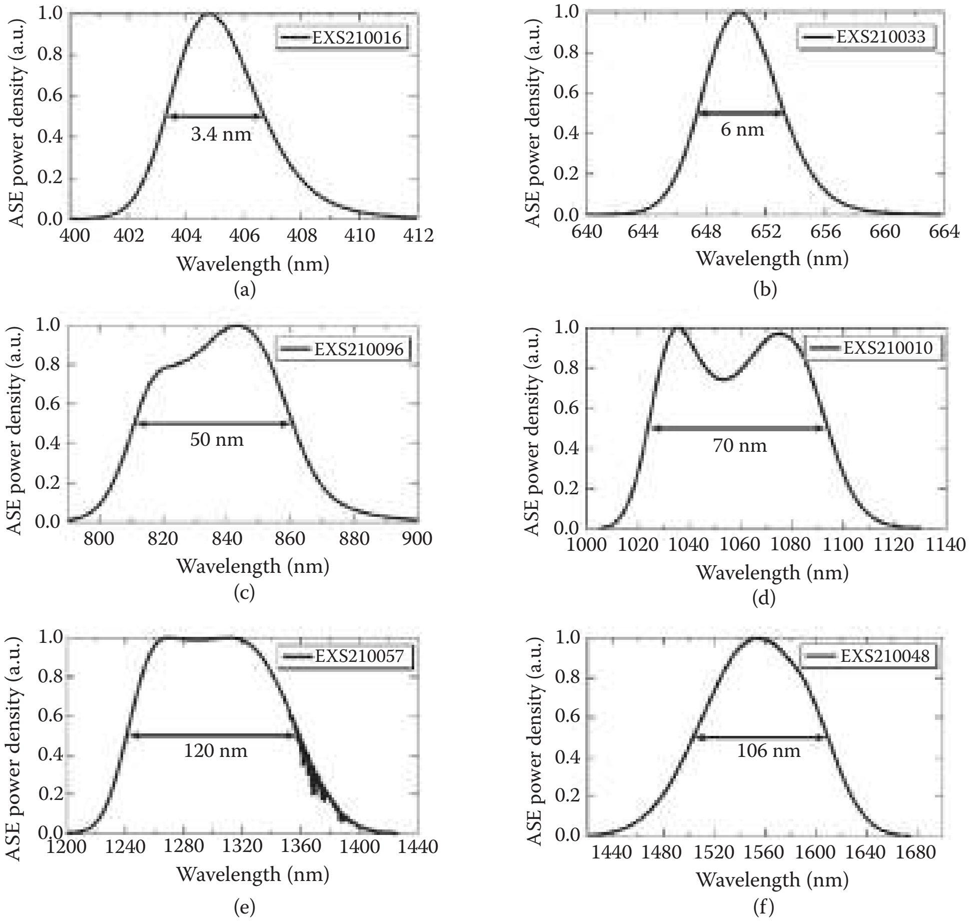


FIGURE 19.1 Measured amplified spontaneous emission (ASE) spectra (on a linear scale) obtained for six different superluminescent light-emitting diode (SLED) modules operating at relevant wavelength ranges from about 400 to 1550 nm. The full-width at half maximum (FWHM) or 3-dB bandwidth is given in the diagrams. The Gaussian-shaped spectra shown in (a), (b), and (f) are based on a single transition epitaxial design, whereas the spectra with two humps shown in (c), (d), and (e) are based on a multitransition epitaxial design, as explained in Section 19.3.1.

Traditionally, the vast majority of SLEDs are employed in systems with light emission in the near infrared from 750 to 1600 nm (see Figure 19.1c through f). Over the last decade, the emerging interest in speckle-free red–green–blue (RGB) applications forced the development of SLEDs operating in the visible wavelength range (see Figure 19.1a and b). Based on the technology for red LDs, which has been well established over the last 30 years, the demonstration of a red SLED realized in GaAs was a straightforward task (Semenov et al. 1993a). In contrast, the development of blue and green SLEDs has been confronted with a couple of intrinsic technological difficulties (Rossetti et al. 2012). One of them is the choice of the substrate material, which can be GaN, SiC, sapphire, or others. For SLEDs and LDs, high-quality substrate materials with good thermal conductivity and low dislocation densities are considered to be of utmost importance. Therefore, past development activities have focused on the growth of the epitaxial layer structure on free-standing (mainly *c*-plane) GaN substrates. The first GaN-based SLEDs emitting in the blue violet spectral region could be demonstrated just a few years ago (Feltin et al. 2009; Rossetti et al. 2010), whereas research and development for the demonstration of green SLEDs is still ongoing. Very recently, the first long-wavelength SLED realized in GaSb for optical sensing at 2.4 μm has been demonstrated (Wootten et al. 2014).

19.2 Design and Modeling of SLED Devices

19.2.1 SLED Geometry and Epitaxial Design

Typically, SLEDs are realized as epitaxially grown layers on a substrate in an index-guided ridge-waveguide geometry similar to Fabry–Pérot-type LDs. The latter consists of a straight waveguide with optical dielectric coatings applied at the facets of the chip, typically a high-reflective (HR) at the back facet and a partial-reflective (PR) coating at the front facet (see, e.g., Macleod 2010). The straight waveguide in combination with reflecting facets defines a resonant cavity that allows for the build-up of longitudinal cavity modes and laser oscillations. In contrast, several tricks are applied to SLEDs for the suppression of resonant cavity modes in order to minimize modulations in the ASE spectrum (so-called spectral ripples) or, even worse, to prevent the starting of lasing. The most popular and very efficient way to achieve this is the tilting of the waveguide by a few degrees with respect to the normal of the facets (Alphonse et al. 1988) in combination with the application of antireflection (AR) coatings (Macleod 2010) at the facet sides. The residual net modal reflectivity of such a design can be calculated using the analytical model presented in Marcuse (1989). Exemplary reflectivity calculations as a function of the tilt angle are given in Matuschek and Duelk (2013). Other methods are based on the incorporation of an absorber section at the back-facet side of the waveguide (Patterson et al. 1994; Kwong et al. 2008), the use of a bent (Semenov et al. 1993b), or tapered (Koyama et al. 1993) waveguide structure, or a combination of various methods (see, e.g., Lee et al. 1973; Nagai et al. 1989; Semenov et al. 1993a; Middlemast et al. 1997).

The ASE process occurs in the active region along the SLEDs waveguide structure. The most common approach to realize a positive material gain is based on a single-quantum well (SQW) or multiquantum well (MQW) active region design (see, e.g., Chuang 1995). For some applications, a bulk layer approach might be advantageous, as it is the case, for example, for SLEDs that require a low polarization extinction ratio (PER),[†] that is, the amount of transverse electric (TE)- and transverse magnetic (TM)-polarized output power should be as equal as possible (Heo et al. 2011). Quantum-dot epi structures for SLEDs have been proposed (Sun et al. 1999) and demonstrated a few times as well. However, this technology has various severe limitations, like a limited wavelength range (so far 1050–1200 nm), low differential gain requiring long SLED chips and large drive currents, manufacturing challenges, and others (Rossetti et al. 2008).

Figure 19.2 shows the schematic front view of a typical SLED structure for an SQW epitaxial layer design. Generally, quantum well (QW) active-region designs allow for the use of strained layers. Compressively strained QW layers are required for SLED designs showing a high PER greater than ~10 dB, which is desired for many applications. Maximum strain values up to 3% (with respect to the lattice constant of the substrate material) may be realized in practice. Higher strain values may lead to the formation of dislocations and are therefore detrimental to achieving a good long-term reliability. In addition to the relative strain of an individual layer, the total integrated net strain for the entire active region must be taken into account too. Hence, strain compensation by incorporation of tensile strained barrier layers is a useful option for strain relaxation over the active region (Tansu and Mawst 2001).

The QW and barrier layers are sandwiched between (undoped or partially doped) waveguide layers that are surrounded by highly doped *n*- and *p*-cladding layers. The high refractive index of the active-region and waveguide layers leads to the confinement of the optical mode in vertical direction. On the other hand, the ridge-waveguide geometry ensures the lateral mode confinement. As a result, the epitaxial layer stack realized in a ridge-waveguide geometry allows for the build-up of transverse optical modes with a lateral extension roughly given by the ridge width and a vertical extension as given by the thickness of the waveguide layers. For almost all applications, single-mode operation on the fundamental optical mode is preferred. Hence, care has to be taken to avoid the occurrence of higher-order modes or leaky substrate

---

[†] The PER is defined as $\mathrm{PER} = 10 \cdot \lg\left(\frac{P_{\mathrm{TE}}}{P_{\mathrm{TM}}}\right)$ dB, where P_{TE} and P_{TM} is the SLED's total output power from the front facet in TE and TM polarization, respectively.

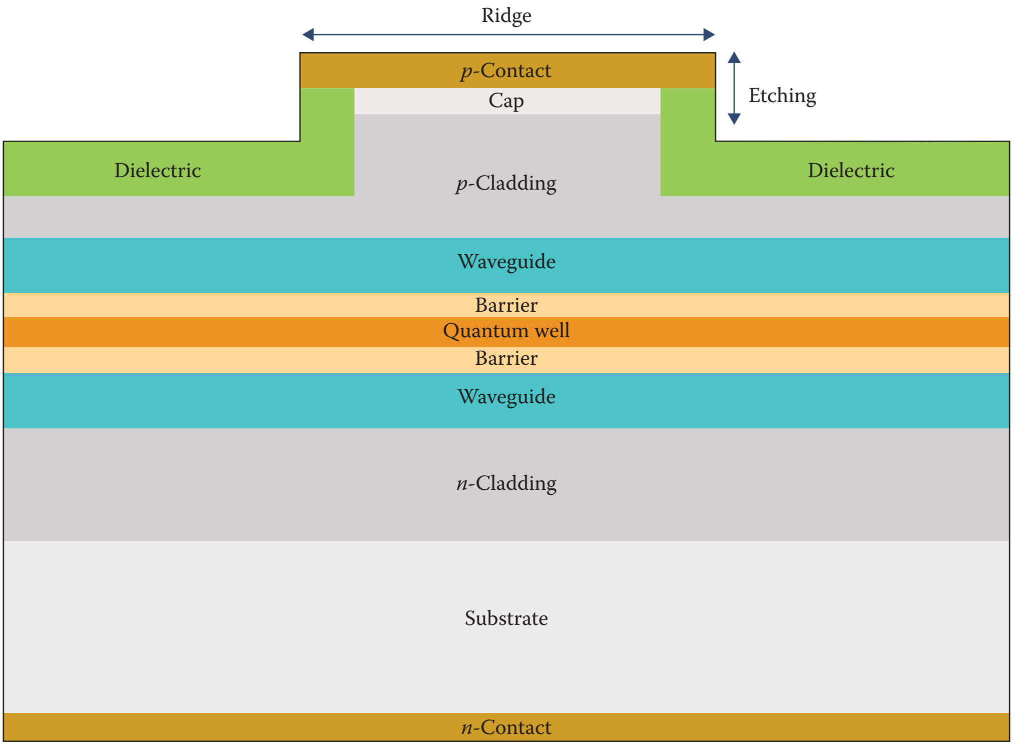


FIGURE 19.2 Basic layer structure of a single-quantum well (SQW) epitaxy. The QW and barrier layers are sandwiched between waveguide layers followed by the cladding layers on both sides. The ridge-waveguide geometry ensures the lateral guiding of the optical mode, whereas the vertical guiding is a consequence of the refractive-index profile defined by the material composition of the different layers.

modes. The latter are particularly a problem for GaN-based epi structures in the visible due to the high refractive index of the substrate material (Laino et al. 2007; Matuschek and Duelk 2013). Moreover, some output power might be lost into higher-order lateral modes if such modes are supported by the lateral waveguide (see, e.g., Coldren et al. 2012). The number of supported lateral modes, and thus, the corresponding power loss scales with the width of the ridge waveguide. Therefore, the stripe width is typically chosen to be smaller than 10 μm.

As mentioned earlier, the appropriate choice of semiconductor materials depends primarily on the wavelength range of interest, and thus, on the bandgap energy of available semiconductor compound material. In Table 19.1, the substrate materials and the most commonly used compound materials for the functional layers of the epitaxial structure, as shown in Figure 19.2, are summarized for the different wavelength regimes.

19.2.2 Modeling and Simulation Approaches

A couple of different approaches exist for the modeling and simulation of SLEDs. Generally, the analytical insight into the electro-optical performance decreases with the complexity of the set of coupled equations taken under consideration. On the other hand, the accuracy increases when more effects are taken into account.

Commercially available packages (Synopsis 2017; Crosslight Software 2017; Photon Design 2017) allow for full three-dimensional (3D) simulations. They are based on the coupling of solvers for the various problems involved in the simulation of an SLED (so-called multiphysics treatment) (Li and Li 2010). Typically,

TABLE 19.1 Materials Used for the Epitaxial Layers of an SLED Grown on a Substrate Depending on the Wavelength Range of Interest

Wavelength Range	390–570 nm	570–1150 nm	1150–2000 nm	2000–2700 nm
Substrate	GaN	GaAs	InP	GaSb
Cladding	$Al_xGa_{1-x}N$	$Al_xGa_{1-x}As$ $Al_xGa_yIn_{1-x-y}P$	InP $Al_xIn_{1-x}As$	$Al_xGa_{1-x}As_{1-y}Sb_y$
Waveguide	GaN	$Al_xGa_{1-x}As$ $Al_xGa_yIn_{1-x-y}P$	$Al_xGa_yIn_{1-x-y}As$ $In_xGa_{1-x}As_{1-y}P_y$	$Al_xGa_{1-x}As_{1-y}Sb_y$
Active region (QWs + barriers)	$In_xGa_{1-x}N$	$Al_xGa_yIn_{1-x-y}P$ $Al_xGa_yIn_{1-x-y}As$ $In_xGa_{1-x}As_{1-y}P_y$	$Al_xGa_yIn_{1-x-y}As$ $In_xGa_{1-x}As_{1-y}P_y$	$Al_xGa_{1-x}As_{1-y}Sb_y$ $In_xGa_{1-x}As_{1-y}Sb_y$

Note: The general definition of the quaternary materials includes the special case that the materials may reduce to ternary or binary materials if the material fractions x and/or y are equal to 0 or 1.

the quantum-mechanical problem is described using the $\mathbf{k}\cdot\mathbf{p}$ method and the optical modes are found from a Helmholtz equation (Chuang 1995). These equations are coupled to the semiconductor equations, for example, as Poisson equation and drift-diffusion equation and to a heat-transfer equation if the thermal properties are also taken into account (Loeser and Witzigmann 2008). All equations are then solved simultaneously and self-consistently for each iteration step, yielding a solution for all local and global variables of the system.

In analytical approaches both the physical dimensionality of the problem and the number of coupled equations are reduced. They are normally based on a traveling-wave equation for the electric field (Park and Li 2006) or the optical power (or photon density) (Matuschek and Duelk 2013; Milani et al. 2015) and on rate equations for the carrier density. The material gain is either calculated from quantum-mechanical equations or using simplified semianalytical equations assuming a logarithmic or linear gain dependence on the carrier density (Coldren et al. 2012). In order to calculate the net modal gain, both the optical confinement factor and the internal loss coefficient are required. The first can be calculated from the overlap integral of the optical modes with the active layers. The latter can be used, for instance, as a free fitting parameter or it can be extracted from an *inverse efficiency versus chip length* plot, which is obtained from the *L–I* characteristics of LDs measured for various chip lengths. The LD design for such measurements differs from the SLED design only in the straight ridge waveguide and its facets are typically uncoated.

19.2.3 *L–I* Characteristics of Reflecting SLEDs

SLEDs operate in the ASE regime, that is, spontaneously emitted photons induce the process of stimulated emission of photons while traveling along the waveguide. Hence, the number of forward- and backward-propagating photons grows exponentially in direction to both facets of the chip. On the other hand, the SLED's *L–I* characteristic follows to good approximation a power law with chip length–dependent exponent because of the logarithmic gain dependence on current density as shown in Matuschek and Duelk (2013). The analytic derivation of this power law was based on the assumption of an ideal SLED without any residual reflectivity at both facet sides.

Here, we want to extend this approach to the general case of SLEDs with arbitrary effective modal reflectivities at the chip's front- and back-facet sides. This derivation includes special cases of standard SLEDs and of so-called reflecting SLEDs (R-SLEDs). Typical SLEDs feature a tilted ridge-waveguide geometry with tilt angles smaller than 15° and high-quality AR coatings deposited at both facet sides in order to avoid any resonant cavity effect. In contrast, R-SLEDs consist of a tilted waveguide section in direction to the front facet with deposited AR coating and a straight waveguide section toward the chip's back facet with a deposited HR coating (Matuschek and Duelk 2014), as shown in Figure 19.3. As explained below in more detail, R-SLEDs are able to produce much higher output powers due to the double-pass geometry

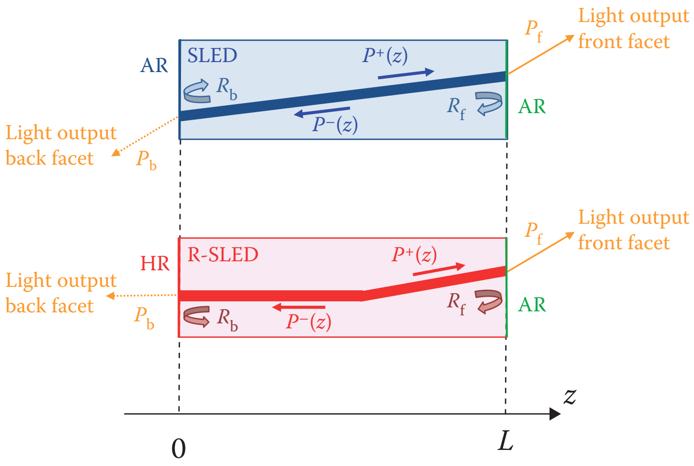


FIGURE 19.3 Schematic top view of an SLED and an R-SLED with effective modal reflectivities R_f and R_b. The power is exponentially amplified in both directions along the (partially) tilted ridge-waveguide structure, where P^+ denotes the power distribution for forward-propagating photons and P^- for backward-propagating photons, respectively. P_f and P_b are the powers, which are coupled out and can be measured at the front- and back-facet sides of the SLED or R-SLED. The z-axis is defined by the propagation direction of photons along the ridge-waveguide structure, so that the single-path length is slightly greater than the vertical distance between the facets due to the tilt.

compared to standard SLEDs under same operating conditions. However, the drawback is that such devices are much more susceptible to the appearance of ripples in the ASE spectrum and/or starting of undesired lasing operation.

In the following, the longitudinal power distribution for two counter-propagating waves is derived under the condition of steady-state continuous wave (CW) operation. Based on a one-dimensional (1D) traveling-wave amplifier approach similar to (Marcuse 1983; Matuschek et al. 2008), the differential equations

$$\frac{\mathrm{d}P^+(z)}{\mathrm{d}z} = \left\{g_{\mathrm{mod}} - \alpha_{\mathrm{i}}\right) \cdot P^+(z) + \frac{n_{\mathrm{sp}}\hbar\omega v_{\mathrm{gr}}}{2L} g_{\mathrm{mod}}\,, \tag{19.1}$$

$$-\frac{\mathrm{d}P^-(z)}{\mathrm{d}z} = \left\{g_{\mathrm{mod}} - \alpha_{\mathrm{i}}\right) \cdot P^-(z) + \frac{n_{\mathrm{sp}}\hbar\omega v_{\mathrm{gr}}}{2L} g_{\mathrm{mod}}\,, \tag{19.2}$$

determine the power distribution $P^{\pm}(z)$ for forward- and backward-propagating photons of the same longitudinal mode with angular frequency ω. $g_{\mathrm{mod}} = \Gamma g_{\mathrm{mat}}$ is the modal gain with g_{mat} being the material gain and Γ the mode confinement factor. α_{i} is the internal loss coefficient, n_{sp} the population inversion factor, L the effective chip length of the tilted ridge-waveguide structure, and $v_{\mathrm{gr}} = c/n_{\mathrm{gr}}$ the group velocity with the speed of light in vacuum, c, and the effective group index, n_{gr}. The independent variable (z-coordinate) is defined in the range $z \in [0;\, L]$, where the boundaries are defined by the chip's front and back facet (see Figure 19.3). The second term on the right-hand side describes the generation of light by spontaneously emitted photons, whereas the first term describes the stimulated amplification process by already existing photons.

It should be noted that, generally, there is one set of equations for each longitudinal mode. Here, we have omitted an index for the number of the longitudinal mode. We solve the equations for the mode closest to the gain maximum. Finally, the total power values are obtained by summing over an effective number of longitudinal modes, which is proportional to the full-width at half maximum (FWHM) or 3-dB bandwidth of the emitted ASE power spectrum (Matuschek and Duelk 2013). As explained in detail in Marcuse (1983), *it is not necessary to treat this ASE process as a coherent signal since it distributes itself continuously over a relatively wide band of wavelengths with random phases between adjacent wavelength components*. Hence, a description using powers instead of fields with definite amplitude and phase is sufficient.

Equations 19.1 and 19.2 can be easily solved if we assume that the gain is homogeneous in longitudinal direction, that is, we neglect any gain saturation effects. This approximation is justified for low-power SLEDs but might be questionable for SLEDs and R-SLEDs operating at high injection currents, and thus, at high-output powers, as will be shown at the end of this section. Then the general solution is given by

$$P^{+}(z) = P_0^{+} \cdot G(z) + a \cdot \{G(z) - 1\}\,, \tag{19.3}$$

$$P^{-}(z) = P_L^{-} \cdot G(L - z) + a \cdot \{G(L - z) - 1\}\,, \tag{19.4}$$

with

$$a = \frac{n_{\mathrm{sp}}\hbar\omega v_{\mathrm{gr}}}{2L} \frac{g_{\mathrm{mod}}}{g_{\mathrm{mod}} - \alpha_{\mathrm{i}}}, \tag{19.5}$$

where we have introduced the gain amplification factor

$$G(z) = \exp\left\{(g_{\mathrm{mod}} - \alpha_{\mathrm{i}})z\right\}. \tag{19.6}$$

The subscripts 0 and L denote power values at positions $z = 0$ and $z = L$ inside the chip.

Equations 19.3 and 19.4 are coupled by the boundary conditions

$$P_L^{-} = R_{\mathrm{f}} \cdot P_L^{+}, \tag{19.7}$$

$$P_0^{+} = R_{\mathrm{b}} \cdot P_0^{-}, \tag{19.8}$$

where R_{f} and R_{b} are the effective modal reflectivities at the front and back sides, respectively. Using these boundary conditions, we can derive the expressions

$$P_L^{+} = a \cdot \frac{1 + G(L)R_{\mathrm{b}}}{1 - G^2(L)R_{\mathrm{b}}R_{\mathrm{f}}} \cdot (G(L) - 1)\,, \tag{19.9}$$

$$P_0^{-} = a \cdot \frac{1 + G(L)R_{\mathrm{f}}}{1 - G^2(L)R_{\mathrm{b}}R_{\mathrm{f}}} \cdot (G(L) - 1)\,, \tag{19.10}$$

for the light powers that hit the facets from inside the chip. It should be noted that Equations 19.9 and 19.10 are symmetric with respect to an exchange of the chip's front- and back-facet sides.

The power coupled out of the chip is obtained by multiplying (Equations 19.9 and 19.10) with the effective transmittance of the front- and back-facet coatings, T_{f} and T_{b}.[†] Thus, taking the values of all variables at the spectral position of the gain maximum, which we indicate by adding a subscript index 0 at each relevant variable and summing over the number of effectively contributing modes finally leads to

---

[†] Generally, the inequality $R_{\mathrm{f/b}} + T_{\mathrm{f/b}} < 1$ holds because of slight absorption that might occur in the applied dielectric coatings. Moreover, for tilted SLED sections light that is not reflected back at the facet into the waveguide but elsewhere, is lost and not covered by the reflectivities $R_{\mathrm{f/b}}$.

$$P_f = \tilde{P} \cdot \frac{1 + G_0(L)R_\mathrm{b}}{1 - G_0\,(2L)\,R_\mathrm{b}R_\mathrm{f}} \cdot \left\{G_0(L) - 1\right) \cdot T_\mathrm{f}, \tag{19.11}$$

$$P_\mathrm{b} = \tilde{P} \cdot \frac{1 + G_0(L)R_\mathrm{f}}{1 - G_0\,(2L)\,R_\mathrm{b}R_\mathrm{f}} \cdot \left\{G_0(L) - 1\right) \cdot T_\mathrm{b}. \tag{19.12}$$

The prefactor $\tilde{P}$ is given by

$$\begin{aligned} \tilde{P} = n_\mathrm{eff}a_0 &= n_\mathrm{eff} \cdot \frac{n_\mathrm{sp}\hbar\omega_0 v_{\mathrm{gr},0}}{2L} \frac{g_{\mathrm{mod},0}}{g_{\mathrm{mod},0} - \alpha_{\mathrm{i},0}} \\ &= \Delta\lambda_\mathrm{3dB} \cdot \frac{n_\mathrm{sp}hc^2}{\lambda_0^3} \frac{g_{\mathrm{mod},0}}{g_{\mathrm{mod},0} - \alpha_{\mathrm{i},0}}, \end{aligned} \tag{19.13}$$

where we have estimated the effective mode number, n_eff, by the 3-dB bandwidth of the ASE output spectrum, $\Delta\lambda_\mathrm{3dB}$, and the longitudinal mode spacing, $\lambda_0^2/(2n_{\mathrm{gr},0}L)$. Strictly speaking, the replacement of the sum by an effective mode number requires that the dispersion of the effective modal facet reflectivities and coating transmittance can be neglected over the relevant spectral range.

Equations 19.11 and 19.12 together with Equation 19.13 represent the general solution for the front- and back-facet output power of an SLED with arbitrary front- and back-side reflectivity. They are particularly useful for analyzing differences observed in the output power from both facets caused by residual facet reflections. In Matuschek et al. (2008), very similar expressions were derived and a short discussion of the consequences on the front-to-back output power ratio was given. In the following, we apply our general results to two special cases of an SLED, namely an ideal standard SLED and an ideal R-SLED.

19.2.3.1 Ideal Standard SLED

An SLED device is assumed where both facets do not have any residual reflection. This condition is nearly fulfilled for real devices with sufficiently great tilt angle and AR-coated facets. In this limit, we can set $R_\mathrm{f} = R_\mathrm{b} = 0$ and $T_\mathrm{f} = T_\mathrm{b} = 1$ and obtain

$$P_f = P_\mathrm{b} = \tilde{P} \cdot \left\{G_0(L) - 1\right) = \Delta\lambda_\mathrm{FWHM} \cdot \frac{n_\mathrm{sp}hc^2}{\lambda_0^3} \frac{g_{\mathrm{mod}\,,0}}{g_{\mathrm{mod}\,,0} - \alpha_{\mathrm{i},0}} \cdot \left(e^{(g_{\mathrm{mod}\,,0} - \alpha_{\mathrm{i},0})\cdot L} - 1\right). \tag{19.14}$$

Inserting a logarithmic gain model in Equation 19.14 yields a power law for the SLED's L–I characteristic with chip length–dependent exponent. The consequences of this power law have been discussed and experimentally proven in detail in Matuschek and Duelk (2013) and are not repeated here.

19.2.3.2 Ideal R-SLED

An SLED device is assumed with a perfect front facet having zero residual reflection, similar to an ideal standard SLED, but with a back facet being a perfect reflector that reflects all incident light back into the waveguide. For real devices, this condition is fairly well realized with effective modal reflectivity values of about 90% or even higher for the HR coating. Moreover, the waveguide at the back facet must be straight or perpendicular with respect to the facet in order to avoid any geometrical reduction of the effective modal reflectivity. In the limit of an ideal R-SLED we can set $R_\mathrm{f} = T_\mathrm{b} = 0$ and $R_\mathrm{b} = T_\mathrm{f} = 1$ and obtain

$$P_f = \tilde{P} \cdot \left\{G_0\,(2L) - 1\right) = \Delta\lambda_\mathrm{FWHM} \cdot \frac{n_\mathrm{sp}hc^2}{\lambda_0^3} \frac{g_{\mathrm{mod}\,,0}}{g_{\mathrm{mod}\,,0} - \alpha_{\mathrm{i},0}} \cdot \left(e^{2\cdot(g_{\mathrm{mod}\,,0} - \alpha_{\mathrm{i},0})\cdot L} - 1\right) \tag{19.15}$$

for the total power out of the front facet. It should be obvious that under those assumptions no power will exit the back facet of the chip, that is, $P_b = 0$.

Comparing Equation 19.15 with Equation 19.14, it follows directly that with respect to output power from the front facet an R-SLED of length L behaves like a standard SLED of twice the length $2L$. This result seems intuitive since the amplification process for photons is extended over a distance, which corresponds to one round-trip in the chip with perfect reflection at the back facet. However, it is important to realize that with respect to current injection the chip is still a short chip of length L. This means that an R-SLED of length L emits the same power out of the front facet as a standard SLED of length $2L$ at same current density values J or, in other words, at half values of the injection current I. Also, under these conditions the average carrier densities are approximately the same and, thus, the emitted ASE output spectra with their corresponding FWHM bandwidth. Consequently, the prefactor of Equation 19.13 has the same value in both cases.

19.2.4 Experimental and Numerical Verification of the Analytical Model

In this section, we want to verify the results derived from the analytical treatment by comparing them with experimental results as well as results obtained from a full 3D simulation of the SLED and R-SLED under consideration. The simulation software is based on a finite-element discretization of the device. Using the $\mathbf{k}\cdot\mathbf{p}$ method, the quantum mechanical system is described by a 6×6 Luttinger–Kohn Hamiltonian and the eigenvalue problem is solved as well as a Helmholtz equation for the optical modes (Chuang 1995). These equations are coupled to the semiconductor equations and thermal equations as described earlier. All equations are then solved self-consistently on a discrete mesh.

As an object for our study, we use SLED and R-SLED chips from the same epi wafer with light emission in the wavelength region around 860 nm. All chips are operated at a heat-sink temperature of 25°C. The MQW epi structure of these chips is based on (Al)GaAs/(In)GaAs layers grown on a GaAs substrate (see Table 19.1). For the simulation, the effective modal reflectivity values have not been set to the ideal values 0 and 1 but to values of 10^{-9} for the AR-coated facets with tilted waveguide section and 0.95 for the back facet with straight waveguide section. These values are quite realistic for well-designed SLED and R-SLED structures. The simulator has been calibrated by adjusting the L–I characteristics and ASE output spectra for short standard SLEDs with chip lengths ranging from 500 to 850 μm.

Figure 19.4 shows the results obtained for the 850-μm-long SLED chip after the calibration procedure. Obviously, the measured L–I curve is almost exactly reproduced by the simulation. Moreover, the detailed shape of the measured ASE spectra is also simulated very accurately for different levels of injection currents. The quality of the analytical model is shown by the dotted L–I characteristic, which is obtained from Equation 19.14 by following the procedure described in Matuschek and Duelk (2013). The center wavelength and 3-dB bandwidth have been taken from the ASE spectrum at an injection current of 100 mA. The modal gain as a function of the current density has been derived from the full 3D simulation by averaging over the z-direction. Apparently, the output power at high-injection currents is somewhat overestimated by the analytical model because the relatively simple model does not take into account self-heating and gain saturation effects of the chip.

After successful calibration, we are able to compare the electro-optical properties of a 1750-μm standard SLED with an 840-μm R-SLED. The chip length ratio is 2.08, which is close to the ideal value of two. Moreover, the back-facet coating of the R-SLED chip is not a real HR coating but only a PR coating with a reflectance value of about 20%. Figure 19.5 shows the measured and simulated front-facet output power as a function of the current density for both types of SLEDs. For our considerations the effective width of current injection is not relevant. Therefore, we scale the current values only by the chip length so that the density values are defined by $J = I/L$. As can be seen, the agreement between all curves is excellent. Only the measured power of the R-SLED (dash-dotted line) is slightly smaller because of the nonideal reflectance of the PR coating. First of all, this proves that an R-SLED behaves like a standard SLED of twice the length

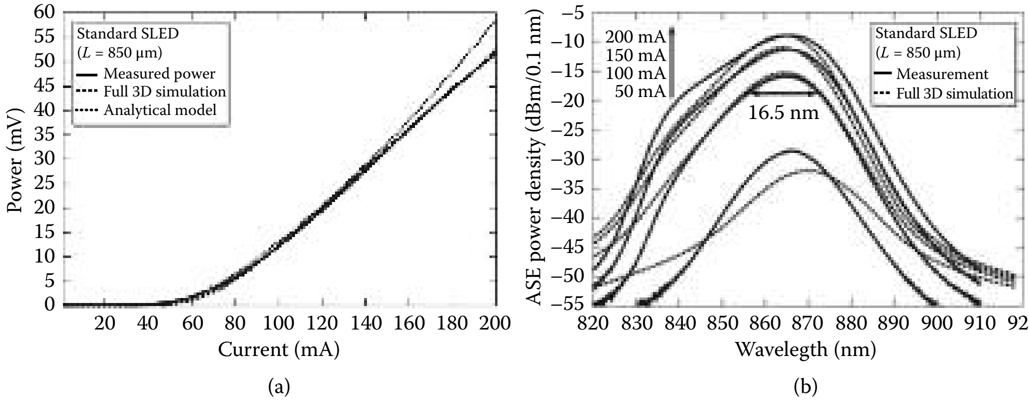


FIGURE 19.4 (a) L–I characteristics and (b) ASE output spectra (on an absolute logarithmic scale) for an 850-μm-long standard SLED chip. Solid lines represent measurement data and dashed lines simulation results obtained from a full 3D simulation after the calibration procedure. The dotted L–I characteristic follows from the analytical model according to Equation 19.14, as explained in the text. The ASE spectra are shown for four different injection currents: 50, 100, 150, and 200 mA. At 100 mA the 3-dB spectral bandwidth is about 16.5 nm. Note that all simulated ASE spectra have been shifted slightly by the same amount in wavelength so that they coincide at their maximum values with the measured spectra. This allows a better comparison of the detailed behavior of the spectral shapes.

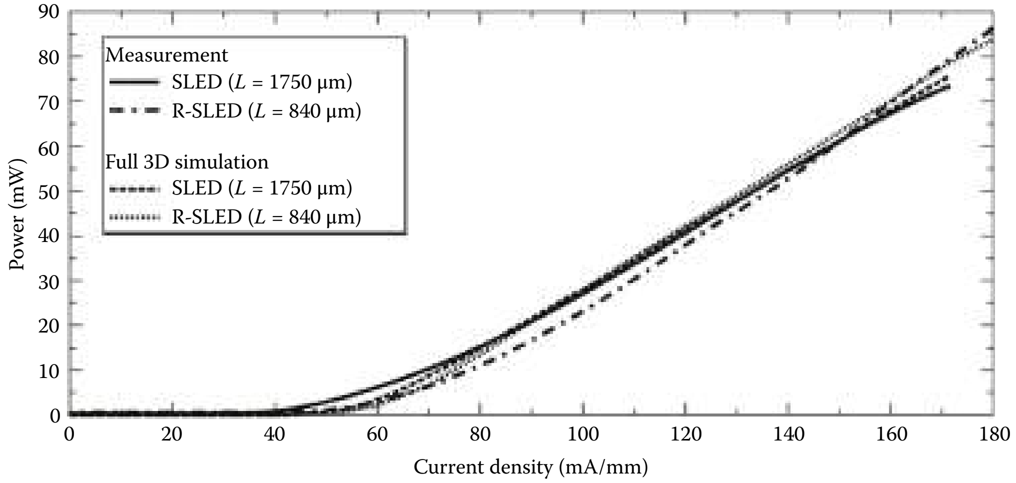
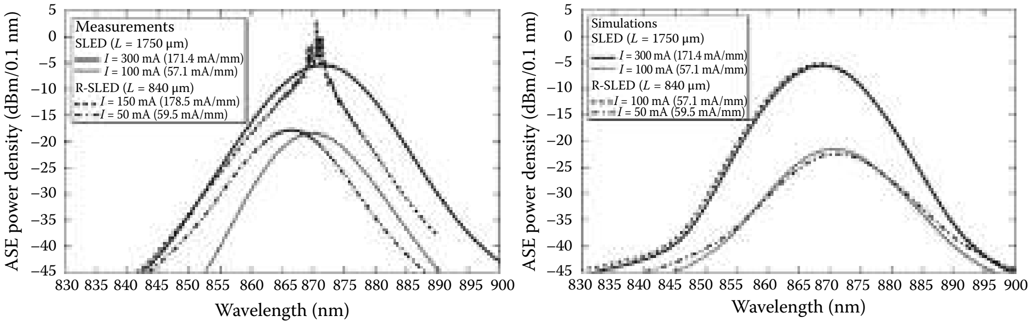


FIGURE 19.5 Measured (solid and dash–dotted line) and simulated (dashed and dotted line) front-facet output power of an SLED chip with length 1750 μm and an R-SLED chip with length 840 μm as function of current density, as described in the text. All curves agree almost perfectly.

at the same current density. Second, experimental results can be accurately reproduced and extrapolated once the simulator is well calibrated.

This is further demonstrated by analyzing the ASE output spectra. Figure 19.6a shows measured ASE spectra and Figure 19.6b shows simulated ASE spectra for both chips at two different levels of current density. In the first case, the current density with values slightly below 60 mA/mm is just above the threshold for ASE operation, which is why the output power is low. In the second case, the current density is beyond 170 mA/mm, resulting in the chips operating at high front-facet output powers of about 75 mW. The simulated curves coincide more or less over the entire spectral range for both current densities, whereas a



(a)

(b)

FIGURE 19.6 (a) Measured and (b) simulated ASE output spectra of an SLED chip with length 1750 μm and an R-SLED chip with length 840 μm at two different levels for the current density. The dotted and dash-dotted lines correspond to a current density slightly below 60 mA/mm and the solid and dashed lines to a value slightly above 170 mA/mm. In contrast to Figure 19.4b, no additional wavelength shift has been applied to the ASE spectra.

small wavelength shift can be observed for the measured curves. This is explained by the fact that even nominally identical chips from the same epi wafer have slightly different center wavelengths of the emitted ASE spectra. Furthermore, the thermal heat sinking toward the submount might not be identical for all chips. A more important difference, which can be observed for the measurements at the high current densities, is that the ASE spectrum of the R-SLED is somewhat narrower and shows strong periodic modulations (spectral ripples) in the center part. The periodicity corresponds to the longitudinal mode spacing defined by the length of the linear cavity. This is a clear indication of a resonant cavity effect (Fabry–Pérot modulation) caused by the high back-side reflectivity and demonstrates the tendency of an R-SLED to start lasing at high levels of current injection. In Section 19.3.2, the influence of spectral ripples on the coherence function will be discussed.

To gain further insight into the physical properties of an SLED and R-SLED, respectively, we analyze the longitudinal distribution of the optical intensity and the carrier density along the z-axis inside the chip based on the results obtained from the full 3D simulation (see Figure 19.7). The intensity profiles have been extracted at the center of the optical mode, and the maximum value at the front facet has been normalized to unity. The carrier density profiles have been extracted at the center of one of the QWs. The z-axis has been defined so that the back facet of the R-SLED coincides with the middle of the standard SLED and the front facets almost coincide, too. With this definition, all chip properties of the standard SLED are symmetric with respect to the origin. As shown in the upper plot, the shapes of the optical intensity profiles are quite similar for the same current density. However, as can be clearly seen, for low current densities (dotted and dash–dotted lines) the growth of the optical intensity in direction to the front facet is almost exponential, whereas the optical intensity saturates for higher current densities (solid and dashed lines) and, hence, higher optical output powers. This gain saturation behavior due to high optical powers is directly linked to the longitudinal carrier density distribution as shown in the lower plot. At low current densities, the carrier density is almost constant along the chip with a value of about 3.3×10^{-18} cm^{-3}. In contrast, at high current densities the carrier density decreases strongly by more than 40% from the origin toward the front facet with a terminal value of about 2.8×10^{-18} cm^{-3}. Low values of the carrier density mean low material gain, and thus, low gain amplification of the optical power.

Our investigations can be summarized as follows:

- At same current densities an R-SLED behaves very similar like a standard SLED with twice the length.
- An important difference is that R-SLEDs show ripples in the ASE spectrum and may have the tendency to start lasing operation.
- For long SLED chips and for R-SLEDs at high-output powers, gain saturation effects are not negligible anymore.

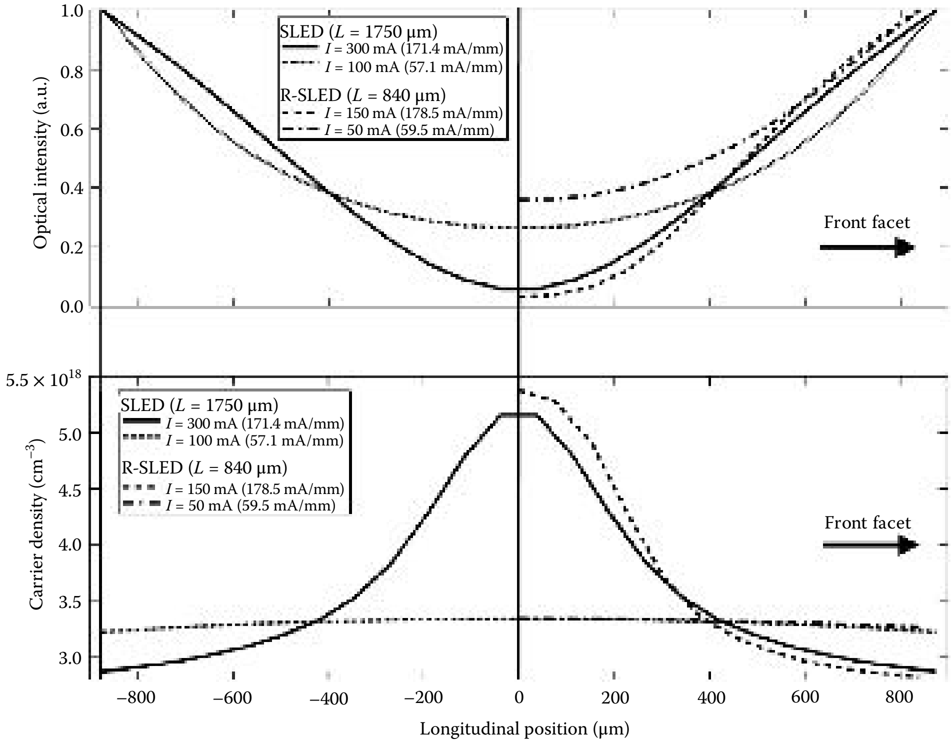


FIGURE 19.7 Longitudinal distribution of the optical intensity (upper plot) and carrier density (lower plot) along the z-axis inside the chip of an SLED chip with length 1750 μm and an R-SLED chip with length 840 μm at two different levels for the current density. The dotted and dash-dotted lines correspond to a current density slightly below 60 mA/mm and the solid and dashed lines to a value slightly above 170 mA/mm.

19.3 SLED Designs with Specified Targets

Commercial SLEDs have to fulfill target specifications for the electro-optical performance (output power, wavelength, bandwidth, spectral shape, ripple, PER, etc.) that are defined by the application. Also, their long-term reliability has to meet certain lifetime requirements (e.g., 10% power drop over 5000 hours of operation or 50% power drop over 100,000 hours of operation) that are dependent on the application and the intended use of operation. The shape and width of the ASE spectrum is directly linked to the coherence function, which is the autocorrelation function of an SLED. This is of particular importance for applications like OCT, as explained below. It is well known that increasing the output power is, in general, accompanied by a reduction in the ASE spectral bandwidth. Hence, methods are required to overcome this limitation.

19.3.1 Broad-Bandwidth SLED Designs

Generation of light in an SLED is based on the recombination from an electronic state in the conduction band with a hole state in the valence band (VB). In a QW, the bound states build a discrete set of subbands. In the simplest case, the SLED's epitaxial structure is based on an SQW active-region design or an MQW structure consisting of uncoupled identical QWs. Such designs can be reduced to the discussion of the quantized SQW states. Here, we restrict our discussion to the VB for heavy holes because ASE spectra of SLEDs consisting of compressively strained QWs are dominated by transitions to the heavy-hole sub-bands.

19.3.1.1 SLEDs with Bell-Shaped (Gaussian) ASE Spectra

The spectral shape of the ASE spectra is mainly determined by the fundamental sub-band transition from the first electronic state to the first heavy-hole state $\{e_1 \rightarrow hh_1)$, as long as the carrier density is sufficiently low, so that the population of other sub-band states can be neglected (Chen et al. 1990). Generally, single transition designs yield bell-shaped (Gaussian) ASE spectra as shown in Figure 19.1a, b, and f. Such spectra are preferred with respect to the avoidance of side lobes in the near optical path length difference (OPD) range of the coherence function (see Section 19.3.2). An increasing carrier density in the QW leads at first to a rising population of high-energy states in the fundamental sub-bands, and thus, to a broadening of the spectrum. If other sub-band states start to become significantly populated, an additional hump tends to appear on the short-wavelength side (as shown in Figure 19.1c). Hence, with increasing population of the sub-bands, the spectral appearance changes continuously from being bell shaped in direction to an M-shaped double-humped spectrum (see Figure 19.1d).

19.3.1.2 SLEDs with M-Shaped (Double-Humped) ASE Spectra

Generally, all approaches for the design of SLED structures showing an ASE spectrum with ultrawide emission bandwidth are based on a multitransition design. This means that more than just one optical transition contributes to the emission spectrum. One has to distinguish between two principally different approaches. The first one is based on the SQW (or MQW) epitaxial design, as just discussed, in which, in addition to the fundamental transition, other transitions fulfilling the quantum-mechanical selection rules are utilized to contribute to the generation of light. If properly designed, such QW structures allow for the enormous enlargement of the spectral bandwidth (Semenov et al. 1993b; Kondo et al. 1992). The second approach is based on the use of nonidentical (so-called chirped) QWs (Lin et al. 2004). Chirping means that at least one QW of an MQW epitaxial layer structure is differently designed compared to the other QW(s), yielding optical transitions at different wavelengths. Chirping can be achieved by a different QW thickness and/or different material compositions for the QW and/or the barrier layers. The basic idea of utilizing chirped QWs is demonstrated in Figure 19.8 for a chirped double-quantum well (DQW) structure. Due to the greater band-gap energy and thinner QW thickness of the right QW, the fundamental transition results in photons with higher energy (shorter wavelength) compared to the left QW. Thus, an ultrawide bandwidth design may be achieved if the spectral separation chosen between both fundamental transitions is sufficiently large.

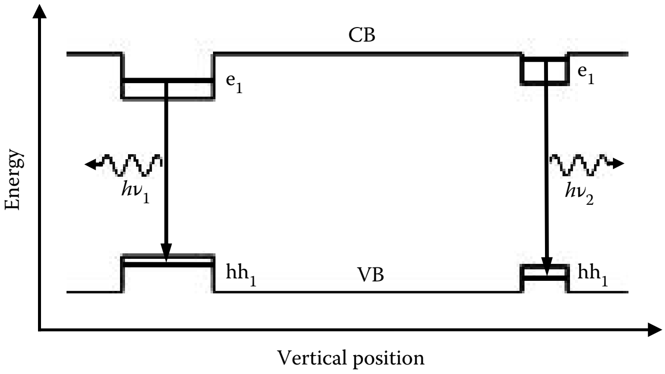


FIGURE 19.8 Energy-band diagram for two uncoupled chirped quantum wells (QWs). The horizontal lines in the wells indicate the first bound electronic state in the conduction band (CB), e_1, and the first bound hole state in the valence band (VB) for each QW separately. For the VB we restrict ourselves to the heavy hole band with its states hh_1. The arrows indicate the fundamental transition from the CB to the VB for both QWs. The electron–hole recombination results in the emission of a long-wavelength photon with energy $h\nu_1$ from the left QW and a short-wavelength photon with energy $h\nu_2$ from the right QW.

As mentioned earlier, ultrawide ASE spectra tend to exhibit a double-humped spectral shape. Such spectra show their maximum spectral bandwidth when both humps are well balanced, i.e., for approximately the same peak strength (the so-called flat-top condition). For a given chip design and a given heat-sink temperature, this situation can be achieved only over a limited range of operation currents. This makes such designs less flexible compared to single-transition designs. Moreover, the non-Gaussian spectral shape is unfavorable with respect to the occurrence of side lobes in the coherence function as mentioned earlier. The stronger the deviation from a Gaussian shape, the higher the side lobes are in the coherence function.

Generally, the width of the ASE spectrum broadens with the energy separation between allowed transitions to the heavy-hole VB. As a consequence of a greater spectral separation, first of all, the dip between both humps obvious in the ASE spectra at flat-top condition increases yielding a negative impact on the coherence function. And second, the SLED has to be driven at higher currents for operation at flat-top condition. The reason is that higher carrier densities in the QWs are required in order to lift the energetically more distant transitions on the short-wavelength side. This might be used as a positive side effect or if it is undesired, this effect can be compensated for by using chips with shorter active segment length. Nevertheless, it is obvious that the approach described here is limited to finding the best compromise between spectral bandwidth and the depth of the spectral dip.

19.3.1.3 Flattening of M-Shaped ASE Spectra

The M-shape approach can be optimized by a flattening of the spectral dip (see Figure 19.1e). This requires the use of at least one additional transition between QW states yielding photons with a wavelength somewhere around the spectral position of the dip. The additional transition fills up the dip between both peaks in the ASE spectrum. It can be realized in different ways, as discussed in the following.

The first possibility is the use of a chirped MQW active region design with at least one of the chirped QWs having its fundamental transition in the spectral range of the dip (Lin et al. 1996). For MQW structures consisting of a large number of QWs an elaborated approach is based on the continuous chirp of the fundamental transition wavelength for each QW over a wide spectral range.

Another possibility is the utilization of coupled QWs, where the coupling is caused by the finite potential walls between the QWs. For a coupled DQW active region design, the coupling leads to a splitting of the fundamental sub-band states, where the coupling strength, and thus, the splitting can be adjusted by the thickness of the barrier layers between the QWs and/or the well depth. The latter can be adjusted by the material composition of the barrier layers. If properly designed, the transition wavelength of the short-wavelength photon appears in the spectral range of the dip.

A third possibility to reduce or even remove the dip is the combination of two or more SLEDs with spectrally shifted ASE spectra. The SLEDs are designed so that the spectral peak of one SLED coincides with the dip of the other SLED. The combined spectrum is flatter and potentially broader if the SLEDs are appropriately designed. The flattening effect is demonstrated by the example of the M-shaped ASE spectrum for the 1050-nm SLED shown in Figure 19.1d. We assume that a second SLED has been designed with similar spectrum just shifted by 24 nm to shorter wavelengths. For simplicity reasons we take the original spectrum (the dashed curve in Figure 19.9) and shift it accordingly in order to obtain the dotted curve. Now, we assume that the output spectra are combined via a dispersion-free 50:50 coupler. This corresponds to a simple linear addition of both spectra with the same weight, resulting in the solid curve shown in Figure 19.9. Obviously, the original dip at around 1050 nm is removed and the resulting spectrum looks rather bell shaped. In the following section, the impact of the SLED combination on the coherence function is discussed.

19.3.2 Coherence Function and Coherence Length

As mentioned earlier, for OCT applications, the coherence function is of major interest. It allows the extraction of the coherence length, which determines the theoretical limit for the axial resolution of an

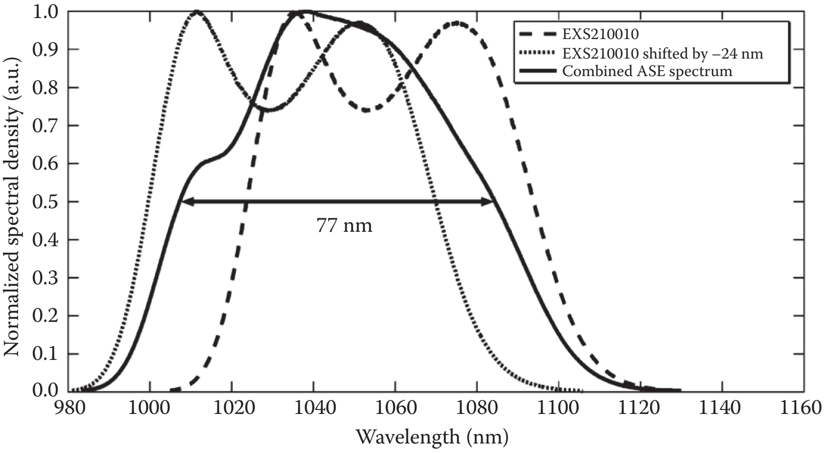


FIGURE 19.9 The dashed line shows the same ASE spectrum as plotted in Figure 19.1d for module EXS210010. The dotted curve shows the same spectrum shifted by 24 nm to shorter wavelengths. The solid curve is obtained as the linear superposition of both spectra. The FWHM bandwidth of the combined spectrum is 7 nm greater compared to the width of each single spectrum.

OCT imaging system (Duelk and Hsu 2015). For a broadband light source, the coherence length can be written as

$$l_{\text{coh}} = \gamma \cdot \frac{2 \cdot \ln(2)}{\pi} \cdot \frac{\lambda_{3\text{dB}}^2}{\Delta\lambda_{3\text{dB}}}, \tag{19.16}$$

with $\lambda_{3\text{dB}}$ being the 3-dB center wavelength and $\Delta\lambda_{3\text{dB}}$ the 3-dB bandwidth of the ASE spectrum. The factor γ depends on the shape of the ASE spectrum. For an ideal Gaussian shape, the factor is unity, $\gamma = 1$, for other shapes, $\gamma > 1$. For SLEDs with double-humped ASE spectra operating at flat-top condition, a value of $\gamma \approx 1.19$ is typically used. This yields a penalty of roughly 20% in coherence length caused by the deviation from an ideal Gaussian shape. As an example, for the ASE spectrum shown in Figure 19.1e, we calculate for the coherence length $l_{\text{coh}} \approx 7.4$ μm using Equation 19.16 with $\lambda_{3\text{dB}} = 1300$ nm, $\Delta\lambda_{3\text{dB}} = 120$ nm, and assuming a flat-top shape with $\gamma = 1.19$.

The coherence length can be directly determined from the coherence function, too. The latter is equivalent to the SLED's autocorrelation function, and thus, according to the Wiener–Khinchin theorem, the Fourier transform of the ASE spectrum (Schmitt 1999). Figure 19.10 shows the coherence function as a function of the OPD obtained for the ASE spectrum just discussed. The coherence function is symmetric with respect to the OPD. Its drop-off from the maximum value at zero OPD to 50% defines the coherence length, that is, the coherence length is the half-width at half maximum (HWHM). From the coherence function shown in Figure 19.10a near zero, we extract a coherence length of 7.9 μm. This value is close to the value estimated from Equation 19.16. Moreover, side lobes are visible at an OPD of 20 μm with a side-lobe suppression ratio (SLSR) of about 10 dB. The side lobes are caused by the non-Gaussian flat-top shape of the ASE spectrum. For applications like OCT, those side lobes may generate imaging artifacts if they are not suppressed by an additional windowing function in the OCT signal processing (Duelk and Hsu 2015). Therefore, SLED designs with bell-shaped ASE spectra are preferred for such applications. Figure 19.11 shows the coherence functions, which correspond to the ASE spectra plotted in Figure 19.9 for module EXS210010 and the combined SLED spectrum. Obviously, going from the M-shaped to the rather bell-shaped ASE spectrum leads to a strong reduction of the side lobes with an increase of the SLSR from 7 to 13 dB. As a side effect, the coherence length decreases by almost 1 μm from 8.6 to 7.7 μm due to the broadening of the combined SLED spectrum by 7 nm.

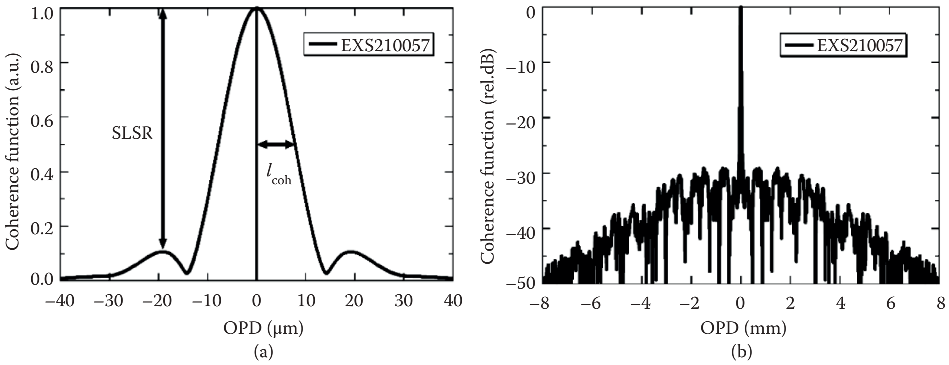


FIGURE 19.10 Coherence function versus optical path length difference (OPD) obtained for the ASE spectrum shown in Figure 19.1e for SLED EXS210057. (a) Linear plot over a small OPD range. The coherence length defined by the half-width at half maximum (HWHM) value is found to be 7.9 μm (in air) and the side-lobe suppression ratio (SLSR) roughly 10 dB. (b) Logarithmic plot over the full OPD range.

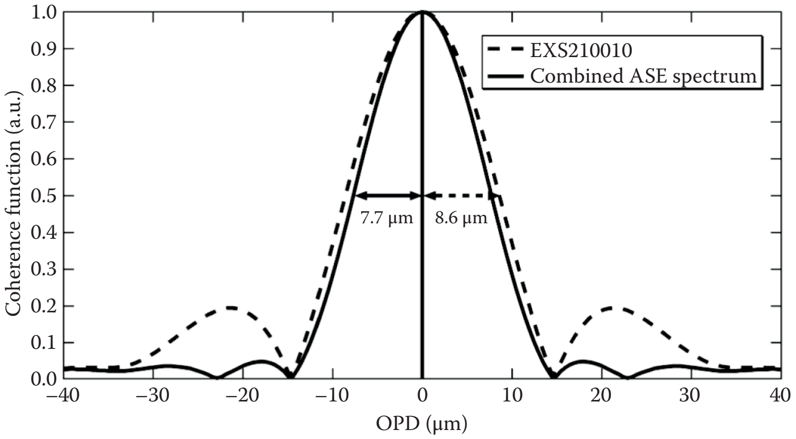


FIGURE 19.11 Coherence function in the near OPD range obtained for the ASE spectra plotted in Figure 19.9 for module EXS210010 (dashed curve) and the combined SLED spectrum (solid curve). From the dashed curve an SLSR of 7 dB and a coherence length of 8.6 μm is extracted, whereas the values found from the solid curve are 13 dB and 7.7 μm, respectively.

Periodic modulations in the ASE spectrum caused by the build-up of a resonant cavity, as discussed in Section 19.2.4, may show up as secondary coherence subpeaks at an OPD of a few millimeters (see, e.g., Figure 18.1 in Duelk and Hsu 2015). Figure 19.10b is free of such secondary coherence peaks as the ASE spectrum is smooth and free of spectral ripple. In order to avoid OCT imaging artifacts (so-called "ghost lines"), a secondary peak suppression ratio (SPSR) of at least 25 dB is needed. The SPSR requirement defines the tolerance for the maximum allowed spectral ripples in the ASE spectrum. As a consequence, this might cause R-SLED designs to be impractical for OCT applications.

19.3.3 High-Power SLED Designs

In addition to the requirements for the ASE output spectrum, SLEDs should typically deliver a minimum amount of output power at a given injection current. A variety of methods exist for increasing the SLED's front-facet output power, and thus achieve a high-power design. In the following, we describe possible approaches and discuss their advantages and drawbacks.

19.3.3.1 Increasing the Injection Current

The most obvious approach is just to increase the injection current. This is possible as long as the point for the thermal roll-over is not reached. But one should take into account that the increasing carrier density may have a strong influence on the shape of the ASE output spectrum. As discussed in Section 19.3.1, ASE spectra resulting from a single optical transition will typically remain bell shaped with broader bandwidth. However, for M-shaped ASE spectra, which may be well balanced for a lower injection current, as shown in Figure 19.1d and e, the low-wavelength peak will strongly grow with increasing injection current. This leads to an unfavorably strong reduction of the spectral bandwidth compared to the preferred flat-top condition.

19.3.3.2 Increasing the Chip Length

The output power can be increased for a given injection current by increasing the active segment length, that is, the length of the gain medium. This method works for operating currents sufficiently above the ASE threshold level because the latter increases with increasing chip length. Longer chips lead to a reduction of both the thermal resistance and the series resistance leading to lower junction temperatures for SLED chips running in CW mode of operation. Moreover, the carrier density is reduced resulting in the opposite effect on the ASE spectra, as explained in Section 19.3.3.1.

19.3.3.3 Using an R-SLED Design

As discussed in Section 19.2.3, a very efficient way to design an ultimate high-power SLED structure is the R-SLED approach. This allows for very high-output powers using relatively short chips. The main drawback of this approach is the enhancement of spectral ripples or even worse the tendency to start lasing due to the build-up of a resonant cavity. This may lead to an increase of the noise level in general and the appearance of strong secondary coherence subpeaks. Hence, the R-SLED design might be the optimum choice for insensitive applications with respect to resonant-cavity effects. For other applications, this approach might not work or is not recommended.

19.3.3.4 High-Power SLEDs by Epitaxial Design

The epitaxial design of the active-region layers has a strong influence on the SLED's *L*–*I* characteristic. Typically, it is easier to design SLEDs delivering higher-output powers by increasing the number of QWs and/or by increasing the thickness of the QW layers. Modifying the epitaxial design in this way has diverse and complicated effects on the expected electro-optical performance. For example, the injection current required for reaching ASE threshold increases in general with increasing total thickness of the active region. Moreover, the carrier density is reduced, having a similar effect on the ASE spectrum as described earlier in Section 19.3.3.2 for an extended chip length.

Therefore, the approach that has to be chosen for the realization of a specific SLED design depends on the whole set of target specification. The expected *L*–*I* characteristics and ASE spectra for a given active region design finally determine the appropriate chip length and injection current for device operation. There are no simple and strictly valid rules for finding the optimal SLED design. However, it can be approached using the discussion presented in this section as a guideline.

19.4 Summary

SLEDs are employed over a wide range of different applications with emitted ASE spectra covering the wavelength range from the visible to the near infrared. The band-gap energy of available semiconductor materials determines the appropriate choice of the substrate and epitaxial layer materials for the design of a specific SLED structure. Most SLEDs are realized in a tilted ridge-waveguide geometry similar to a Fabry–Pérot LD, which allows for the formation of optical modes. The target specifications mainly on the output power and the ASE spectrum determine the active-region design. SQW and MQW structures are most commonly used to achieve material gain.

The performance of SLED structures can be simulated like other opto-electronic devices using full 3D simulation tools. Such complex tools couple the different problems involved in the physical device description (multiphysics treatment). The simulation results are rather accurate at the expense of analytical insight. Analytical approaches based on traveling-wave equations describing the longitudinal propagation of light reduce the dimensionality of the problem and the number of coupled equations. They are particularly useful to derive an analytical expression for the SLED's *L–I* characteristic. We have presented a simple model that includes the influence of arbitrarily strong modal reflectivities at both facets on the output power. Using our analytical results, we have compared the performance of standard SLEDs with the new class of so-called R-SLEDs. Furthermore, our model has been verified by a comparison against experimental data and results obtained from full 3D simulations.

The design of SLEDs with specific target requirements has been discussed. In particular, the appearance of single-peaked (bell-shaped) and double-humped (M-shaped) ASE spectra for broad bandwidth SLED structures and how it affects the corresponding coherence functions has been analyzed in detail. Various methods for boosting the output power have been illustrated with their implications particularly on the ASE spectra. These findings can be finally used as a guideline for the design of a specific broad-bandwidth high-power SLED structure.

References

Alphonse, G. A., D. B. Gilbert, M. G. Harvey, et al. 1988. High-power superluminescent diodes. *IEEE J. Quant. Electron*. 24, no. 12: 2454–2457.

Bohnert, K., P. Gabus, J. Nehring, et al. 2002. Temperature and vibration insensitive fiber-optic current sensor. *IEEE J. Lightwave Technol.* 20, no. 2: 267–275.

Burns, W. K., C. Chen, and R. P. Moeller. 1983. Fiber-optic gyroscopes with broad-band sources. *IEEE J. Lightwave Technol.* 1, no. 1: 98–105.

Chen, T. R., L. Eng, Y. H. Zhuang et al. 1990. Quantum well superluminescent diode with very wide emission spectrum. *Appl. Phys. Lett.* 56, no. 14: 1345–1346.

Chuang, S. L. 1995. *Physics of Optoelectronic Devices*. New York: John Wiley & Sons.

Coldren, L. A., S. W. Corzine, and M. L. Masanovic. 2012. *Diode Lasers and Photonic Integrated Circuits*, 2nd ed. Hoboken, NJ: John Wiley & Sons.

CrosslightSoftware2017

Crosslight Software, Inc. Crosslight Software, 2017. http://www.crosslight.com/

Drexler, W. and J. G. Fujimoto, eds. 2008. *Optical Coherence Tomography: Technology and Applications*, 2nd ed. Berlin: Springer-Verlag.

Duelk, M. and K. Hsu. 2015. SLEDs and swept source laser technology for OCT. In *Optical Coherence Tomography: Technology and Applications*, 2nd ed., Drexler, W. and J. G. Fujimoto (eds.), 527–562. Berlin: Springer-Verlag.

Dufour, M. L., G. Lamouche, V. Detalle, et al. 2005. Low coherence interferometry—An advanced technique for optical metrology in industry. *Insight–Non-Destructive Testing and Condition Monitoring* 47, no. 4: 216–219.

Exalos, A. G. EXALOS: High-performance broadband SLEDs, swept sources and sub-systems, 2017. http://www.exalos.com/sled-modules/

Feltin, E., A. Castiglia, G. Cosendey, et al. 2009. Broadband blue superluminescent light-emitting diodes based on GaN. *Appl. Phys. Lett.* 95: 081107.

Heo, D., I.-K. Yun, J.-S. Lee, et al. 2011. High-power SLD-based BLS module for WDM-PON applications. *J. Korean Phys. Soc.* 58, no. 3: 429–433.

Kaminow, I. P. and D. Marcuse. 1983. Superluminescent LED with efficient coupling to optical waveguide. U.S. Patent 4 376 946.

Kondo, S., H. Yasaka, Y. Noguchi, et al. 1992. Very wide spectrum multiquantum well superluminiscent diode at 1.5 µm. *Electron. Lett.* 28, no. 2: 132–134.

Koyama, F., K. Y. Liou, A. G. Dentai, et al. 1993. Multiple-quantum-well GaInAs/GaInAsP tapered broad-area amplifiers with monolithically integrated waveguide lens for high-power applications. *IEEE J. Photonics Technol. Lett.* 5, no. 8: 916–919.

Kwong, N. S. K., K.-Y. Lau, and N. Bar-Chaim. 2008. High-power, high-efficiency GaAlas superluminescent diodes with integrated absorber for lasing suppression. *IEEE J. Quant. Electron.* 25, no. 4: 696–704.

Laino, V., F. Roemer, B. Witzigmann, et al. 2007. Substrate modes of (Al,In)GaN semiconductor laser diodes on SiC and GaN substrates. *IEEE J. Quant. Electron.* 43, no. 1: 16–24.

Lee, T. P., C. A. Burrus, and B. I. Miller. 1973. A stripe-geometry double-heterostructure amplified-spontaneous-emission (superluminescent) diode. *IEEE J. Quant. Electron.* 9, no. 8: 820–828.

Li, Z. Q. and Z. M. S. Li. 2010. Comprehensive modeling of superluminescent light-emitting diodes. *IEEE J. Quant. Electron.* 46, no. 4: 454–461.

Lin, C.-F., Y.-S. Su, C.-H. Wu, et al. 1996. Broad-band superluminescent diodes fabricated on a substrate with asymmetric dual quantum wells. *IEEE J. Photonics Technol. Lett.* 8, no. 11: 1456–1458.

Lin, C.-F., Y.-S. Su, C.-H. Wu, et al. 2004. Influence of separate confinement heterostructure on emission bandwidth of InGaAsP superluminescent diodes/semiconductor optical amplifiers with nonidentical multiple quantum wells. *IEEE J. Photonics Technol. Lett.* 16, no. 6: 1441–1443.

Loeser, M. and B. Witzigmann. 2008. Multi-dimensional electro-opto-thermal modeling of broadband optical devices. *IEEE J. Quant. Electron.* 44, no. 6: 505–514.

Macleod, H. A. 2010. *Thin-Film Optical Filters*. Boca Raton, FL: Taylor & Francis.

Marcuse, D. 1983. Computer model of an injection laser amplifier. *IEEE J. Quant. Electron.* 19, no. 1: 63–73.

Marcuse, D. 1989. Reflection loss of laser mode from tilted end mirror. *J. Lightwave Technol.* 7, no. 2: 336–339.

Matuschek, N. and M. Duelk. 2013. Modeling and simulation of superluminescent light emitting diodes (SLEDs). *IEEE J. Sel. Top. Quant. Electron.* 19, no. 5: 7800307.

Matuschek, N., T. Pliska, and N. Lichtenstein. 2008. Properties of pump-laser modules exposed to polarization-dependent and wavelength-selective feedback from fiber Bragg gratings. *IEEE J. Quant. Electron.* 44, no. 3: 262–274.

Matuschek, N. and M. Duelk. 2014. Modeling and simulation of reflecting SLEDs. *14th International Conference on Numerical Simulation of Optoelectronic Devices* (NUSOD '14), Postdeadline Poster MPD43, Mallorca, Spain.

Middlemast, I., J. Sarma, and S. Yunus. 1997. High power tapered superluminescent diodes using novel etched deflectors. *Electron. Lett.* 33, no. 10: 903–904.

Milani, N. M., V. Mohadesi, and A. Asgari. 2015. A novel theoretical model for broadband blue InGaN/GaN superluminescent light emitting diodes. *J. Appl. Phys.* 117: 054502.

Nagai H., Y. Noguchi, and S. Sudo. 1989. High-power, high-efficiency 1.3 µm superluminescent diode with buried bent absorbing guide structure. *Appl. Phys. Lett.* 54, no. 18: 1719–1721.

Park, J. and X. Li. 2006. Theoretical and numerical analysis of superluminescent diodes. *IEEE J. Lightwave Technol.* 24, no. 6: 2473–2480.

Patterson, B. D., J. E. Epler, B. Graf, et al. 1994. A superluminescent diode at 1.3 µm with very low spectral modulation. *IEEE J. Quant. Electron.* 30, no. 3: 703–712.

Photon Design. Photon Design - Your source of photonics CAD tools, 2017. https://www.photond.com/

Rossetti, M., J. Dorsaz, R. Rezzonico, et al. 2010. High power blue-violet superluminescent light emitting diodes with InGaN quantum wells. *Appl. Phys. Express.* 3: 061002.

Rossetti, M., L. H. Li, A. Fiore, et al. 2008. Quantum dot superluminescent diodes. In *Handbook of Self Assembled Semiconductor Nanostructures for Novel Devices in Photonics and Electronics*, ed. Henini, M., 565–599. Oxford: Elsevier.

Rossetti, M., J. Napierala, N. Matuschek, et al. 2012. Superluminescent light emitting diodes—The best out of two worlds. *SPIE Photonics West, MOEMS and Miniaturized Systems XI*, 8252–8206.

Schmitt, J. M., 1999. Optical Coherence Tomography (OCT): A Review. *IEEE J. Sel. Top. Quant. Electron.* 5, no. 4: 1205–1215.

Semenov, A. T., V. R. Shidlovski, S. A. Safin, et al. 1993a. Superluminescent diodes for visible (670 nm) spectral range based on AlGaInP/GaInP heterostructures with tapered grounded absorber. *Electon. Lett.* 29, no. 6: 530–532.

Semenov, A. T., V. R. Shidlovski, and S. A. Safin. 1993b. Wide-spectrum SQW superluminescent diodes at 0.8 μm with bent optical waveguide. *Electon. Lett.* 29, no. 10: 854–856.

Senior, J. M. 2009. *Optical Fiber Communications Principles and Practice*. Harlow: Pearson Education Limited.

Sun, Z. Z., D. Ding, Q. Gong, et al. 1999. Quantum-dot superluminescent diode: A proposal for an ultra-wide output spectrum. *Opt. Quant. Electron*. 31: 1235–1246.

Synopsis, Inc. Synopsys, 2017. http://www.synopsys.com/

Tansu, N. and L. J. Mawst. 2001. High-performance strain-compensated InGaAs–GaAsP–GaAs (λ = 1.17 μm) quantum-well diode lasers. *IEEE J. Photonics Technol. Lett*. 13, no. 3: 179–181.

Wild, G. and S. Hinckley. 2009. Distributed optical fibre smart sensors for structural health monitoring: A smart transducer interface module. *5th International Conference Intelligent Sensors, Sensor Networks and Information Processing*, Melbourne, Australia, 373–378.

Wootten, M. B., J. Tan, Y. J. Chien, et al. 2014. Broadband 2.4 μm superluminescent GaInAsSb/AlGaAsSb quantum well diodes for optical sensing of biomolecules. *Semicond. Sci. Technol*. 29, no. 11: 115014.

V

Semiconductor Optical Amplifiers (SOAs)

20 Semiconductor Optical Amplifier Fundamentals *Michael J. Connelly* 611
Introduction • Basic Principles • Quasi-Analytic Static Model • Bulk SOA Static Model Including ASE • Time-Domain Model Including ASE • Pulse Amplification Analytic Model • FWM Analytic Model • Conclusion

21 Traveling-Wave and Reflective Semiconductor Optical Amplifiers *Angelina Totović and Dejan Gvozdić* 631
Introduction • General Structure and Operation Principles of TW- and R-SOA • Modeling of Material and Structural Parameters • Rate Equations • Overview of Steady-State and Dynamic TW- and R-SOA Models • Conclusion

22 Tapered Semiconductor Optical Amplifiers *José-Manuel G. Tijero, Antonio Pérez-Serrano, Gonzalo del Pozo, and Ignacio Esquivias* 697
Introduction • Modeling Approaches • Model Implementation • Simulation Example: 1.5-μm InGaAsP/InP Tapered Amplifier • Summary

23 Quantum-Dot Semiconductor Optical Amplifiers *Benjamin Lingnau and Kathy Lüdge* ... 715
Introduction • QD Semiconductor Optical Amplifier Model • Application Examples • Conclusion

24 Wave Mixing Effects in Semiconductor Optical Amplifiers *Simeon N. Kaunga-Nyirenda, Michal Dlubek, Jun Jun Lim, Steve Bull, Andrew Phillips, Slawomir Sujecki, and Eric Larkins* 747
Introduction • Review of Nonlinear Phenomena in SOAs and Applications • Challenges in Modeling Nonlinear Effects • Improved Modeling of Nonlinear Phenomena in SOAs • Case Study: FWM • Summary and Conclusions

25 Semiconductor Optical Amplifier Dynamics and Pattern Effects *Zoe V. Rizou and Kyriakos E. Zoiros* 771
Introduction • SOA Dynamics Background • SOA Model Formulation • SOA Model Simplification • SOA Optical Gain Modulation • Pattern Effect I: Direct Optical Amplification • SOA Electrical Gain Modulation • Pattern Effect II: Direct Current Modulation

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

20

Semiconductor Optical Amplifier Fundamentals

20.1 Introduction 611
20.2 Basic Principles 612
Typical Bulk SOA Structure • Quantum Well, Dot and Dash SOAs
20.3 Quasi-Analytic Static Model 615
20.4 Bulk SOA Static Model Including ASE 617
Material Gain • Traveling-Wave and Carrier Density Rate Equations • Algorithm and Simulations
20.5 Time-Domain Model Including ASE 621
XGM Wavelength Converter
20.6 Pulse Amplification Analytic Model 623
SOA XPM Mach–Zehnder Wavelength Converter
20.7 FWM Analytic Model 627
20.8 Conclusion 628

Michael J. Connelly

20.1 Introduction

The rapid growth in the development of optical networks requires small, inexpensive, and easy-to-integrate optical amplifiers for use as basic amplifiers (power boosters, in-line amplifiers to compensate for fiber loss, and optical receiver preamplifiers) and also as optoelectronic signal processing devices such as wavelength converters, optical switches, intensity and phase modulators, logic gates, and dispersion compensators.

There are two main classes of optical amplifier: optical fiber amplifiers (OFAs) and semiconductor optical amplifiers (SOAs). OFAs are optical amplifiers that use optical fiber as the gain medium, which in most cases is a glass fiber doped with rare-earth ions such as erbium (for operation in the 1.55-μm telecommunications band). The active dopant is supplied with energy by an external pump laser. OFAs are relatively large devices, with advantages such as wide optical bandwidth (tens of nm), high gain (>20 dB), high saturation output power (>20 dBm), defined as the output signal power at which the gain is half that of the unsaturated gain, low-noise figure, and low polarization sensitivity. When an OFA is used to amplify an optical data signal, its slow gain dynamics (the lifetime of the excited energy levels are typically in the tens of ms range) are a significant advantage as the amplifier only experiences the average power of the data signal. This results in low intersymbol interference and, when used to amplify wavelength division multiplexed signals, low interchannel cross-talk. However, OFA's slow dynamics preclude its use in optical signal processing applications.

To achieve optical gain, an SOA uses an electrically pumped semiconductor material, as is the case for a semiconductor laser, such that a population inversion occurs between the material conduction and valence bands. An incoming light wave is amplified when the resulting stimulated emission exceeds losses due to stimulated absorption and other material or structural losses. By appropriate choice of the gain material and its bandgap energy, SOAs can be designed to operate in the wavelength region of choice, typically the 1.3- and 1.55-μm optical communication windows (Connelly, 2002a; Dutta and Wang, 2013). Various SOA designs have been shown to achieve gain, noise figure, saturation output power, and optical bandwidths comparable to those for OFAs; however, SOAs can exhibit significant polarization sensitivity. The main advantages of SOAs over fiber amplifiers are their small size and compatibility with photonic integrated circuit (PIC) technology. SOAs have much faster dynamics than OFAs, which can lead to data signal distortion and interchannel cross-talk when operated in the gain-saturated regime. However, high-speed dynamics and nonlinearities can be exploited to realize all-optical signal processing functions.

In this chapter, the basic principles and types of SOA are reviewed followed by descriptions and implementations of relatively simple models that are used to gain insight into important SOA static, dynamic, and nonlinear behavior. The models described are (1) quasi-analytic static model, (2) bulk SOA static model including amplified spontaneous emission (ASE), (3) time-domain model including ASE, which is used to simulate pattern effects and a cross-gain modulation (XGM)-based wavelength converter, (4) pulse amplification analytic model, which is also used to simulate a cross-phase modulation (XPM)-based wavelength converter, and (5) a four-wave mixing (FWM) analytic model. Further chapters describe more detailed descriptions and models of particular types of SOA.

20.2 Basic Principles

The principle of operation of a traveling-wave SOA is shown in Figure 20.1. The amplifier consists of an electrically pumped active waveguide. An incoming light wave is amplified as it propagates through the waveguide. Antireflection (AR) coatings are used to suppress the end facet reflections, which are approximately 32% in uncoated devices. Through the use of AR coatings and optimized waveguide designs (such as using a tilted waveguide geometry), reflectivities as low as 10^{-5} or less can be achieved. The models described in this chapter assume zero facet reflectivities. The input signal experiences a single-pass power gain $G = \exp(gL)$, where g is the gain coefficient at the signal wavelength and L the amplifier length. Although SOA waveguides are designed to be single mode, they support two orthogonal polarization modes, the transverse electric (TE) and transverse magnetic (TM) modes.

The amplification process adds broadband noise (spontaneous emission) to the propagating signal, caused by spontaneous recombination of conduction band electrons to valence band holes. This noise is subsequently amplified, leading to ASE. In the linear operating region, where the gain is constant and the ASE statistics are not affected by the input signal, the output ASE power spectral density (W/Hz) in a single polarization state at the signal energy E_s is given by

$$\rho_{\text{ASE}} = n_{\text{sp}} E_s (G - 1), \tag{20.1}$$

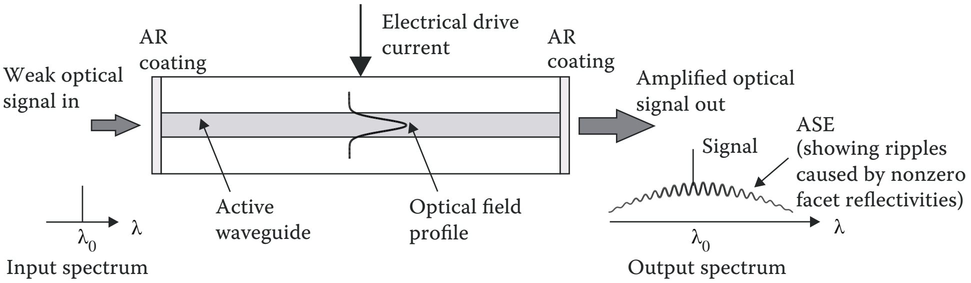


FIGURE 20.1 Semiconductor optical amplifier (SOA) basic structure.

where n_{sp} is the population inversion factor having a maximum possible value of 1. The amplifier noise factor F is defined as the ratio of the amplifier input to output signal-to-noise ratios (SNRs). F quantifies the degradation of the SNR, as determined in terms of the signal and noise levels in the photocurrent of an ideal photodetector placed in the signal path, due to the insertion of the amplifier in a signal transmission system (Baney et al., 2000; Kweon, 2002). The ideal photodetector responsivity $R = e/E$ (having a quantum efficiency equal to 1), where e and E are the electronic charge and light wave photon energy, respectively. Assuming that the input signal is shot-noise limited, the input SNR is equal to $P_{\mathrm{s}}/(2E_{\mathrm{s}}B_{\mathrm{e}})$, where P_{s} is the input signal power and B_{e} the detector electrical bandwidth. In practical receivers, the photodetector is preceded by an optical bandpass filter of bandwidth B_{o} that passes the signal and greatly reduces the ASE. The average signal and total ASE powers after the optical filter are $P_{\mathrm{s}}G$ and $2\rho_{\mathrm{ASE}}B_{\mathrm{o}}$, respectively, with corresponding photodetector shot-noise current variances (A^2), after electrical filtering, of $\sigma_{\mathrm{sig}}^2 = 2eRGP_{\mathrm{s}}B_{\mathrm{e}}$ and $\sigma_{\mathrm{sp}}^2 = 2eR\rho_{\mathrm{ASE}}B_{\mathrm{o}}B_{\mathrm{e}}$ (Olsson, 1989; Kweon, 2002). Since the photodetector is a square-law type detector, the signal also beats with the ASE giving rise to a signal-spontaneous beat noise current variance $\sigma_{\mathrm{s\text{-}sp}}^2 = 4R^2GP_{\mathrm{s}}\rho_{\mathrm{ASE}}B_{\mathrm{e}}$. The beating of the ASE with itself leads to spontaneous–spontaneous beat noise having a current variance $\sigma_{\mathrm{sp\text{-}sp}}^2 = 2R^2\rho_{\mathrm{ASE}}^2B_{\mathrm{e}}(2B_{\mathrm{o}} - B_{\mathrm{e}})$. The output SNR, $\mathrm{SNR}_{\mathrm{out}}$, is the ratio of the square of the signal photocurrent $(RGP_{\mathrm{s}})^2$ to the sum of the noise variances. If the optical filter bandwidth is small enough such that $\sigma_{\mathrm{sp\text{-}sp}}^2$ and σ_{sp}^2 are negligible compared to the other two noise sources (the signal-spontaneous beat noise limit), then

$$\mathrm{SNR}_{\mathrm{out}} = \frac{(RGP_{\mathrm{s}})^2}{4R^2GP_{\mathrm{s}}\rho_{\mathrm{ASE}}B_{\mathrm{e}} + 2eRGP_{\mathrm{s}}B_{\mathrm{e}}}. \tag{20.2}$$

The noise factor is then

$$F = \frac{2\rho_{\mathrm{ASE}}}{GE_{\mathrm{s}}} + \frac{1}{G}. \tag{20.3}$$

The noise figure (NF) is F expressed in decibel units, $\mathrm{NF} = 10\log_{10} F$ dB. For large gains, $F \approx 2n_{\mathrm{sp}}$. Since the maximum value of n_{sp} is 1, the minimum noise figure possible is 3 dB. Usually, the shot-noise term in Equation 20.3 is much smaller than the signal-spontaneous beat noise term so

$$F = \frac{2\rho_{\mathrm{ASE}}}{GE_{\mathrm{s}}}. \tag{20.4}$$

Nonzero reflectivities lead to multiple passes of the signal and ASE, with constructive interference occurring at discrete optical frequencies ν_q that are integer multiples r of the ratio of the speed of light in the amplifier c/n_{eff}, where n_{eff} is the waveguide effective index (defined as the ratio of the propagation constant in the waveguide to the free space propagation constant), to the round-trip distance $2L$. As the latter quantity is much greater than the wavelength, r is very large. The frequency spacing $\Delta\nu$ between adjacent resonances is $\Delta\nu = c/(2n_{\mathrm{eff}}L)$. For a typical SOA with $L = 1$ mm and $n_{\mathrm{eff}} = 3.5$ and operating in the 1.55-μm region, the wavelength spacing is 1.2 nm. When the effect of nonzero reflectivities is significant, the SOA is termed a Fabry–Pérot SOA (FP-SOA) because of its similarity to a Fabry–Pérot resonator. The signal gain of an FP-SOA at optical frequency ν is given by

$$G_{\mathrm{FP}}(\nu) = \frac{(1-R_1)(1-R_2)G}{(1-\sqrt{R_1R_2}G)^2 + 4\sqrt{R_1R_2}G\sin^2[\pi(\nu-\nu_r)/\Delta\nu]}, \tag{20.5}$$

where R_1 and R_2 are the SOA input and output facet reflectivities, respectively. FP resonances cause undesirable ripples in the signal gain and ASE spectrums (Figure 20.1). If the single-pass gain is high enough such that the denominator in Equation 20.5 approaches 0, G_{FP} will become very large and the SOA will begin to oscillate, at which point it behaves as a laser and not as an amplifier.

A reflective semiconductor optical amplifier (RSOA) can be formed by applying an AR coating to the input facet and a highly reflective coating to the opposite facet. RSOAs have applications such as modulators in passive optical networks (Lee et al., 2005).

20.2.1 Typical Bulk SOA Structure

A bulk material SOA operating in the 1550-nm region is shown in Figure 20.2 (Connelly, 2007). The active region is sandwiched between two separate confinement heterostructure (SCH) layers, which have a lower refractive index than the active region, and hence confine the light. The p–n junctions formed by the p-type and n-type InP layers act as current blocks, thereby providing good confinement of the injected carriers in the active region. An important SOA geometrical parameter is the optical confinement factor Γ defined as the fraction of the transverse (to the propagation direction) optical intensity overlapping with the active region. In general, Γ is polarization dependent, so the TE and TM confinement factors Γ_{TE} and Γ_{TM} are unequal. They are only equal in an unstrained bulk material SOA having a square cross-section waveguide. Such SOAs are difficult to fabricate as they require optical confinement in both transverse directions. Most commercial SOAs use rectangular cross-section waveguides and consequently, $\Gamma_{TE} \neq \Gamma_{TM}$. Confinement factors can be determined using formulas (Chuang, 2009) or commercial mode solvers. If $\Gamma_{TE} \neq \Gamma_{TM}$ and the material gain is polarization independent, then the SOA gain will be polarization dependent. In Figure 20.2, the introduction of tensile strain between the active region and SCH layers is used to increase the TM to TE material gain ratio, to compensate for the higher Γ_{TE} value caused by the waveguide asymmetry, and thereby reduce polarization sensitivity. The tapered regions act as mode expanders that couple light from the active waveguide to an underlying passive waveguide, thereby simplifying coupling to external optical fibers.

20.2.2 Quantum Well, Dot and Dash SOAs

SOAs that use bulk materials require high transparency current densities (at which the SOA has unity gain). The active layer in quantum well (QW) SOAs have much smaller thicknesses (~5–10 nm) than in conventional bulk structures (~100 nm). In bulk material, the injected carriers can move in three dimensions whereas in a QW they are confined to two dimensions. The enhanced quantum effects result in a significantly different band structure and material gain compared to bulk material. The small thickness of the QW allows many QWs to be stacked to form a multiple quantum well (MQW) active region. MQW SOAs have a number of advantages compared to bulk SOAs. Because of the small active region volume compared to bulk SOAs, a reduced injection current is sufficient to create a large carrier density in the

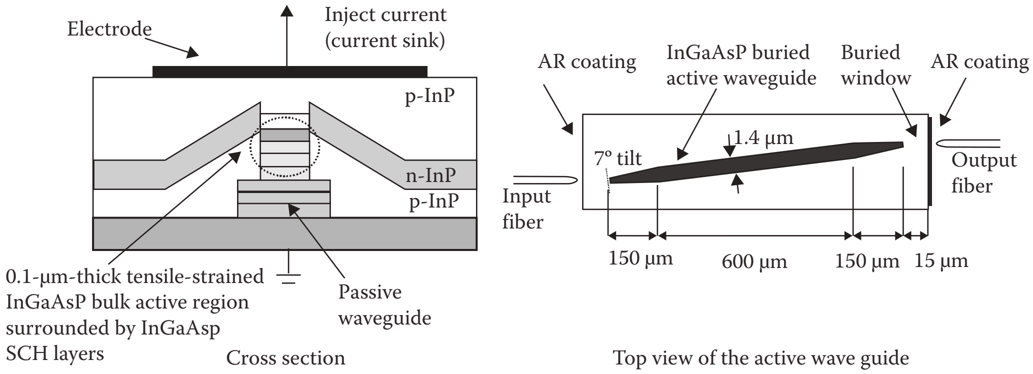


FIGURE 20.2 Typical SOA structure (Connelly, 2007). The buried windows and 7° tilted waveguide (with respect to the end facets) further reduce the effective facet reflectivity.

active region resulting in a broader gain spectrum and shorter carrier lifetime (related to the average time for conduction band electrons to recombine with valence band holes). Shorter carrier lifetimes result in shorter gain recovery times, which is of particular importance in reducing pattern effects when the SOA is used to amplify high-speed data. Furthermore, the loss coefficient in MQW active regions is significantly smaller than that in bulk devices, which leads to an improvement in noise performance. The polarization sensitivity of QW structures can be reduced by using strained QWs. Strain-induced band structure modifications also result in a reduction in loss mechanisms such as Auger recombination and intervalence band absorption (IVBA). The linewidth enhancement factor (LEF) is also reduced, which results in amplified pulses experiencing less spectral broadening, leading to superior high-speed performance as compared to bulk and unstrained QW SOAs. Bulk and QW SOAs have typical gain recovery lifetimes in the range of hundreds of picoseconds; so, when used to amplify data signals, significant pattern effects can be present at baud rates in the region of tens of Gbaud/s. Pattern effects also induce nonlinear phase noise through the self-phase modulation (SPM) effect, which can be a more serious performance degradation factor when advanced modulation formats are used. The amount of SPM is directly related to the LEF.

A quantum dot (QD) is a semiconductor nanostructure with dimensions that typically range from 2 to 10 nm, which confines the motion of injected carriers to all three spatial directions (Akiyama et al., 2007). Compared to bulk and QW SOAs, QD SOAs have been shown to have improved gain bandwidth, noise figure, and saturation output power as well as enhanced nonlinear effects such as FWM. Polarization insensitive QD SOAs have been realized using optimized QD shapes along with close stacking of the QD layers. Of particular significance is the ultrafast gain recovery lifetime, typically of the order of a few picoseconds, a factor of ten lower than that for bulk and QW SOAs so that QD SOAs are capable of amplifying ultrafast data signals with little or no pattern effects (Ben-Ezra et al., 2005). QD SOA LEFs can be much lower than for bulk and QW SOAs and so are particularly suitable for amplifying very high baud rate advanced modulation data signals.

Quantum dash (QDash) SOAs are of interest as an alternative to QD SOAs, since they have some dot-like properties and can more easily be made to operate in the 1.55-μm region, although they also have been shown to have longer gain recovery times in the range of hundreds of picoseconds (Lelarge et al., 2007).

20.3 Quasi-Analytic Static Model

As the input signal power to an SOA is increased, the gain reduces as the electrically pumped conduction band carriers (electrons) in the active material are depleted by stimulated recombination with holes in the valence band. To determine the factors that influence SOA gain at high input powers, a simple traveling-wave-based model can be used (Connelly, 2002a). The carrier density n dependent material gain g_{m} per unit length at the signal wavelength is assumed to be a linear function $g_{\mathrm{m}} = a(n - n_{\mathrm{t}})$, where a is the differential of the material gain with respect to n and n_{t} is the transparency carrier density. The carrier density rate equation is

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{\eta I}{eV} - \frac{n}{\tau_{\mathrm{c}}} - \frac{\Gamma a(n - n_{\mathrm{t}})P}{AE_{\mathrm{s}}}, \tag{20.6}$$

where I is the bias current and E_{s} the signal photon energy. The injection efficiency η is the fraction of the SOA current entering the active region. The active region cross-section area, $A = dW$, where d and W are the active waveguide thickness and width, respectively, and the active region volume $V = AL$. τ_{c} is the interband carrier lifetime due to nonradiative and radiative spontaneous recombination processes including trap, Auger, and spontaneous emission. The propagation of the signal power P is described by the traveling-wave equation

$$\frac{\mathrm{d}P}{\mathrm{d}z} = \left[\Gamma a(n - n_{\mathrm{t}}) - \alpha\right] P, \tag{20.7}$$

where z is the propagation direction (along the amplifier axis) measured from the input and α is the loss coefficient. Under steady-state conditions, the time differential in Equation 20.6 is zero, from which

$$n = \left(\frac{\tau_c \eta I}{eV}\right)\frac{P_{sat}}{P + P_{sat}} + n_t \frac{P}{P + P_{sat}}. \tag{20.8}$$

The saturation power P_{sat} is defined as

$$P_{sat} = \frac{AE_s}{\Gamma a \tau_c}, \tag{20.9}$$

Defining the spatially dependent normalized power $p(z) = P(z)/P_{sat}$, the unsaturated carrier density $n_0 = \tau_c \eta I/(eV)$ and the unsaturated gain coefficient $g_0 = \Gamma a (n_0 - n_t)$, Equation 20.8 can be written as

$$n(z) = \frac{n_0}{1+p} + \frac{n_t p}{1+p}. \tag{20.10}$$

Inserting Equation 20.10 into Equation 20.7 gives

$$\frac{dp}{dz} = \left(\frac{g_0}{1+p} - \alpha\right)p. \tag{20.11}$$

The unsaturated gain $G_0 = \exp[(g_0 - \alpha)L]$. Equation 20.11 can be solved numerically, using, e.g., the Runge–Kutta method, with the boundary condition $p(0) = P_{in}/P_{sat}$, where P_{in} is the input signal power. The amplifier gain G is calculated as the ratio of the output power $P_{out} = p(L)P_{sat}$ to the input power. The calculated gain versus output power is shown in Figure 20.3a for various values of G_0 with $\exp(\alpha L) = 10$. The carrier density is calculated using Equation 20.10. For a given SOA, G_0 corresponds to a particular value of bias current. The normalized saturation output power $p_{o,sat}$ is defined as the normalized output power at which the amplifier gain is half the unsaturated gain. From Figure 20.3a, it can be seen that $p_{o,sat} \approx -3.8$ dB for the 20 dB unsaturated gain curve and almost independent of G_0. The saturation output power $P_{o,sat} = p_{o,sat}P_{sat}$, which from the form of Equation 20.9 can be increased by reducing a and τ_c. In real SOAs, these two parameters are not constants but for a given structure and material depend on wavelength and carrier density.

Consider a square cross-section bulk SOA with parameters $W = d = 0.4$ μm, $L = 600$ μm, $\Gamma = 0.45$, $\tau_c = 0.5$ ns, $a = 1 \times 10^{-20}$ m^2, $n_t = 2.5 \times 10^{24}$ m^{-3}, and $\exp(\alpha L) = 10$. If the unsaturated gain is 20 dB at a bias current of 150 mA and assuming 100% injection efficiency, then $n_o = 4.9 \times 10^{24}$ m^{-3}, $g_0 = 1.1 \times 10^4$ m^{-1} and for a signal wavelength of 1550 nm, $P_{sat} = 9.6$ dBm and so $P_{o,sat} = 5.8$ dBm. Figure 20.3b shows the signal power and carrier density distributions for various values of normalized input power. At input powers $\ll P_{sat}$, the carrier density is uniform throughout the amplifier. Models that include ASE show that below saturation, the carrier density is not uniform but is symmetrical with minima at the SOA ends and a maximum at the center. As the level of saturation increases, the carrier distribution becomes less uniform, until the amplifier is highly saturated, at which the carrier density throughout the amplifier approaches its transparency value. As Figure 20.3b shows, the spatial distribution of the carrier density in a SOA can vary greatly. A further disadvantage of this model is that the spontaneous recombination term in Equation 20.4 is assumed to be a linear function of n. In practice, spontaneous recombination is more accurately modeled as a polynomial function of carrier density, so τ_c is actually carrier dependent.

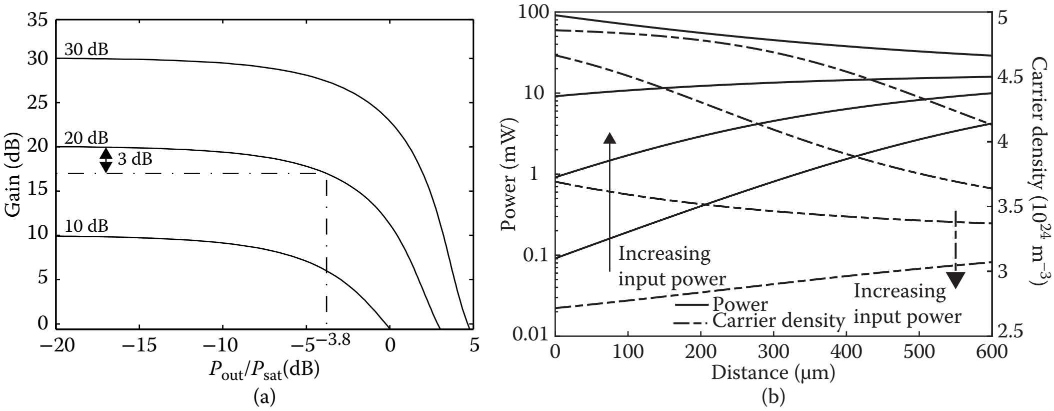


FIGURE 20.3 (a) Gain characteristics for various unsaturated gains. (b) Carrier density and power distributions for normalized input powers of 0.01, 0.1, 1, and 10. The unsaturated gain is 20 dB.

20.4 Bulk SOA Static Model Including ASE

The simple model described above illustrates SOA operational principles, in particular, saturation effects, but it is of limited use in a real SOA analysis. Most importantly, it does not include gain coefficient wavelength dependency or ASE, which contributes to gain saturation and determines the amplifier noise performance. In this section, a modified version of the wideband models for an unstrained bulk SOA that includes ASE (Connelly, 2001, 2007) is considered. The SOA has the same active region geometry and confinement factor as used in the quasi-analytic static model. The bulk material is unstrained, and thereby polarization independent. In the mathematical formulation, Γ_{TE} and Γ_{TM} are assumed to be different but are assumed to be equal in the simulations.

20.4.1 Material Gain

The material gain g_m and spontaneous gain g_{sp} (used to determine the additive spontaneous emission) of the unstrained bulk material are calculated using methods detailed in Jones and O'Reilly (1993) and Connelly (2007). Typical gain spectra are shown in Figure 20.4. The net gain coefficient, at energy E and polarization p (TE or TM), $g_p(n, E) = \Gamma_p g_{m,p} - \alpha_p$ and the loss coefficient $\alpha_p(n) = \alpha_0 + \Gamma_p\,\alpha_1\,n$, where α_0 is the waveguide scattering loss and α_1 the IVBA coefficient loss coefficient (Suematsu and Adams, 1994), with values of 3000 m^{-1} and 7500×10^{-24} m^2, respectively (Connelly, 2001).

20.4.2 Traveling-Wave and Carrier Density Rate Equations

A traveling-wave approach is used to determine signal and the TE and TM ASE photon rates. Transverse variations in the optical field are not considered, and as such, the model has a one-dimensional spatial dependency. The traveling-wave equation for the forward propagating signal photon rate N_s is

$$\frac{dN_s}{dz} = g_s\,N_s, \tag{20.12}$$

where g_s is the net gain coefficient at the signal polarization and photon energy E_s. The boundary condition at the input $N_s(z = 0) = P_{in}/E_s$, where P_{in} is the input signal power. For a given carrier density spatial

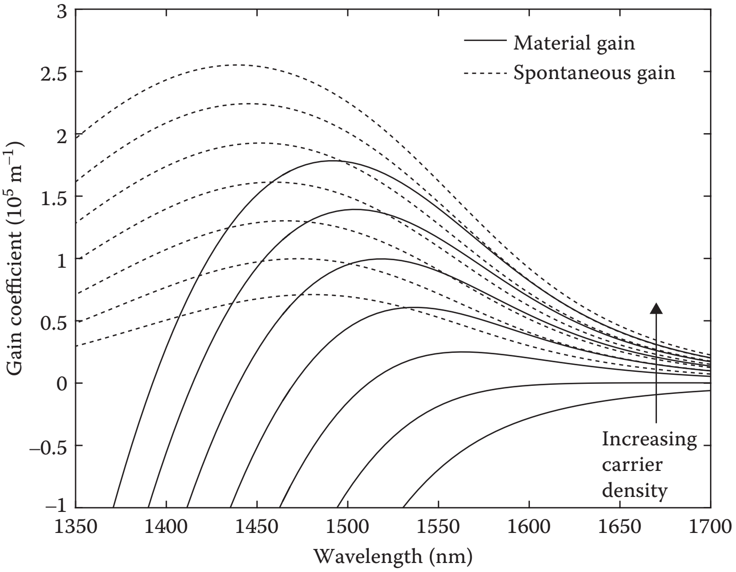


FIGURE 20.4 Unstrained bulk material gain and spontaneous emission gain spectra. The carrier density ranges from 2 to 5×10^{24} m^{-3} in increments of 0.5×10^{24} m^{-3}. The model takes into account intraband relaxation effects and carrier density–dependent bandgap shrinkage. The bandgap wavelength at zero carrier density is 1.55 μm.

distribution, Equation 20.12 has the solution

$$N_s(z) = \frac{P_{in}}{E_s} \exp\left(\int_0^z g_s \mathrm{d}z \right), \tag{20.13}$$

which can be calculated using numerical integration. To model ASE, a spectral slicing scheme is used in which the ASE is split into M spectral slices, each of energy width ΔE. The traveling-wave equations for the forward (+) and backward (−) propagating kth spectral slice $N^{\pm}_{p,k}$ with polarization p and centered at energy E_k are given by

$$\frac{\mathrm{d}N^{\pm}_{p,k}}{\mathrm{d}z} = \pm g_{p,k} N^{\pm}_{p,k} \pm \Gamma_p g_{sp,p,k} \Delta E, \tag{20.14}$$

with boundary conditions $N^{+}_{p,k}(z = 0) = 0$ and $N^{-}_{p,k}(z = L) = 0$. The last term on the right-hand side (RHS) of Equation 20.14 is the additive spontaneous emission into the kth spectral slice. For a given carrier density spatial distribution, Equation 20.14 has the solution

$$N^{+}_{p,k}(z) = \exp\left(-\int_0^z g_{p,k}\, \mathrm{d}z \right) \int_0^z \left[\exp\left(\int_0^z g_{p,k}\, \mathrm{d}z \right) \Gamma_p\, g_{sp,p,k} \Delta E \right] \mathrm{d}z \tag{20.15}$$

$$N^{-}_{p,k}(z) = \exp\left(-\int_z^L g_{p,k}\, \mathrm{d}z \right) \int_z^L \left[\exp\left(\int_z^L g_{p,k}\, \mathrm{d}z \right) \Gamma_p\, g_{sp,p,k} \Delta E \right] \mathrm{d}z, \tag{20.16}$$

which can be calculated using numerical integration. The carrier density rate equation is

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{\eta I}{eV} - R(n) - \frac{1}{\mathrm{d}W}\left[\Gamma_{\mathrm{s}}\, g_{\mathrm{m},p}N_{\mathrm{s}} + \sum_{p=\mathrm{TE,TM}}\sum_{k=1}^{M}\Gamma_p\, g_{\mathrm{m},p,k}(N^{+}_{p,k} + N^{-}_{p,k})\right]. \tag{20.17}$$

The first term on the RHS of Equation 20.17 is the carrier pumping due to the injected bias current. The injection efficiency η is assumed to be equal to 1. Γ_{s} is the confinement factor of the signal polarization. The nonstimulated emission recombination rate $R(n)$ is given by

$$R(n) = A_{\mathrm{tr}}n + R_{\mathrm{rad}}(n) + Cn^3, \tag{20.18}$$

where A_{tr} and C are the trap and Auger coefficients, respectively. Trap recombination is mainly caused by Shockley–Read–Hall recombination at defects in the material (Fukuda, 1991). The spontaneous radiative recombination rate $R_{\mathrm{rad}}(n)$ is commonly assumed to be a bimolecular process approximated by $R_{\mathrm{rad}} = Bn^2$, where the bimolecular recombination coefficient B is often taken to be a fitting parameter; however, it can be directly calculated from the material gain (Chuang, 2009). For the material under consideration, $B \approx 2.8 \times 10^{-16} - 0.24 \times 10^{-40}\mathrm{n}\ \mathrm{m}^3\mathrm{s}^{-1}$. Auger recombination involves three carriers: an electron and hole, which recombine in a band-to-band transition and give off the resulting energy to another electron or hole (Chuang, 2009). Because SOAs usually operate with high carrier (electron) densities, the hole density is almost equal to the carrier density so the Auger recombination rate is proportional to n^3. The Auger coefficient for a particular material or structure can be calculated from first principles, but this is difficult and there is often a high degree of uncertainty in values quoted in the literature. In most models, typical values are used or it is treated as a fitting parameter. Here, C is taken to be equal to $3 \times 10^{-41}\ \mathrm{m}^6\mathrm{s}^{-1}$. At the high carrier densities present in SOAs and if the defect level is low, trap-induced recombination is often assumed to be negligible compared to the other two processes and can be ignored, as is the case here. The third term on the RHS of Equation 20.17 is the amplified signal-induced carrier density depletion rate. $\Gamma_{\mathrm{s}}N_{\mathrm{s}}$ is the photon rate in the active region, which is multiplied by $g_{\mathrm{m},p}$ to obtain the conduction band to valence band carrier recombination rate per unit length. Division by the active region cross-section area results in the carrier density depletion rate. The fourth term on the RHS of Equation 20.17 is the ASE-induced carrier density depletion rate, which is obtained in a similar fashion to that for the amplified signal.

20.4.3 Algorithm and Simulations

Under static conditions, the time differential in Equation 20.17 is equal to zero; the aim of the numerical algorithm is to determine the signal, ASE photon rates, and the carrier density such that this is the case (Connelly, 2007). The model equations cannot be solved analytically, thereby requiring a numerical solution (Connelly, 2001, 2007). Fifty spatial points are used for numerical integration and 150 spectral slices covering a range of 1250–1650 nm. First, the material gain and spontaneous gain are calculated for the spectral slice energies and particular values of carrier density within the expected range. The bimolecular recombination coefficient is calculated for the same carrier density values. Initially, the carrier density in the amplifier is set to some reasonable value (e.g., $3 \times 10^{24}\ \mathrm{m}^{-3}$). The signal intensity and ASE photon rates are then estimated using Equations 20.13, 20.15, and 20.16. An updated value for n at each spatial section is then estimated using Equation 20.17 with the time derivative set to zero, and subsequently the gain and spontaneous recombination rate throughout the amplifier, using interpolation of the full calculations, followed by the signal and ASE photon rates. This process is continued until the values of the ASE photon rates converge to a suitable tolerance (0.1%). Amplifier characteristics such as gain, ASE spectrum, and noise figure can then be calculated. The polarization-dependent ASE output spectral density (W/Hz) at E_k is given by

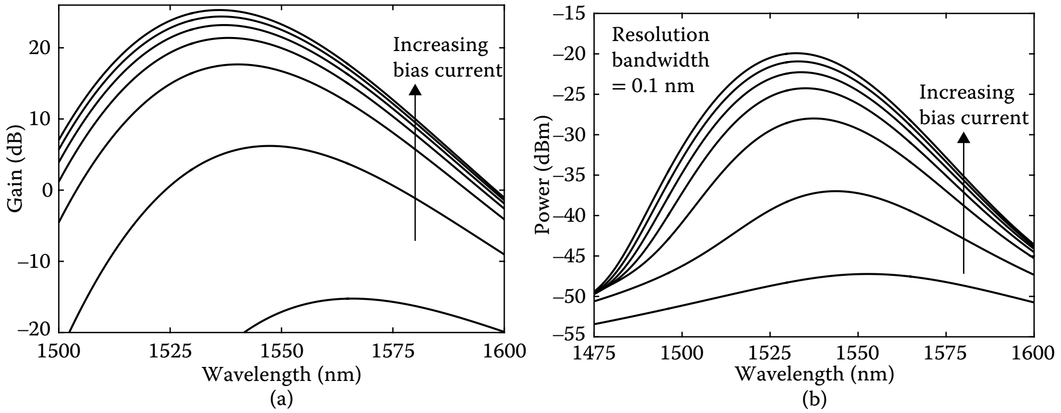
$$\rho_{p,k} = \frac{h E_k}{\Delta E} N^{+}_{p,k}(z = L). \tag{20.19}$$

In the signal-spontaneous beat noise limit, the noise figure spectrum is given by

$$NF_{p,k} = 10 \log_{10}\left(\frac{2\rho_{p,k}}{E_k G_{p,k}}\right), \tag{20.20}$$

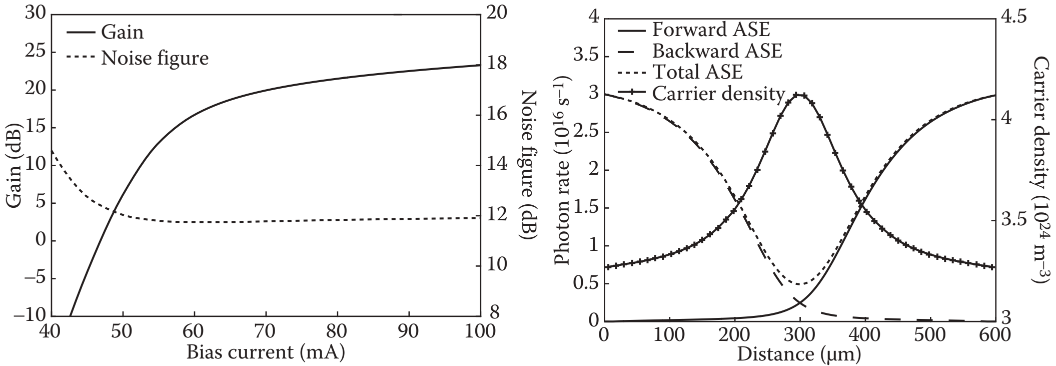
where $G_{p,k}$ is the amplifier gain at E_k (Baney et al., 2000; Olsson, 1989). The maximum gain achievable is mainly limited by ASE. The small-signal gain and ASE spectra are shown in Figure 20.5 for bias currents ranging from 40 to 100 mA. The gain and ASE spectrum peaks shift to shorter wavelengths as the bias current increases, an effect which is present in real SOAs, and they both saturate at high bias currents. The 3-dB optical gain bandwidth at a bias of 100 mA is 30.6 nm. The corresponding ASE spectrum has a 3-dB bandwidth of 34.4 nm; in general, the ASE spectral width is a reasonably good measure of the amplifier gain bandwidth. Figure 20.6a shows the small-signal gain and noise figure bias current dependency. The gain begins to saturate at bias currents at which the ASE saturates; the noise figure reaching a limiting value of approximately 11.7 dB. The ASE and carrier density spatial distributions are shown on Figure 20.6b for no input signal. The latter has a symmetrical profile with a maximum at the center, in contrast to the quasi-analytic model, which does not include ASE. The gain, output signal, and ASE powers versus input power characteristics are shown in Figure 20.7a for a bias current of 100 mA at which the unsaturated gain is 23.3 dB and $P_{\mathrm{o,sat}} = 4.9$ dBm. The reduction in ASE power is clearly seen as the amplifier is driven into saturation by the input signal. As the input power level is increased, the amplified signal begins to compete for the available gain and the carrier density distribution becomes more asymmetrical, as shown in Figure 20.7b for a moderate saturation level.

The model can be used to investigate the effects of different material and geometrical parameters on SOA performance and is thereby useful in SOA design.



(a)
(b)

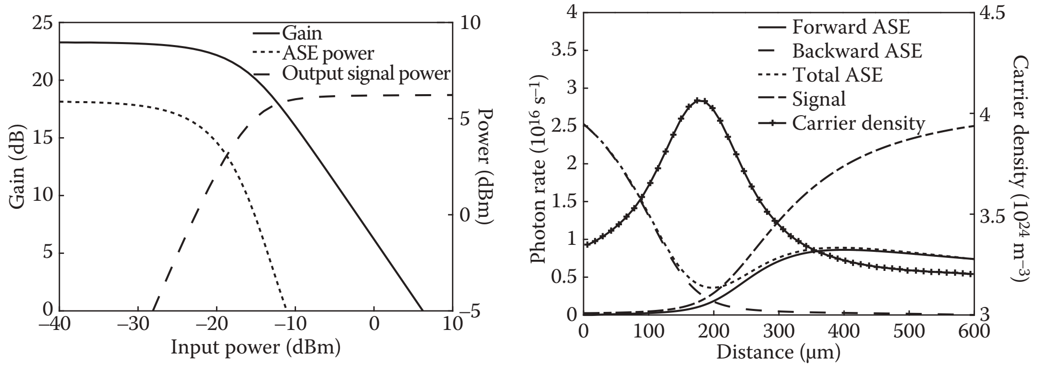
FIGURE 20.5 (a) Gain spectra and (b) amplified spontaneous emission (ASE) spectra (summed over both polarizations) in the absence of an input signal for bias currents of 40–100 mA in increments of 10 mA.



(a)

(b)

FIGURE 20.6 Small-signal (a) gain and noise figure versus bias current at 1550 nm, (b) ASE and carrier density spatial distributions. The bias current and signal wavelength are 100 mA and 1550 nm, respectively.



(a)

(b)

FIGURE 20.7 (a) Gain, output signal and ASE powers versus input power. (b) ASE and carrier density spatial distributions for an input power of −15 dBm, at which the gain has been reduced from its unsaturated value of 23.3–20 dB. The bias current and signal wavelength are 100 mA and 1550 nm, respectively.

20.5 Time-Domain Model Including ASE

The most common application of SOAs is for modulated signal amplification, so it is of interest to model time-domain behavior such as pattern effects, XGM, SPM, and XPM (Obermann et al., 1998; Durhuus et al., 1992; Asghari et al., 1997; Occhi et al., 2003). By including the time differential term in Equation 20.17, the static model described above can be used to model SOA dynamics having timescales as short as hundreds of picoseconds (Connelly, 2002b). The carrier density is initialized to some suitable value and the initial values of the signal and ASE photon rates are set to zero. The carrier density at the next time step is determined by solving Equation 20.17 using the modified Euler method, which is a second-order technique (Press et al., 2007) and for a given time step is significantly more accurate than simply replacing the time differential by a first-order finite difference (Euler method). Nonetheless, an appropriate time-step size must be chosen to prevent stability issues and ensure accuracy. Assuming that the propagation time through the amplifier is much less than the time variation of the optical input and current, the ASE and signal photon rates can be determined using Equations 20.13, 20.15, and 20.16. The new values of the photon rates are then used in Equation 20.17 to determine the carrier density at the next time step until the time span of interest is completed.

The carrier density rate Equation 20.17 can be expressed as

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{\eta I}{eV} - n\left(\frac{1}{\tau_\mathrm{c}} + \frac{1}{\tau_\mathrm{s}}\right), \tag{20.21}$$

where the inverse of the nonstimulated emission carrier lifetime τ_c is given by

$$\frac{1}{\tau_\mathrm{c}} = \frac{R(n)}{n}, \tag{20.22}$$

and the inverse of the stimulated recombination carrier lifetime τ_s is given by

$$\frac{1}{\tau_\mathrm{s}} = \frac{1}{n\,\mathrm{d}W}\left[\Gamma_\mathrm{s}\, g_{\mathrm{m},p} N_\mathrm{s} + \sum_{p=\mathrm{TE,TM}} \sum_{k=1}^{M} \Gamma_p\, g_{\mathrm{m},p,k}(N_{p,k}^{+} + N_{p,k}^{-})\right]. \tag{20.23}$$

The total effective carrier lifetime τ_eff is

$$\tau_\mathrm{eff} = \left(\frac{1}{\tau_\mathrm{c}} + \frac{1}{\tau_\mathrm{s}}\right)^{-1}, \tag{20.24}$$

which depends on the amplifier operating point and input optical power and also has a strong spatial dependency. The smaller the τ_eff, the faster the gain recovery time. Figure 20.8 shows the output power, spatially averaged effective carrier lifetime $\tau_\mathrm{eff,m}$, and instantaneous gain for an amplified 5-Gb/s non-return-to-zero (NRZ) data stream with high and low powers of 0.1 mW and 0.1 μW, respectively. The bias current, unsaturated gain, and signal wavelength are 100 mA, 23.3 dB, and 1550 nm, respectively. The amplified data distortion is due to the dynamic changes of the carrier and photon distributions causing pattern effects on a timescale approximately given by $\tau_\mathrm{eff,m}$, which for the simulations shown in Figure 20.8 can vary between 370 and 850 ps. At higher bit rates the severity of pattern effects increases and consequently their effect on system performance.

20.5.1 XGM Wavelength Converter

XGM occurs when an intensity-modulated input data signal (the pump) causes dynamic changes in the carrier density and thereby the gain experienced by a copropagating or counter-propagating light wave (the probe) at a different wavelength to the data signal. Carrier density changes also cause changes in the active region refractive index and thereby the phase of both the pump and probe leading to SPM and XPM. XGM can be used to realize wavelength converters in which data is copied from an input pump data signal to an input continuous wave (CW) probe. An XGM-based 2.5-Gb/s wavelength converter with copropagating pump and probe light waves is shown in Figure 20.9. When the pump power is high, the gain experienced by the probe is reduced and vice versa, hence the inverted pump data is transferred to the probe. The above time-domain model is used to simulate the converter, as shown in Figure 20.9. The performance degrades as the bit rate increases, as the dynamic gain is unable to follow the variations in the input pump power. Because of their superior dynamics, compared to bulk or QW SOAs, QD SOAs can be used as XGM wavelength converters at much higher bit rates.

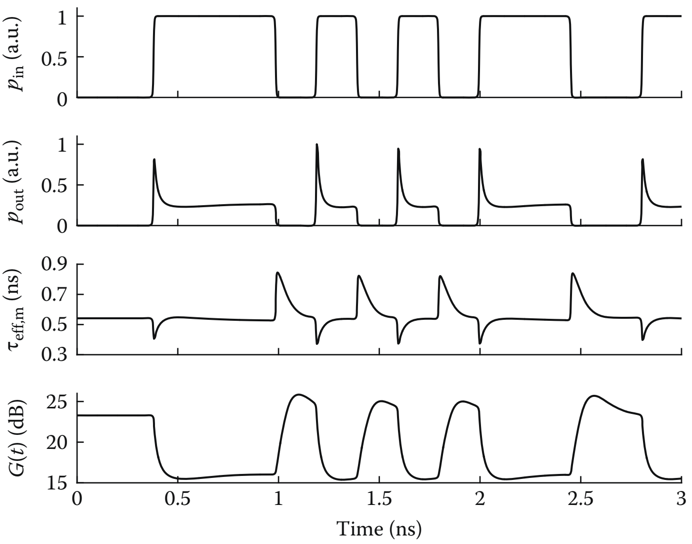


FIGURE 20.8 Simulated SOA output power, spatially averaged effective carrier lifetime, and instantaneous gain for an input NRZ 5-Gb/s data sequence. The data high- and low-level powers are 0.1 mW and 0.1 μW, respectively. The bias current, unsaturated gain, and signal wavelength are 100 mA, 23.3 dB, and 1550 nm, respectively. The simulation time step is 2.5 ps.

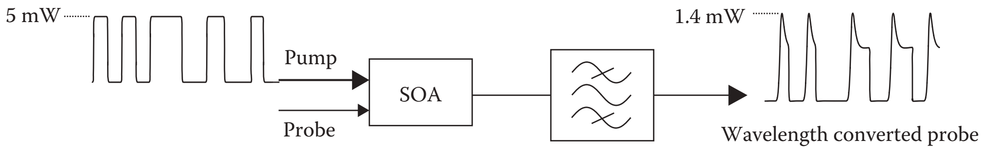


FIGURE 20.9 Cross-gain modulation (XGM)-based wavelength converter and simulated wavelength converted probe output. The bandpass filter is used to select the probe signal. The pump input is a 30-dB extinction ratio NRZ 2.5-Gb/s data sequence with a peak power and wavelength of 5 mW and 1550 nm, respectively. The input probe power and wavelength are 10 μW and 1560 nm, respectively. The bias current is 100 mA.

20.6 Pulse Amplification Analytic Model

Because SOAs have wide optical bandwidths and fast gain recovery lifetimes, they can be used to amplify optical pulses with full width at half-maximum (FWHM) pulse widths in the picosecond range. By exploiting SOA nonlinearities using materials with intrinsically faster gain recovery times, it is possible to amplify femtosecond pulses. Many pulse amplification models are described in the literature; here a simple model (Agrawal and Olsson, 1989), which does not include ASE, is described. The model is applicable to pulse widths greater than the intraband relaxation time (typically <0.1 ps). The amplifier is assumed to be polarization independent. The optical field can be written as $A(t) = \sqrt{P(t)}\,\exp[j\varphi(t)]$, where P and φ are the temporal power and phase. The gain coefficient g at the signal wavelength is modeled as $g = \Gamma a\,(n - n_{tr})$. The power, phase, and gain coefficient dynamics are determined by solving

$$\frac{\partial P}{\partial z} = (g - \alpha)P \tag{20.25}$$

$$\frac{\partial\varphi}{\partial z} = -\frac{1}{2}\alpha_{\mathrm{H}}\, g \tag{20.26}$$

$$\frac{\partial g}{\partial\tau} = \frac{(g - g_0)}{\tau_{\mathrm{c}}} - \frac{gP}{E_{\mathrm{sat}}}. \tag{20.27}$$

τ is the pulse local time measured with respect to a time frame moving with the pulse. The temporal dependence of g, induced by the propagating pulse power, leads to dynamic changes in the pulse phase, i.e., SPM. τ_{c} is the interband carrier lifetime, which is typically the order of hundreds of picoseconds in bulk and QW SOAs, and α_{H} is the LEF—a proportionality factor relating phase changes to changes in the gain (Henry, 1982). α_{H} is often taken to be a constant, but for a given material is carrier density and wavelength dependent. The saturation energy $E_{\mathrm{sat}} = E_{\mathrm{s}} Wd/(\Gamma a)$. To determine the evolution of the pulse shape and phase, a numerical solution of Equations 20.25 through 20.27 is required. However, if it is assumed that $\alpha << g$ then Equations 20.23 through 20.25 can be solved to obtain a closed-set of equations for the output pulse power and phase given by

$$P_{\mathrm{out}}(\tau) = G(\tau)P_{\mathrm{in}}(\tau) \tag{20.28}$$

$$\varphi_{\mathrm{out}}(\tau) = \varphi_{\mathrm{in}}(\tau) - \frac{\alpha_H}{2}h(\tau), \tag{20.29}$$

where the instantaneous gain $G(\tau) = \exp[h(\tau)]$. If the input FWHM pulse width $\tau_p << \tau_c$, then the integrated gain $h(\tau)$ is given by

$$h(\tau) = -\ln\left\{1 - \left(1 - \frac{1}{G_0}\right)\exp\left[-\frac{U_{\mathrm{in}}(\tau)}{E_{\mathrm{sat}}}\right]\right\}, \tag{20.30}$$

where $G_0 = \exp(g_0 L)$ is the unsaturated gain, g_0 is the unsaturated gain coefficient, and

$$U_{\mathrm{in}}(\tau) = \int_{-\infty}^{\tau} P_{\mathrm{in}}(\tau')\,\mathrm{d}\tau'. \tag{20.31}$$

The output pulse frequency chirp, defined as the time derivative of its instantaneous phase divided by 2π, is given by

$$\Delta\nu_{\mathrm{out}}(\tau) = \Delta\nu_{\mathrm{in}}(\tau) + \frac{\alpha(G_0 - 1)}{4\pi G_0}\frac{P_{\mathrm{out}}(\tau)}{E_{\mathrm{sat}}}\exp\left[\frac{-U_{\mathrm{in}}(\tau)}{E_{\mathrm{sat}}}\right], \tag{20.32}$$

where $\Delta\nu_{\mathrm{in}}$ is the input pulse chirp. The model is applicable to pulse widths of the order of picoseconds to tens of picoseconds. For pulse widths >1 ps, nonlinear intraband processes such as carrier heating (CH) and spectral hole burning (SHB) that have characteristic time constants in the subpicosecond range are important; so, models that take such effects into account are required (Hong et al., 1994; Mecozzi and Mork, 1997). Once the output pulse power and phase are known, its power spectrum can be calculated. To show the effects of an SOA on an amplified pulse consider an unchirped zero phase Gaussian input pulse,

$$P_{\mathrm{in}}(\tau) = \frac{E_{\mathrm{in}}}{\tau_0\sqrt{\pi}}\exp\left[-\left(\frac{\tau}{\tau_0}\right)^2\right], \tag{20.33}$$

where E_{in} is the pulse energy. τ_o is related to τ_p by $\tau_p \approx 1.665\tau_0$. Simulated amplified pulse shape, chirp, and spectrum are shown in Figure 20.10. The amplified pulse is asymmetric because the leading edge of

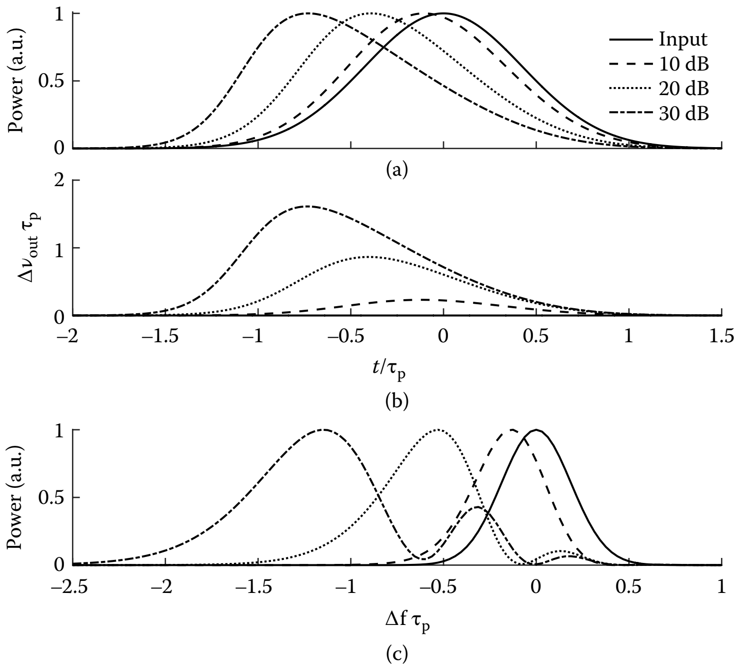


FIGURE 20.10 Amplified pulse; (a) power, (b) chirp, and (c) spectrum. Δf is the frequency deviation from the input light wave center frequency. The input is a zero-chirp Gaussian pulse with $E_{in}/E_{sat} = 0.1$. The parameter is the unsaturated gain and $\alpha_H = 5$.

the pulse experiences a larger gain than the trailing edge. The amplified pulse is chirped and its spectrum is broader than the input spectrum. At high gains, the amplified pulse spectrum exhibits a multipeak structure caused by SPM-induced frequency chirp.

20.6.1 SOA XPM Mach–Zehnder Wavelength Converter

XGM-based wavelength converters suffer from the disadvantage that high-input pump powers are required to obtain the large SOA gain modulation depth necessary to achieve a large converted signal extinction ratio. The associated high-carrier depletion rates lead to a significant increase in the gain recovery lifetime, which limits the data rates at which such converters can operate. To overcome the trade-off between speed and modulation depth SOA XPM can be used (Asghari et al., 1997). XPM occurs when an intensity-modulated input data signal (the pump) causes dynamic changes in the carrier density and, consequently, the active region refractive index and thereby the phase of a copropagating or counter-propagating probe signal at a different wavelength to the pump. In an SOA XPM-based wavelength converter, the probe phase changes are converted into amplitude changes by incorporating one or two SOAs in an interferometric structure such as the Mach–Zehnder interferometer (MZI) shown in Figure 20.11. If sufficient phase shift is achieved with a low-gain modulation index, and the interferometer visibility is high, then the XPM converter performance will be superior to XGM-based converters. An optical delay line is used to adjust the relative delay between the modulated input pump pulses reaching each SOA. The interferometer output probe power, assuming that destructive interference occurs between the recombining probe light waves when their relative phases are equal, is given by

$$p(t) = \frac{P_{probe}}{4}\left[G_1(t) + G_2(t) - 2\sqrt{G_1(t)G_2(t)}\cos\left(\varphi_1(t) - \varphi_2(t)\right)\right], \tag{20.34}$$

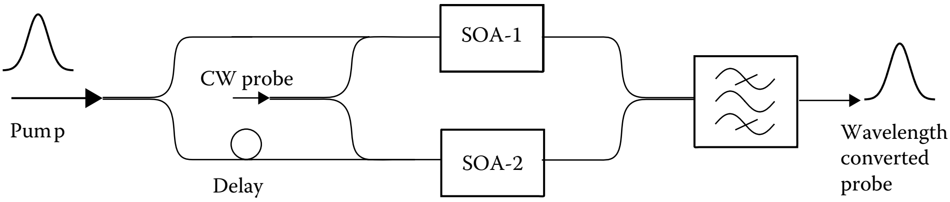


FIGURE 20.11 SOA Mach–Zehnder interferometer (MZI) wavelength converter.

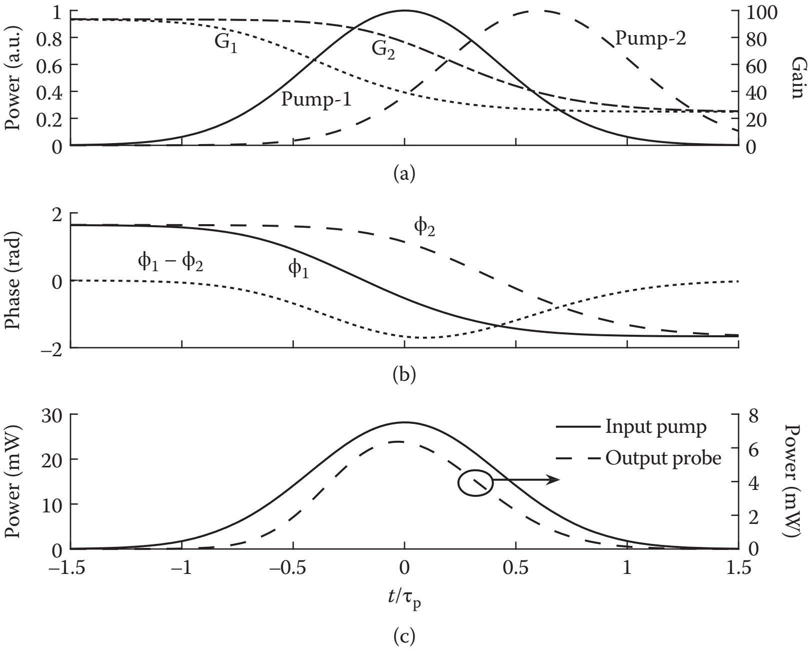


FIGURE 20.12 SOA MZI wavelength converter simulations. (a) SOA input pump powers and dynamic gains, (b) induced SOA probe phase shifts and difference, (c) comparison between the input pump and converted probe pulse shapes.

where P_{probe} is the input CW probe power, $G_1(t)$, $\varphi_1(t)$, $G_2(t)$, and $\varphi_2(t)$ are the gain and phase changes of the SOA in the upper and lower interferometer arms, respectively. The couplers have splitting and coupling ratios of 3 dB. The pulse amplification model described above can also be used to model XPM in the case where the input pump pulse width is much less than the intraband lifetime and assuming that the gain has fully recovered since the previous pulse. Including the effects of relatively slow interband processes requires a numerical solution. In the following simulations, an MZI with two identical SOAs having gains of 20 dB and $\alpha_H = 5$ is considered. The input pump is a zero-chirp Gaussian pulse with $E_{\text{in}}/E_{\text{sat}} = 0.03$. The CW input probe power is −10 dBm. The delay between the pump pulses entering SOA-1 and SOA-2 is equal to 0.6 of the input pump pulse width. These settings were chosen to obtain a converted probe pulse shape similar to the input pump pulse, as shown in Figure 20.12. The input pump pulse causes dynamic gain saturation and associated probe phase shifts. In the unsaturated state, the phase difference between the probe light waves in the interferometer arms is zero, leading to total destructive interference and thereby a zero output probe signal. The arrival of the pump pulses in the SOAs creates a window in which the magnitude of the phase difference increases, leading to constructive interference and thereby a nonzero converted probe signal pulse.

20.7 FWM Analytic Model

FWM is a *coherent* nonlinear process that can occur in an SOA between two copolarized input optical fields, a strong pump and a weaker probe at angular frequencies ω_0 and $\omega_0 - \Omega$, where Ω is the pump and probe detuning Agrawal (1988). Beating between the optical fields results in the establishment of gain and refractive index pulsations in the SOA at the detuning frequency. In most practical applications of FWM, in particular wavelength conversion, the detuning is much larger than the inverse of the interband carrier lifetime (typically in the nanosecond range) and the pulsations are caused by fast intraband processes, in particular CH and SHB. The gain and index pulsations result in nondegenerate FWM, in which energy from the pump and probe is coupled into a new copropagating field at frequency $\omega_0 + \Omega$, called the conjugate, so termed because its dynamic phase is the opposite of the signal phase.

SHB is caused by the injected pump signal creating a hole in the *intraband* carrier distribution. This modulates the occupation probability of carriers within a band leading to fast gain modulation. CH is caused by stimulated emission and free carrier absorption. Stimulated emission subtracts carriers that are cooler than average while free carrier absorption moves carriers to higher-energy levels in the band. The resulting increase in temperature decreases the gain. Two characteristic times are associated with CH: the carrier–phonon scattering time τ_1, which is the average time carriers require to cool down to the semiconductor lattice temperature, and the carrier–carrier scattering time τ_2, which is the average time taken by the carrier population to reach a heated equilibrium from the initial nonheated equilibrium. The latter time constant is also associated with SHB. The characteristic times are of the order of tens to hundreds of femtoseconds, so FWM-based wavelength converters can operate with detuning frequencies as high as hundreds of GHz. A comprehensive analysis of FWM in SOAs requires a numerical solution to a set of coupled-mode equations Agrawal (1988). The electric field in the SOA can be expressed as

$$E(z,t) = E_1 \exp\left\{j[k_1 z - (\omega_0 - \Omega)t]\right\} + E_0 \exp\left[j(k_0 z - \omega_0 t)\right] + E_2 \exp\left\{j[k_2 z - (\omega_0 + \Omega)t]\right\}, \quad (20.35)$$

where the copropagating and copolarized pump, probe, and conjugate field amplitudes are E_0, E_1, and E_2, respectively, with corresponding frequencies of ω_0, $\omega_0 - \Omega$, and $\omega_0 + \Omega$, and propagation coefficients k_0, k_1, and k_2. If it is assumed that the conjugate power is small relative to the copropagating input pump and probe powers, then an analytic model can be used to predict the SOA FWM characteristics (Mecozzi et al., 1995). In this model, the FWM conversion efficiency η defined as the ratio of the output conjugate to input probe powers is given by

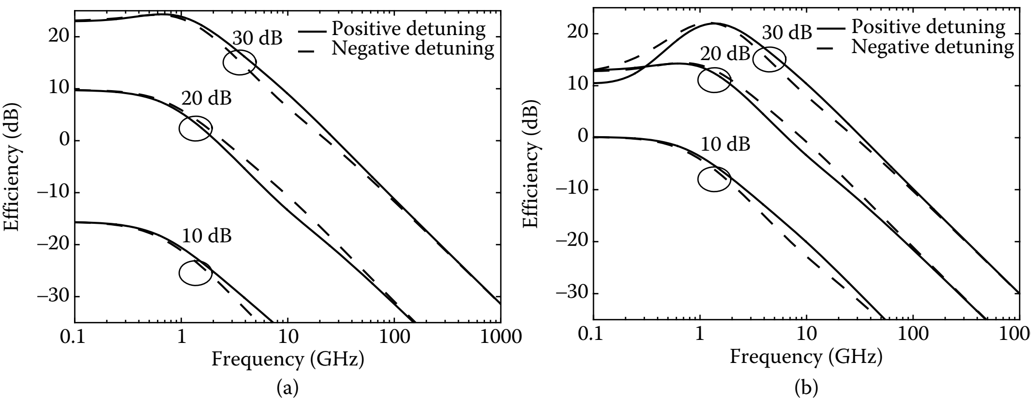
$$\eta = G\left|G_1\right|^2. \quad (20.36)$$

G is the saturated amplifier gain, which can be calculated using Equation 20.11, where p is taken to be the sum of the pump and probe powers. The unsaturated gain $G_0 = \exp[(g_0 - \alpha)L]$, where g_0 and α are the unsaturated gain and loss coefficients, respectively.

$$G_1 = \frac{-1 + j\alpha_H}{\alpha_H} \exp\left[-\frac{1}{2}\sigma F_{\mathrm{cd}}(\Omega)\right] \sin\left[\frac{\alpha_H}{2}\sigma F_{\mathrm{cd}}(\Omega)\right] - \frac{1}{2}\varepsilon_{\mathrm{sh}} P_{\mathrm{sat}} H_{\mathrm{sh}}(\Omega)\,\sigma F_{\mathrm{sh}} - \frac{1}{2}\varepsilon_{\mathrm{ch}} P_{\mathrm{sat}} H_{\mathrm{ch}}(\Omega)\,\sigma F_{\mathrm{ch}} \quad (20.37)$$

$$F_{\mathrm{cd}}(\Omega) = \frac{1}{1 - j\Omega\tau_{\mathrm{c}}\zeta}\left[\ln\left(\frac{1 + GP_{\mathrm{T}}/P_{\mathrm{sat}} - j\Omega\tau_{\mathrm{c}}}{1 + P_{\mathrm{T}}/P_{\mathrm{sat}} - j\Omega\tau_{\mathrm{c}}}\right) + \zeta \ln\left(\frac{G_0}{G}\right)\right] \quad (20.38)$$

$$F_{\mathrm{ch}} = -\frac{1}{\zeta}\left[\frac{P_T}{P_{\mathrm{sat}}}(G-1) - \ln\left(\frac{G_0}{G}\right)\right] \quad (20.39)$$



(a)

(b)

FIGURE 20.13 Four-wave mixing (FWM) efficiency versus absolute value of frequency detuning for input pump and probe powers of (a) 200 and 20 μm, and (b) 20 and 2 μm, respectively.

$$F_{\mathrm{sh}} = \ln\left(\frac{G_0}{G}\right) \tag{20.40}$$

$$S(0) = P_{\mathrm{sat}}\left(\frac{1-\zeta}{\zeta}\right)\frac{1-(G/G_0)^{\zeta}}{G-(G/G_0)^{\zeta}} \tag{20.41}$$

$$\sigma = \frac{P_0}{P_0+P_1}. \tag{20.42}$$

$\varepsilon_{\mathrm{ch}}$ and $\varepsilon_{\mathrm{sh}}$ are parameters characterizing the strengths of CH and SHB processes, respectively, and $\zeta = \alpha/g_0$. P_0 and P_1 are the input pump and probe powers, respectively, and the total input power $P_T = P_0 + P_1$. The Fourier transforms H_{ch} and H_{sh} of the nonlinear gain responses due to CH and SHB, respectively, are given by

$$H_{\mathrm{ch}}(\Omega) = \frac{1}{(1-j\Omega\tau_1)(1-j\Omega\tau_2)} \tag{20.43}$$

$$H_{\mathrm{sh}}(\Omega) = \frac{1}{1-j\Omega\tau_2}. \tag{20.44}$$

The spatial dependence of the saturated gain G can be determined using Equation 20.9, where p is taken to be the sum of the normalized pump and probe powers. Figure 20.13 shows the calculated efficiency for different values of amplifier unsaturated gain and input pump/probe powers with model parameters: $P_{\mathrm{sat}} = 10$ mW, $\alpha_H = 4.0$, $\tau_c = 0.25$ ns, $\tau_1 = 750$ ps, $\tau_2 = 15$ ps, $\varepsilon_{\mathrm{sh}} = 10.0$ W^{-1}, $\varepsilon_{\mathrm{ch}} = 2.5$ W^{-1}, and exp(αL) = 5 dB. The efficiency characteristics are asymmetric and show a more peaked structure as the amplifier gain is increased. The efficiency can be greater than unity for frequencies up to hundreds of GHz, i.e., frequency conversion with gain, so FWM-based SOA wavelength converters are of particular importance in wavelength division multiplexed systems with wide wavelength channel spacing.

20.8 Conclusion

SOAs have many uses in optical networks, not only as basic amplifiers but also as fundamental components in all-optical signal processing applications. Although SOAs have been investigated for many years, the development of new SOA structures and incorporation in PICs, advanced materials such as QDs, and exploitation of ultrafast nonlinearities make SOA research and development an exciting field of study.

The successful advancement of SOA technology requires the development of accurate models, which can be a significant challenge requiring a deep knowledge and understanding of many diverse areas such as photonic materials, device design, optoelectronic nonlinearities, noise modeling, and communication systems design. This chapter has given an overview of SOA basics and used relatively simple models to explain important SOA characteristics and applications. Further chapters consider more advanced models applied to particular types of SOAs.

References

Agrawal, A. P. 1988. Population pulsations and nondegenerate four-wave mixing in semiconductor lasers and amplifiers. *J. Opt. Soc. Am. B* 5:147–159.

Agrawal, A. P. and Olsson, A. 1989. Self-phase modulation and spectral broadening of optical pulses in semiconductor laser amplifiers. *IEEE J. Quantum Electron*. 25:2297–2306.

Akiyama, T., Sugawara, M. and Arakawa, Y. 2007. Quantum-dot semiconductor optical amplifiers. *Proc. IEEE* 95:1757–1766.

Asghari, M., White, I. H. and Penty, R. V. 1997. Wavelength conversion using semiconductor optical amplifiers. *J. Lightwave Technol.* 15: 1181–1190.

Baney, D. M., Gallion, P. and Tucker, R. S. 2000. Theory and measurement techniques for the noise figure of optical amplifiers. *Opt. Fiber Technol.* 6:122–154.

Ben-Ezra, Y., Haridim, M. and Lembrikov, B. I. 2005. Theoretical analysis of gain-recovery time and chirp in QD-SOA. *IEEE Photonics Technol. Lett*. 17:1803–1805.

Chuang, S. L. 2009. *Physics of Optoelectronic Devices*. New York, NY: Wiley.

Connelly, M. J. 2001. Wide-band semiconductor optical amplifier steady-state numerical model. *IEEE J. Quant. Electron*. 37:439–447.

Connelly, M. J. 2002a. *Semiconductor Optical Amplifiers*. Boston, MA: Springer.

Connelly, M. J. 2002b. Wideband dynamic numerical model of a tapered buried ridge stripe semiconductor optical amplifier gate. *IEE Proc. Circ. Dev. Syst*. 149:173–178.

Connelly, M. J. 2007. Wide-band steady-state numerical model and parameter extraction of a tensile-strained bulk semiconductor optical amplifier. *IEEE J. Quant. Electron*. 43:47–56.

Durhuus, T., Mikkelsen, B. and Stubkjaer, K. E. 1992. Detailed dynamic model for semiconductor optical amplifiers and their crosstalk and intermodulation distortion. *J. Lightwave Technol.* 10:1056–1065.

Dutta, N. K. and Wang, Q. 2013. *Semiconductor Optical Amplifiers*. Singapore: World Scientific Publishing.

Fukuda, M. 1991. *Reliability and Degradation of Semiconductor Lasers and LEDs*. Boston, MA: Artech House.

Henry, C. H. 1982. Theory of the linewidth of semiconductor lasers. *IEEE J. Quant. Electron*. 18:259–264.

Hong, M. Y., Chang, Y. H., Dienes, A., Heritage, J. P. and Delfyett, P. J. 1994. Subpicosecond pulse amplification in semiconductor laser amplifiers: Theory and experiment. *IEEE J. Quant. Electron*. 30:1122–1131.

Jones, G. and O'Reilly, E. P. 1993. Improved performance of long-wavelength strained bulk-like semiconductor lasers. *IEEE J. Quant. Electron*. 29:1344–1345.

Kweon, G. 2002. Noise figure of optical amplifiers. *J. Korean Phys. Soc*. 41:617–628.

Lee, W., Park, M. Y., Cho, S. H., Lee, J., Kim, B. W., Jeong, G. and Kim, B. W. 2005. Bidirectional WDM-PON based on gain-saturated reflective semiconductor optical amplifiers. *IEEE Photon. Technol. Lett*. 17:2460–2462.

Lelarge, F., Dagens, B., Renaudier, J., Brenot, R., Accard, A., van Dijk, F., Make, D., Le Gouezigou, O., Provost, J. G., Poingt, F. and Landreau, J. 2007. Recent advances on InAs/InP quantum dash based semiconductor lasers and optical amplifiers operating at 1.55 μm. *IEEE J. Sel. Top. Quantum Electron*. 13:111–124.

Mecozzi, A. and Mork, J. 1997. Saturation induced by picosecond pulses in semiconductor optical amplifiers. *J. Opt. Soc. Am. B* 14:761–770.

Mecozzi, A., Scotti, S., D'Ottavi, A., Iannone, E. and Spano, P. 1995. Four-wave mixing in traveling-wave semiconductor optical amplifiers. *IEEE J. Quant. Electron*. 31:689–699.

Obermann, K., Kindt, S., Breuer, D. and Petermann, K. 1998. Performance analysis of wavelength converters based on cross-gain modulation in semiconductor-optical amplifiers. *J. Lightwave Technol*. 16:78–85.

Occhi, L., Schares, L. and Guekos, G. 2003. Phase modeling based on the α-factor in bulk semiconductor optical amplifiers. *IEEE J. Sel. Top. Quantum Electron*. 9:788–797.

Olsson, N. A. 1989. Lightwave systems with optical amplifiers. *J. Lightwave Technol*. 7:1071–1082.

Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery, B. P. 2007. *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. New York, NY: Cambridge University Press.

Suematsu, Y. and Adams, A. R. 1994. *Handbook of Semiconductor Lasers and Photonic Integrated Circuits*. London: Chapman & Hall.

21

Traveling-Wave and Reflective Semiconductor Optical Amplifiers

21.1 Introduction 631
21.2 General Structure and Operation Principles of TW- and R-SOA 632
21.3 Modeling of Material and Structural Parameters 633
Material Gain, Radiative Spontaneous Recombination, and Refractive Index Variation • Waveguide Design and Polarization Insensitivity
21.4 Rate Equations 642
Basic Rate Equations • Amplified Spontaneous Emission • Extended Rate Equation Model
21.5 Overview of Steady-State and Dynamic TW- and R-SOA Models 656
Steady-State Models • Dynamic Models • Case Study: Steady-State Wideband Self-Consistent Numerical Model • Case Study: Steady-State Semianalytical Model • Case Study: Dynamic Propagation Model • Case Study: TMW Small-Signal Model • Case Study: Small-Signal Model for Transparent SOA
21.6 Conclusion 688

Angelina Totović
and
Dejan Gvozdić

21.1 Introduction

Eras in modern human history have often been labeled by their most significant and widely adopted technological achievements. It is no wonder that the last several decades proudly bear the name "The Information Age." Information, in its many forms, is an invaluable asset and its dissemination, collection, storage, and analysis have been the primary focus of researchers in numerous fields. Ever-increasing hunger for information exchange has driven supporting technologies to new levels. The transition from electric to optical domain has proved to be one of the most significant milestones of modern communications. According to TeleGeography and Huawei Marine Networks, as of early 2017, it has been estimated that close to 1.1 million kilometers of submarine optical cables exist, each bearing several optical fibers, and approximately 99% of all international telecommunications traffic is carried over this infrastructure. Starting from the overloaded backbone links, optical technology was gradually implemented in the lower levels of network hierarchy, with the ultimate goal of reaching individual users in the foreseeable future. Decreasing attenuation, eliminating electromagnetic interference, reducing the power consumption, and diminishing the possibility of eavesdropping are just some of the advantages the new technology has realized (Ramaswami, 2002). With advantages also come challenges, as we are gravitating toward increased efficiency, which includes increased capacity over larger distances with decreased signal degradation and power consumption, and aggregation of fixed and mobile networks into a single network architecture (Carapellese et al., 2014). Aside from information exchange, information processing and storage in the optical domain remain in focus as we pursue all-optical solutions to mitigate unnecessary electro-optic and opto-electric conversion (Ramaswami, 2002; Sygletos et al., 2008). Many steps have already been taken, both in the physical layer, where new materials and production technologies significantly influenced performance, and in the higher levels of abstraction, where new modulation formats and protocols have been introduced (Ramaswami, 2002; Gladisch et al., 2006; Sygletos et al., 2008). Devices that are simple, but can perform multiple functions, have always been highly regarded. One example is the semiconductor optical amplifier (SOA)—a photonic device that under adequate conditions can amplify the input optical signal. SOAs have been extensively studied and perfected over the past three decades. Similar to laser diodes, they are compact and easy to produce and integrate with other photonic devices on a single chip (Xing et al., 2004). Wide amplification spectrum makes them suitable for use in wavelength-division multiplexed (WDM) optical networks (Mecozzi and Wiesenfeld, 2001; Zimmerman and Spiekman, 2004; Kani, 2010). SOAs consume relatively little power (Xing et al., 2004; Sygletos et al., 2008; Koenig et al., 2014), they are transparent to the optical signal modulation format (Schmuck et al., 2013; Koenig et al., 2014), and they can be designed to be polarization insensitive (Mathur and Dapkus, 1992; Carlo et al., 1998; Michie et al., 2006).

Aside from being used as standalone amplifiers, SOAs can be used for numerous other important functions (Olsson, 1989; Mecozzi and Wiesenfeld, 2001), such as switching (Stabile and Williams, 2011; Figueiredo et al., 2015), modulation/remodulation (Totović et al., 2011; Pham, 2014), wavelength conversion (Joergensen et al., 1997; Dailey and Koch, 2009), signal regeneration (Sygletos et al., 2008), all-optical flip-flops (Pitris et al., 2015), and all-optical random access memory (RAM) cells (Vyrsokinos et al., 2014), to name just a few.

21.2 General Structure and Operation Principles of TW- and R-SOA

Although SOAs come in many different forms, they all share several common features. Each SOA consists of an active region and a cladding and can have either highly reflective or antireflective facets, or a combination of the two, as shown in Figure 21.1. The two most commonly used SOA structures are the traveling-wave (TW) SOA and the reflective (R) SOA.

In the TW-SOA, both facets are coated with an antireflective layer, and the optical signal travels only once through the active region. In other words, input and output ports of TW-SOA are on opposite sides. This is preferable configuration for using SOA as a standalone amplifier, whether as a booster, in-line, or preamplifier (Olsson, 1989; Xing et al., 2004). On the other hand, the R-SOA has one antireflective and one highly reflective facet, which causes the signal to be reflected and travel through the active region twice. Both input and output ports are on the same side, namely at the antireflective facet. The rear facet can be made semitransparent, rather than highly reflective, which is usually the case if the interface between the cleaved semiconductor facet and air is responsible for reflection. This may be a useful feature, since now two output signals can be used, one at the rear and one at the front facet. An SOA with both facets highly reflective is the Fabry–Pérot (FP) SOA (Adams et al., 1985; Thylén, 1988). Due to the feedback loop provided by the highly reflective facets, the FP-SOA closely resembles a laser diode, although it is not typically used to generate coherent signal by itself (Yariv and Yeh, 2007). Nevertheless, its spectrum does exhibit pronounced resonances and antiresonances, unlike TW- and R-SOAs, where ripples in spectrum are much more subtle and are the result of residual reflectivity of the antireflective facets (Yariv and Yeh, 2007).

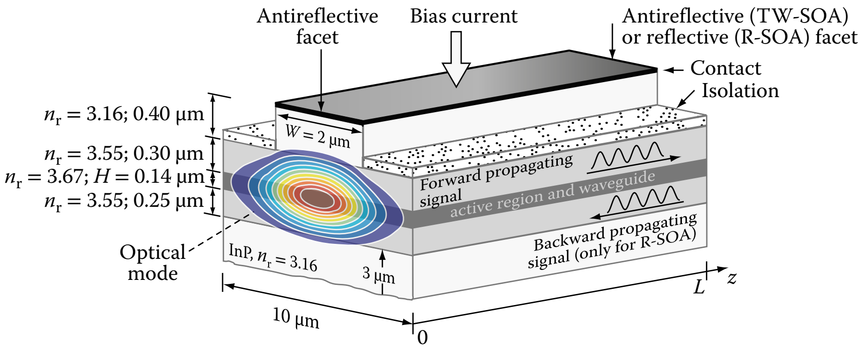

Antireflective facet
Bias current
Antireflective (TW-SOA) or reflective (R-SOA) facet
Contact
Isolation
$n_r = 3.16$; 0.40 μm
$n_r = 3.55$; 0.30 μm
$n_r = 3.67$; $H = 0.14$ μm
$n_r = 3.55$; 0.25 μm
$W = 2$ μm
Forward propagating signal
active region and waveguide
Backward propagating signal (only for R-SOA)
Optical mode
InP, $n_r = 3.16$
3 μm
10 μm
0
L
z

FIGURE 21.1 Schematic representation of the semiconductor optical amplifier (SOA), with relevant geometrical parameters and signals denoted. The active region is based on the unstrained bulk $In_{0.53}Ga_{0.47}As$ and designed to be polarization insensitive. (R-SOA, reflective SOA; TW-SOA, traveling-wave SOA.)

The active region is responsible for the signal amplification through the mechanism of stimulated emission, where the electric bias current is used to achieve population inversion. The purpose of the built-in waveguide is to confine the optical signal that propagates along the SOA and prevent its leaking into the cladding. The output signal of any SOA should ideally be an amplified replica of the input signal, although, in practice, a certain amount of noise is always superimposed, due to the always-present spontaneous emission. Aside from the photons initially generated through the process of spontaneous emission, noise propagation leads to its amplification, which finally results in amplified spontaneous emission (ASE) noise. Although the noise is typically regarded as the unwanted by-product of the device operation, there are numerous examples of using the noise either to generate or deliberately modify the optical signal (Yamatoya and Koyama, 2004; de Oliveira Ribeiro et al., 2005; Kang et al., 2006; Gebrewold et al., 2015), or enhance the overall SOA performance (Valiente et al., 1996).

SOAs can also be classified based on the material used for the active region, where each material has advantages and disadvantages. The simplest choice is bulk semiconductor (Michie et al., 2006; Connelly, 2007; Mazzucato et al., 2015; Totović et al., 2015). However, due to the several drawbacks of bulk SOA, such as high recovery time (Zilkie et al., 2007), and broad ASE spectrum (Totović et al., 2013), other solutions have emerged.

Depending on the desired improvement, different materials with a higher degree of confinement can be used (Zilkie et al., 2007), starting from multiple quantum well (MQW) (Nagarajan et al., 1992; Keating et al., 1999; Qin et al., 2012), and continuing to quantum dash (QDash) (Gioannini, 2004; Reithmaier et al., 2005; Qasaimesh, 2013) and quantum dot (QD) materials (Berg et al., 2001; Qasaimesh, 2003; Kim et al., 2009). SOAs with the active region based on these materials usually exhibit better performance in one or more aspects, at the expense of increased complexity and production cost. For example, QD SOA has fast recovery time, which is crucial for ultrafast signal processing, but due to the dot size distribution, it has moderate material gain and wide ASE spectrum, with full-width at half-maximum (FWHM) reaching double the values reported for MQW and QDash SOAs (Zilkie et al., 2007). On the other hand, QDash SOA has high differential gain and requires much lower driving currents in comparison with other active region types and outperforms bulk SOA in terms of time needed for gain recovery (Reithmaier et al., 2005).

21.3 Modeling of Material and Structural Parameters

Before proceeding to the SOA analysis, it is necessary to understand the physical principles responsible for its operation and quantify them through the parameters typically used for description of photon–electron interaction in the active media. In addition, the waveguide should be optimized to provide polarization-insensitive operation, and the parameters describing it should also be studied.

21.3.1 Material Gain, Radiative Spontaneous Recombination, and Refractive Index Variation

Some of the most important optical properties of an SOA's active layer, determining its functional performance and parameters, are material gain, radiative spontaneous recombination (spontaneous emission) rate, and refractive index. Material gain, or gain per unit length, together with several other structural parameters, is responsible for defining the transmission (device) gain, i.e., the ratio of powers of the output and input optical signals, which is a figure of merit for the amplifier's capability to amplify the input optical signal by stimulated emission of photons (Connelly, 2002). Radiative spontaneous recombination rate and its spectral distribution represent the defining parameters that sculpt ASE spectrum and the SOA's noise (Silver et al., 2000; Connelly, 2002), whereas the refractive index variation directly affects the SOA linewidth enhancement factor (LEF) (Wang et al., 2007), further influencing the SOA's performance in cross-phase modulation (XPM) systems and optical switching.

In SOAs, stimulated emission has interband character, meaning that it occurs between the conduction band (CB) and the valence band (VB) of the semiconductor. Since these electronic bands consist of dispersed energy levels with different density per unit energy and occupation probability, the material gain significantly differs from gain in other solid-state or gas optically active media, which have a few (usually three or four) energy levels involved in the process of stimulated emission and its provisioning. The spreading of energy levels in the bands leads to a broad spectrum of material gain, which increases with the increase of injected carrier density or, equivalently, the current density. The radiative stimulated recombination rate is generally proportional to material gain (Chuang, 1995; Coldren et al., 2012), similar to the radiative spontaneous recombination rate (Chuang, 1995; Coldren et al., 2012). The material gain, as well as the radiative spontaneous recombination rate and the refractive index change, is, generally speaking, the function of three quantities: the joint density of states, electron/hole occupation probability, and a constant corresponding to a single optical transition between the two energy levels, one in CB and the other in VB.

The model of material gain, and the other two parameters, depends on the degree of carrier confinement in the active region of the SOA and the semiconductor strain introduced by the confinement. However, the models may differ with respect to degree of accuracy, i.e., approximations adopted in the model. The common framework for calculating optical properties of semiconductor bulk and quantum confined structures is the envelope function approximation (Meney et al., 1994). Within this framework, a relatively simple, so-called "two-band" model is used for material gain calculation (Chuang, 1995). It is based on the assumption of parabolic band dependence $E(k)$, where E is the energy and k is the wave vector of electron (hole) in the conduction (valence) band. Each band is treated separately by the effective mass Schrödinger equation (Chuang, 1995). More sophisticated multiband envelope function models account for the VB mixing effect, which essentially means that the character of sub-bands is a mixture of heavy hole (HH), light hole (LH), and split-off (SO) band. This effect occurs as a result of quantum confinement and becomes more prominent as the degree of confinement increases.

The most common multiband envelope function models are the Luttinger–Kohn (LK) 4×4 (Chuang, 1991) and 6×6 models (Chao and Chuang, 1992). The former model (LK 4×4) accounts for HH and LH band mixing, whereas the latter (LK 6×6) includes SO band, in addition to HH and LH bands. A more advanced multiband approach is based on the 8×8 $\mathbf{k}\cdot\mathbf{p}$ model (Meney et al., 1994; Liu and Chuang, 2002; Gvozdić and Ekenberg, 2006), which comprises all three VBs and the CB. In addition to nonparabolicity of VB, this model accounts for nonparabolicity of CB as well. However, its fundamental drawback is the occurrence of the spurious solutions, which can be avoided to some extent by various techniques (Foreman, 1997; Cartoixa et al., 2003; Kolokolov et al., 2003; Veprek et al., 2007).

21.3.1.1 Two-Band Model

Although relatively rough, the two-band model might be used in case of bulk, and even other types of quantum confined structures, but with modest accuracy. It is suitable for the unstrained bulk active region since the optical matrix element is isotropic in this case (Chuang, 1995), which means that material gain is the same for any direction of the amplified light polarization. Contrarily, in case of quantum confined semiconductor structures, material gain is polarization dependent. Nevertheless, the two-band model can still be implemented even in the case of quantum wells. If the optical dipole matrix element is averaged over the azimuthal angle corresponding to the plane of the quantum well, it is possible to derive its angular dependence with respect to the confinement direction of the quantum well (Chuang, 1995).

Here, the detailed expressions for material gain g, spontaneous emission rate per unit energy r_{sp} and refractive index variation Δn_{r} are presented for the unstrained bulk active region, according to the two-band model (Chuang, 1995). They are all functions of photon energy $\hbar\omega$ and carrier density n, which is included in expressions through the carrier-dependent quasi-Fermi levels:

$$g\left(n,\hbar\omega\right)=C_{\mathrm{g}}\frac{M_{\mathrm{b}}^{2}}{4\pi^{2}}\sum_{i=\mathrm{HH,LH,SO}}\left(\frac{2\mu_{i}}{\hbar^{2}}\right)^{3/2}\int\limits_{E_{\mathrm{g},i}}^{\infty}\frac{\gamma}{\pi}\frac{\sqrt{x-E_{\mathrm{g},i}}}{(x-\hbar\omega)^{2}+\gamma^{2}}\left(f_{\mathrm{FD}}^{\mathrm{c}}-f_{\mathrm{FD}}^{\mathrm{v}}\right)\mathrm{d}x, \tag{21.1}$$

$$r_{\mathrm{sp}}\left(n,\hbar\omega\right)=\frac{8\pi n_{\mathrm{r}}^{2}\left(\hbar\omega\right)^{2}}{h^{3}c^{2}}C_{\mathrm{g}}\frac{M_{\mathrm{b}}^{2}}{4\pi^{2}}\sum_{i=\mathrm{HH,LH,SO}}\left(\frac{2\mu_{i}}{\hbar^{2}}\right)^{3/2}\int\limits_{E_{\mathrm{g},i}}^{\infty}\frac{\gamma}{\pi}\frac{\sqrt{x-E_{\mathrm{g},i}}}{(x-\hbar\omega)^{2}+\gamma^{2}}f_{\mathrm{FD}}^{\mathrm{c}}\cdot\left(1-f_{\mathrm{FD}}^{\mathrm{v}}\right)\mathrm{d}x, \tag{21.2}$$

$$\Delta n_{\mathrm{r}}\left(n,\hbar\omega\right)=\frac{q^{2}\hbar^{2}}{n_{\mathrm{r}}\varepsilon_{0}m_{0}^{2}}\frac{M_{\mathrm{b}}^{2}}{2\pi^{2}}\sum_{i=\mathrm{HH,LH,SO}}\left(\frac{2\mu_{i}}{\hbar^{2}}\right)^{3/2}\int\limits_{E_{\mathrm{g},i}}^{\infty}\frac{\left(x-\hbar\omega\right)\sqrt{x-E_{\mathrm{g},i}}}{x\left(x+\hbar\omega\right)\left[\left(x-\hbar\omega\right)^{2}+\gamma^{2}\right]}\left(f_{\mathrm{FD}}^{\mathrm{v}}-f_{\mathrm{FD}}^{\mathrm{c}}\right)\mathrm{d}x. \tag{21.3}$$

In previous equations, $C_{\mathrm{g}}=\pi q^{2}/(n_{\mathrm{r}}c\varepsilon_{0}m_{0}^{2}\omega)$, where q is the absolute electron charge, m_0 is the free-electron mass, n_{r} is the refractive index of the active region, c is the speed of light in vacuum, ε_0 is the vacuum permittivity, $\omega=2\pi\nu$ is the angular photon frequency, with ν being its frequency, and $\hbar=h/(2\pi)$ is the reduced Planck's constant, with h being the Planck's constant. In Equations 21.1 through 21.3, $M_{\mathrm{b}}^{2}=m_{0}E_{\mathrm{p}}/6$ stands for the bulk momentum matrix element squared, where E_{p} is the interband matrix element, $\mu_{i}=(m_{\mathrm{e}}^{-1}+m_{i}^{-1})^{-1}$ and $E_{\mathrm{g},i}$ are the reduced effective mass and the energy gap between the CB bottom E_{c} and the VB top $E_{\mathrm{v},i}$, respectively, where i stands for the HH, LH, and SO band. With $f_{\mathrm{FD}}^{\mathrm{c}}(E)=f_{\mathrm{FD}}^{\mathrm{c}}[E_{\mathrm{c}}+\mu_{i}/m_{\mathrm{e}}\cdot(x-E_{\mathrm{g},i})]$ and $f_{\mathrm{FD}}^{\mathrm{v}}(E)=f_{\mathrm{FD}}^{\mathrm{v}}[E_{\mathrm{v},i}-\mu_{i}/m_{i}\cdot(x-E_{\mathrm{g,i}})]$ we denote the Fermi–Dirac distribution $f_{\mathrm{FD}}^{(j)}(E)=\{1+\exp[(E-E_{\mathrm{f}}^{(j)}/k_{\mathrm{B}}T)]\}^{-1}$ for the CB ($j=c$) and the VB ($j=v$), characterized by the corresponding quasi-Fermi levels $E_{\mathrm{f}}^{(j)}$, where k_{B} is the Boltzmann constant and T is the temperature. The half linewidth of the Lorentzian function is denoted by γ.

21.3.1.2 Multiband Model

The LK 4 × 4 and 6 × 6 Hamiltonians are usually implemented in the analysis of quantum confined structures. However, this approach can be very useful in the calculation of gain of strained bulk material (Connelly, 2007), which is polarization dependent in this case. As shown by Connelly (2007), and Mazzucato et al. (2015), eigenvectors of multiband Hamiltonian can be used to find the components of the optical matrix elements along the directions corresponding to polarizations of incoming light.

Totović et al. (2013) present an analysis of the material gain, radiative spontaneous recombination rate spectrum, and refractive index variation in MQWs based on the 8 × 8 $\mathbf{k}\cdot\mathbf{p}$ method (Liu and Chuang, 2002; Gvozdić and Ekenberg, 2006). The analysis is done for an MQW active region consisting of six coupled 0.13% tensile-strained $In_{0.516}Ga_{0.484}As$ quantum wells, which are strain-compensated by the $In_{0.9}Ga_{0.1}As_{0.3}P_{0.7}$ 0.26% compressively strained barriers. The well thickness is $L_w = 19$ nm, while the barriers between the wells are $L_b = 10$ nm thick, making the total thickness of the structure $L_z = 6L_w + 5L_b = 164$ nm. The electronic band structure calculation accounts for Burt–Foreman Hermitianization (Foreman, 1993) and biaxial strain generated by lattice-mismatched growth of the well-barrier layers. By using an appropriate basis set, the 8×8 Hamiltonian is decoupled into two 4×4 Hamiltonians (H_U and H_L) corresponding to the upper (U) and lower (L) block Hamiltonians (Liu and Chuang, 2002). In calculating the band structure, the finite difference method (FDM) is employed. For determining g, r_{sp}, and Δn_r, the following relations are used (Chuang, 1995; Liu et al., 2001; Liu and Chuang, 2002):

$$g(n,\hbar\omega) = \frac{q^2\pi}{n_r c\varepsilon_0 m_0^2 \omega L_z} \sum_{\eta=U,L}\sum_{\sigma=U,L}\sum_{n,m}\int \left|\mathbf{e}\cdot\mathbf{M}_{nm}^{\eta\sigma}\right|^2 \frac{\gamma}{\pi} \frac{f_{\eta n}^{c} - f_{\sigma m}^{v}}{\left(E_{nm}^{\eta\sigma} - \hbar\omega\right)^2 + \gamma^2} \frac{k_t \mathrm{d}k_t}{2\pi}, \tag{21.4}$$

$$r_{sp}(n,\hbar\omega) = \frac{q^2 n_r \omega}{\pi\hbar c^3 \varepsilon_0 m_0^2 L_z} \sum_{\eta=U,L}\sum_{\sigma=U,L}\sum_{n,m}\int \left|M_{sp}\right|^2 \frac{\gamma}{\pi} \frac{f_{\eta n}^{c}\cdot\left(1 - f_{\sigma m}^{v}\right)}{\left(E_{nm}^{\eta\sigma} - \hbar\omega\right)^2 + \gamma^2} \frac{k_t \mathrm{d}k_t}{2\pi}, \tag{21.5}$$

$$\Delta n_r(n,\hbar\omega) = \frac{q^2\hbar^2}{n_r \varepsilon_0 m_0^2 L_z} \sum_{\eta=U,L}\sum_{\sigma=U,L}\sum_{n,m}\int \left|\mathbf{e}\cdot\mathbf{M}_{nm}^{\eta\sigma}\right|^2 \frac{E_{nm}^{\eta\sigma} - \hbar\omega}{E_{nm}^{\eta\sigma}\cdot\left(E_{nm}^{\eta\sigma} + \hbar\omega\right)} \frac{f_{\eta n}^{v} - f_{\sigma m}^{c}}{\left(E_{nm}^{\eta\sigma} - \hbar\omega\right)^2 + \gamma^2} \frac{k_t \mathrm{d}k_t}{2\pi}, \tag{21.6}$$

where η and σ stand for one of the two 4 × 4 Hamiltonians (H_U or H_L) and corresponding to the nth conduction and the mth valence sub-band, respectively, and k_t is the in-plane wave vector, while $f_{\eta n}^{c}(k_t)$ and $f_{\sigma m}^{v}(k_t)$ are the Fermi–Dirac distributions for CB and VB, respectively. In these relations, $E_{nm}^{\eta\sigma} = E_{nm}^{\eta\sigma}(k_t)$ is the energy difference between the eigenenergy of the nth conduction sub-band corresponding to the η-Hamiltonian and mth valence sub-band of the σ-Hamiltonian, whereas $\mathbf{M}_{nm}^{\eta\sigma} = \mathbf{M}_{nm}^{\eta\sigma}(k_t)$ is the corresponding momentum matrix element vector (Liu and Chuang, 2002). Depending on the unit polarization vector **e**, the components of the $\mathbf{M}_{nm}^{\eta\sigma}$ vector correspond to the transverse-electric (TE) mode ($\mathbf{e} = \mathbf{x}$ or $\mathbf{z}$) or the transverse-magnetic (TM) mode ($\mathbf{e} = \mathbf{y}$). In the calculation of r_{sp}, the average matrix element for the three polarization directions is used, i.e., $|M_{sp}(k_t)|^2 = (2|\mathbf{M}_{TE}(k_t)|^2 + |\mathbf{M}_{TM}(k_t)|^2)/3$.

The spectral dependencies of material gain, radiative spontaneous recombination rate, and refractive index variation are shown in Figure 21.2. In Figure 21.2a through c, the results are given for the bulk unstrained $In_{0.53}Ga_{0.47}As$ active region, calculated using the two-band model and the parameters listed in Table 21.1. In Figure 21.2d through f, the same quantities, calculated using the multiband model, are given for an MQW active region consisting of six coupled 0.13% tensile-strained $In_{0.516}Ga_{0.484}As$ quantum wells, which are strain-compensated by the $In_{0.9}Ga_{0.1}As_{0.3}P_{0.7}$ 0.26% compressively strained barriers. The spectral dependencies for the MQW active region are given only for TE polarization and are determined based on the parameters given in Table 21.2.

In addition to bulk and QWs, the SOA's active region often comprises self-assembled QDs. The calculation of the optical properties of QDs can be based on sophisticated **k·p** models, such as 8 × 8 **k·p** model (Grundmann et al., 1995; Stier et al., 1999). However, due to the statistical nature of the QD size, very precise calculation is usually meaningless. Therefore, some simple methods can also be sufficiently effective as in the two-band model (Kim and Chuang, 2006; Kim et al., 2008).

21.3.2 Waveguide Design and Polarization Insensitivity

In addition to high transmission gain, high optical output power, low-noise, and low-energy consumption, in many SOA applications polarization insensitivity of modal gain is required. This demand can be difficult to fulfill since polarization dependence of modal gain is caused by SOA's waveguide geometry, through confinement factor, and material choice, through the optical transitions. Therefore, careful analysis and design of the waveguide structure and material gain are needed (Labukhin and Li, 2006). Numerical

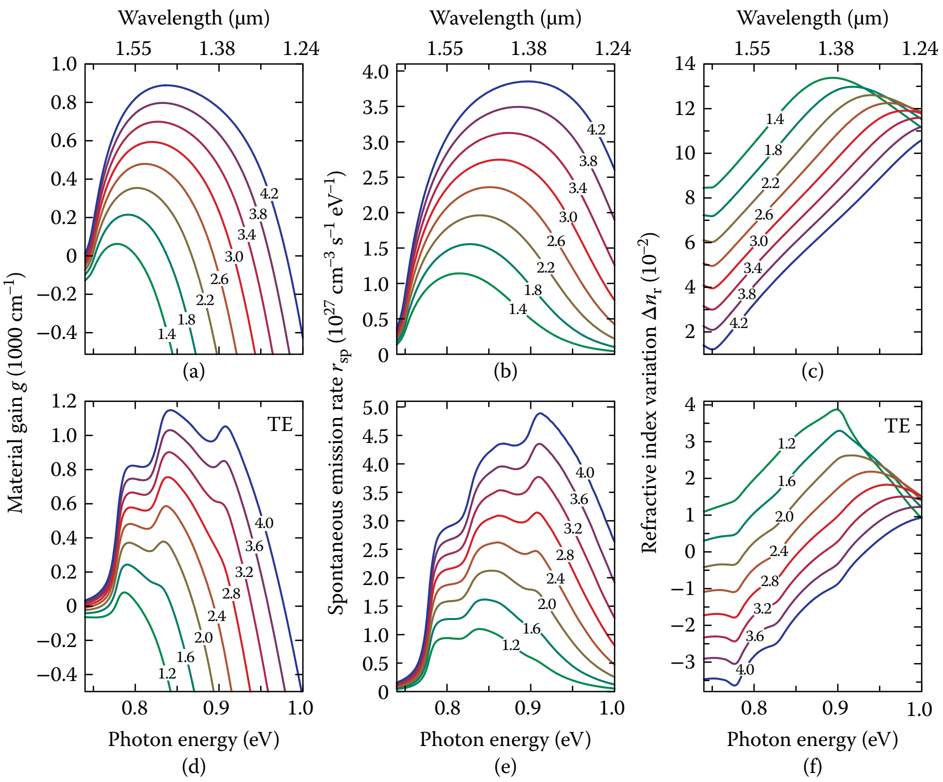


FIGURE 21.2 Spectral dependency of (a, d) material gain, (b, e) radiative spontaneous recombination rate per unit energy, and (c, f) refractive index variation for different carrier densities n (10^{18} cm^{-3}) for (a) through (c) bulk and (d) through (f) multiple quantum well (MQW) active region.

TABLE 21.1 Material Parameters Used for Band Calculation for Bulk SOA

Symbol	Quantity	$In_{0.53}Ga_{0.47}As$	$In_{0.76}Ga_{0.24}As_{0.52}P_{0.48}$
E_g	Energy gap CB–VB (HH, LH)	748 meV	1001 meV
$E_{g,SO}$	Energy gap CB–VB (SO)	1115 meV	1246 meV
Δ_{SO}	SO energy	367 meV	245.3 meV
E_p	Interband matrix element	24.93 eV	23.26 eV
$m_e{}^*$	Electron effective mass	0.0389 m_0	0.0541 m_0
$m_{HH}{}^*$	HH effective mass	0.3410 m_0	0.4069 m_0
$m_{LH}{}^*$	LH effective mass	0.0567 m_0	0.0889 m_0
$m_{SO}{}^*$	SO effective mass	0.1393 m_0	0.1709 m_0

SOA, semiconductor optical amplifier; LH, light hole; HH, heavy hole; SO, split-off; CB, conduction band; VB, valence band.

modeling of the active region's susceptibility, the real and imaginary parts of which correspond to refractive index and material gain, in combination with the calculation of the waveguide propagation constant and confinement factor is an inevitable and important step in the quest for the optimal, polarization-insensitive design of SOA. Waveguide polarization sensitivity comes from significant disproportion of its cross-section dimensions, where the width is usually one order of magnitude larger than its thickness, raising the difference in TE and TM wave propagation and the corresponding confinement factors. Moreover, as shown in

TABLE 21.2 Material Parameters Used for Band Calculation for MQW SOA

Symbol	Quantity	$In_{0.516}Ga_{0.484}As$	$In_{0.9}Ga_{0.1}As_{0.3}P_{0.7}$
E_g	Energy gap	765.2 meV	1107.3 meV
$E_{v,av}$	Average VB position	−6.6689 eV	−6.8986 eV
Δ_{SO}	SO energy	366.3 meV	189.2 meV
E_p	Interband matrix element	25.03 eV	21.91 eV
m_e^*	Electron effective mass	0.0395 m_0	0.0617 m_0
γ_1	Luttinger parameters	10.1534	6.0906
γ_2		3.6114	1.9545
γ_3		4.5580	2.6785
c_{11}	Elastic stiffness constants	1020.7 GPa	996.79 GPa
c_{12}		507.48 GPa	536.03 GPa
a_c	CB deformation potential	−6.091 eV	−5.941 eV
a_v	VB deformation potential	−1.077 eV	−0.802 eV
b	Shear deformation	−1.897 eV	−1.918 eV

SOA, semiconductor optical amplifier; MQW, multiple quantum well; SO, split-off; CB, conduction band; VB, valence band.

Section 21.3.1, quantum-confined and stressed bulk semiconductors in active region exhibit polarization-dependent dipole optical matrix elements, leading to additional variation of TE and TM modes. An unstrained bulk material is polarization isotropic and in this case square waveguide cross-section may provide SOA polarization insensitivity (Connelly, 2001; Michie et al., 2006). Unfortunately, such waveguide design is technologically very demanding (Connelly, 2001; Michie et al., 2006). Moreover, it leads to large far-field divergence and consequently to poor coupling efficiency from the SOA to optical fiber (Connelly, 2002). This can be overcome by tapering the active waveguide near the amplifier facets (Connelly, 2002; Michie et al., 2006).

Nevertheless, there are other solutions for polarization-insensitive operation of SOA proposed in the literature. The majority of these proposals concern MQW SOA and rely on any of the following concepts: MQWs with tensile barriers (Magari et al., 1991, 1994), tensile-strained QWs (Ito et al., 1998; Carrère et al., 2010), tensile-strained QWs with compressive barriers (Godefroy et al., 1995; Zhang and Ruden, 1999), alternation of tensile and compressive QWs (Mathur and Dapkus, 1992; Joma et al., 1993; Tiemeijer et al., 1993; Tishinin et al., 1997; Silver et al., 2000), and the delta-strained QW (Carlo et al., 1998; Cho and Choi, 2001). The concept based on MQWs with tensile barriers (Magari et al., 1991, 1994) essentially relies on considerable increase of barrier refractive index for TM mode in comparison with TE mode, as a consequence of LH band contribution, which modifies the bandgap and shifts the wavelength toward the longer wavelength side due to the tensile strain. The approach based on tensile-strained MQWs with compressive barriers, proposed by Zhang and Ruden (1999), involves the barrier width optimization such that the HH sub-bands are grouped tightly and the LH sub-bands are widely separated in energy. The uppermost valence sub-bands, which have large occupation probability and strongly contribute to the gain, consist of a single LH sub-band and a group of coupled HH sub-bands, giving rise to balanced gains for the TE and TM polarizations. The method of alternation of separate tensile and compressive QWs (Silver et al., 2000) requires thick QWs (around 14 nm) and with high strain (around −0.8%). Due to the thick tensile layers, the difference in density of states of the two types of wells leads to significant charge redistribution within the active region. The space charge modifies the band profile and forces holes to move from the tensile well to the compressive well, increasing TE gain. In order to compensate for hole redistribution, the number of tensile wells can be increased so that they outnumber the compressive wells.

Otherwise, the thickness of the compressive wells can be increased to minimize the charge redistribution effects.

In unstrained MQWs, HH band sub-bands are usually higher than those of LH band, resulting in higher TE gain than the TM one. In delta-strained QWs (Cho and Choi, 2001), the delta layer introduces larger VB discontinuity for HH bands than LH bands, and the quantized energy levels for HH bands experience a shift downward, in contrast to those corresponding to LH bands, which shift upward. As a consequence, the TM transition strength for the transition between the first conduction sub-band and the top LH-like valence sub-band is much larger than the TE transition strength.

In addition to polarization insensitivity, SOA has to provide as high as possible transmission gain. Totović et al. (2013), presented a case study, where both design criteria have been implemented: high device gain and polarization insensitivity. The study analyzes two types of SOA waveguides, with active regions described in Section 21.3.1: unstrained bulk and tensile-strained MQWs, which are strain compensated by compressively strained barriers. In order to satisfy the first design criterion, the upper limit of the confinement factor is set to approximately 30%, thereby preventing the strong influence of ASE on SOA saturation (de Valicourt, 2012). Due to the gain isotropy, polarization-insensitive operation of bulk SOA requires equal confinement factors for both polarizations ($\Gamma_{TE} = \Gamma_{TM}$) (Connelly, 2001; Labukhin and Li, 2006; Michie et al., 2006). This requires optimization of the waveguide structure since confinement factor Γ_{TE} is usually larger than Γ_{TM} (Connelly, 2002; Labukhin and Li, 2006; Michie et al., 2006). Instead of using the unstrained, it is possible to implement a tensile-strained bulk active region, which leads to the polarization and gain anisotropy (Michie et al., 2006; Connelly, 2007). If the difference between the confinement factors is sufficiently small, a carefully chosen amount of tensile strain (Michie et al., 2006; Connelly, 2007) may compensate for the difference in the confinement factors Γ_{TE} and Γ_{TM} without the waveguide optimization. This can provide an efficient way for polarization-insensitive operation of bulk SOAs.

Similarly, tensile-strained MQWs in the active region, for which TM gain dominates over TE gain, may compensate for the difference in confinement factors Γ_{TE} and Γ_{TM}, leading to balanced modal gains for both polarizations, $\Gamma_{TE}g_{TE} = \Gamma_{TM}g_{TM}$, in the wavelength range of interest. The design of the MQW active region and the waveguide as a whole can be performed by a self-consistent iterative procedure involving calculation of the confinement factors and material gain, which ultimately, should provide optimized well/barrier dimensions, strain, number of wells, and waveguide cladding. The self-consistent procedure becomes important if the calculation of the confinement factor accounts for the refractive index variation of the active region, associated with variation of the material gain, and caused by carrier injection into the MQW structure (Totović et al., 2013). For the MQW active region from Section 21.3.1, the procedure leads to the confinement factor ratio $\Gamma_{TM}/\Gamma_{TE} = 0.806$, with $\Gamma_{TE} = 22\%$. The same procedure is applied in the optimization of the waveguide for unstrained bulk SOA. The optimization is performed for 1.55 μm and an average carrier density of 2×10^{18} cm^{-3}, for both bulk and MQW SOAs. The modal gains for TE and TM polarization of the optimized MQW active region are compared in Figure 21.3 for a range of photon energies and carrier densities.

It should be brought to attention that the temperature variation can significantly affect the refractive index and consequently the confinement factor. If efficient heat sink is not available for SOA, temperature effects should be accounted for in the design of the waveguide. More details on temperature effects will be given in Section 21.4.3.4.

The proposed design of the MQW waveguide is similar to those described by Wünstel et al. (1996) and Wolfson (2001) in both optical confinement and layer thicknesses. A rigorous treatment of the confinement factor should take into account the variation of the refractive index imaginary part, i.e., gain, in the active region (Visser et al., 1997). However, in this case study, the focus was set only on the variation of the real part of refractive index Δn_r due to the carrier injection. Although for other carrier densities and photon energies confinement factors somewhat detune from the optimal values and ratio, they partially compensate for the material gain ratio detuning, toning down the difference in modal gains for the TE and TM modes.

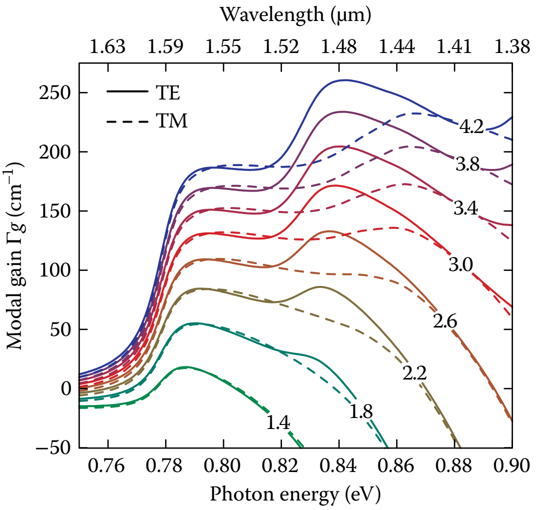


FIGURE 21.3 Spectral dependency of modal gain for polarization insensitive multiple quantum well (MQW) active region for the transverse-electric (TE) (solid lines) and transverse-magnetic (TM) (dashed lines) modes, for different carrier densities n (10^{18} cm^{-3}).

Distribution of the transversal components of the electric $\mathbf{E}(x, y)$ and magnetic $\mathbf{H}(x, y)$ fields is determined by a system of Helmholtz equations (Kawano and Kitoh, 2001):

$$\nabla_{\perp}^{2}\mathbf{E}\left(x,y\right)+\left[\left(n_{\mathrm{r}}+\Delta n_{\mathrm{r}}\right)^{2}k_{0}^{2}-\beta^{2}\right]\mathbf{E}\left(x,y\right)=0, \tag{21.7}$$

$$\left(n_{\mathrm{r}}+\Delta n_{\mathrm{r}}\right)^{2}\boldsymbol{\nabla}_{\perp}\left[\frac{1}{\left(n_{\mathrm{r}}+\Delta n_{\mathrm{r}}\right)^{2}}\boldsymbol{\nabla}_{\perp}\mathbf{H}\left(x,y\right)\right]+\left[\left(n_{\mathrm{r}}+\Delta n_{\mathrm{r}}\right)^{2}k_{0}^{2}-\beta^{2}\right]\mathbf{H}\left(x,y\right)=0, \tag{21.8}$$

where $\Delta n_{\mathrm{r}} = \Delta n_{\mathrm{r}}(n, \hbar\omega)$ is the refractive index variation, which is nonzero only in the active region of SOA, k_0 is the wave vector, $\beta = n_{\mathrm{eff}}k_0$ is the wave propagation constant, and n_{eff} is the effective index of refraction, also dependent on n and $\hbar\omega$. Equations are solved using the finite element method (FEM), with all boundaries set to the Neumann boundary conditions, except for the top ridge, which is set to the Dirichlet boundary condition.

Confinement factors for TE and TM polarizations are defined as (Kawano and Kitoh, 2001)

$$\Gamma_{\mathrm{TE}}=\int_{\mathrm{a.r.}}\left|\mathbf{E}\left(x,y\right)\right|^{2}\mathrm{d}x\mathrm{d}y\Big/\int_{-\infty}^{+\infty}\left|\mathbf{E}\left(x,y\right)\right|^{2}\mathrm{d}x\mathrm{d}y, \tag{21.9}$$

$$\Gamma_{\mathrm{TM}}=\frac{1}{n_{\mathrm{a.r.}}^{2}}\int_{\mathrm{a.r.}}\left|\mathbf{H}\left(x,y\right)\right|^{2}\mathrm{d}x\mathrm{d}y\Big/\sum_{i}\frac{1}{n_{i}^{2}}\int_{i}\left|\mathbf{H}\left(x,y\right)\right|^{2}\mathrm{d}x\mathrm{d}y, \tag{21.10}$$

where a.r. stands for the active region and i stands for each subdomain of the waveguide, including the active region. It should be noted that the background part of the refractive index $n_{\mathrm{r},i}$, with excluded Δn_{r} for the active region, is derived from Adachi's interpolation formulas (Adachi, 1989), and it is consequently a function of photon energy, i.e., $n_{\mathrm{r},i} = n_{\mathrm{r},i}(\hbar\omega)$. Since this dependence has abrupt changes, quantities shown in Figure 21.4 cannot exhibit smooth behavior.

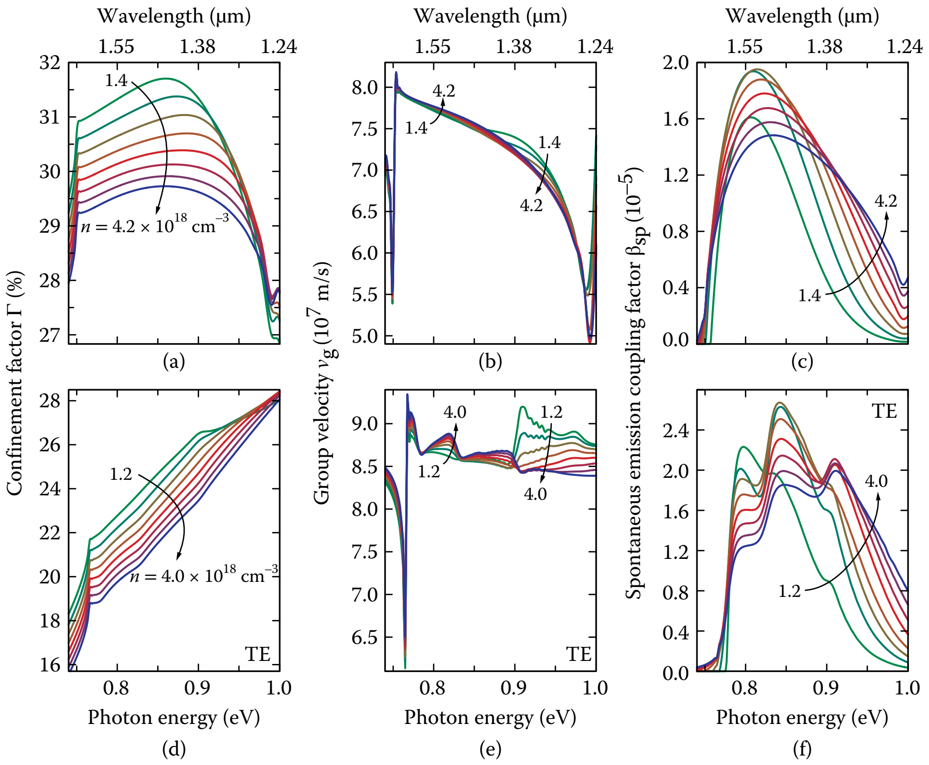


FIGURE 21.4 Spectral dependency of (a, d) confinement factor, (b, e) group velocity, and (c, f) spontaneous emission coupling factor for different carrier densities n (10^{18} cm^{-3}) for (a) through (c) bulk and (d) through (f) multiple quantum well (MQW) active region.

Another important parameter for ASE noise modeling in SOAs is the spontaneous emission coupling factor $\beta_{sp}(n, \hbar\omega)$, which is defined as the spontaneous emission rate coupled to the one optical mode, relative to the total spontaneous emission rate, $R_{sp}(n) = \int r_{sp}(n, \hbar\omega)\mathrm{d}(\hbar\omega)$, and can be calculated according to the following expression (Coldren et al., 2012):

$$\beta_{sp} = \frac{\Gamma v_g}{V}\frac{g n_{sp}}{R_{sp}}, \tag{21.11}$$

where V is the active region volume and Γ and g are the averaged values of confinement factor and material gain over polarizations, respectively, obtained in the same way as the average optical matrix element for radiative recombination rate. Here, $n_{sp}(n, \hbar\omega)$ is the population inversion factor defined by (Coldren et al., 2012)

$$n_{sp} = \left[1 - \exp\left(\frac{\hbar\omega - \Delta E_f}{k_B T}\right)\right]^{-1}, \tag{21.12}$$

where $\Delta E_f = E_f^{(c)} - E_f^{(v)}$ is the difference between the quasi-Fermi levels of CB and VB. In order to calculate β_{sp}, in addition to confinement factor, it is necessary to calculate the group velocity of the light $v_g(n, \hbar\omega)$:

$$v_g = \frac{\partial\omega}{\partial\beta} = \frac{1}{\hbar}\cdot\left[\frac{\partial\beta}{\partial(\hbar\omega)}\right]^{-1}. \tag{21.13}$$

Spectral dependencies of the confinement factor, group velocity, and spontaneous emission coupling factor are shown in Figure 21.4, for both bulk and the MQW active region, optimized to be polarization insensitive. Therefore, the quantities for the MQW active region are given only for TE polarization.

21.4 Rate Equations

In order to analyze the SOA's performance, it is first necessary to identify the main physical processes underlying its operation, which are related to the interaction mechanisms between photons and carriers. These are then typically presented in the form of a coupled system of differential equations denoted as the rate equation system. Optical signal can be described either in terms of complex electric field or in terms of photon density and phase. On the other hand, carriers are modeled either via carrier density (Dailey and Koch, 2007; Totović et al., 2013) or material gain (Shtaif et al., 1998), or, in some cases, using the integrated material gain along the direction of signal propagation (Agrawal and Olsson, 1989; Cassioli et al., 2000; Antonelli and Mecozzi, 2013).

For TW-SOA, it usually suffices to analyze only the forward-propagating optical signal, with respect to the longitudinal axis. However, for R-SOA, both forward and backward-propagating optical signals need to be accounted for. Although the model's complexity increases when counterpropagating optical signals are analyzed, it is often good practice to develop the generalized model since it can easily be customized to account for any SOA type by simply changing the power reflectivity coefficient of the front and/or rear facet.

21.4.1 Basic Rate Equations

During the decades of SOA study, a variety of models of different complexity emerged. Some of them are based on signal power (Agrawal and Olsson, 1989), intensity (Adams et al., 1985), or photon density analysis (Totović et al., 2014), which may or may not include a separate phase equation, whereas the others treat the optical signal using the complex electric field envelope (Melo and Petermann, 2008; Schrenk, 2011; Antonelli and Mecozzi, 2013). In its basic form, the SOA model describes the interaction between the carrier density $n(z, t)$ and the two counterpropagating optical signals, forward (+) and backward (−), with respect to the longitudinal z-axis, all of which are dependent on both time t and the z-coordinate. For SOA operation, it is necessary to achieve population inversion, which is done using the electric (bias) current I. When the bias current is low, carriers are mainly drained by spontaneous recombination, which can be nonradiative, yielding phonons (Ghafouri-Shiraz, 2004), or radiative, which produces noncoherent spectrally wide optical signal, regarded as noise. As the current increases beyond the transparency value, enough carriers exist in the active region to make the stimulated emission the dominant process and the input signal can be amplified during propagation. For the simplest case of bulk SOA, carrier dynamics can be described as follows (Coldren et al., 2012; Totović et al., 2013):

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{I}{qV} - \left[An + R_{\mathrm{sp}}(n) + Cn^3\right] - R_{\mathrm{st}}(n). \tag{21.14}$$

The first term on the right-hand side (RHS) of Equation 21.14 describes the rate at which the carriers are "pumped" into the active region, where q is the elementary charge, and $V = WHL$ is the active region volume, dependent on the active region width W, height H, and length L. The second term accounts for the loss of carriers through spontaneous recombination, where $R_{\mathrm{sp}}(n)$ is the total radiative spontaneous recombination rate, and A and C are the Shockley–Read–Hall and Auger coefficients, respectively, related to nonradiative spontaneous recombination (Coldren et al., 2012). In some instances, carrier loss due to radiative and nonradiative spontaneous recombination is modeled as n/τ, where τ is the carrier lifetime, which is treated either as constant (Thylén, 1988; Shtaif et al., 1998; Totović et al., 2016) or dependent on carrier density (Melo and Petermann, 2008; Xia and Ghafouri-Shiraz, 2016). Finally, the last term on the RHS of Equation 21.14, $R_{st}(n)$, accounts for the loss of carriers due to stimulated emission, either through the amplification of the signal alone or the signal and noise combined.

21.4.1.1 Electric Field Envelope Model

Each optical signal can be represented by its electric field, which can be separated into the carrier signal and the slowly varying envelope along the propagation direction. The complex envelope can then be analyzed either as is (Loudon et al., 2005), or, more commonly, it can be normalized such that its squared magnitude represents photon density (Connelly, 2007; Totović et al., 2011), power (Agrawal and Olsson, 1989; Cassioli et al., 2000), or intensity (Henry, 1982). Regardless of the choice for normalization, the system of equations remains the same since the underlying spatiotemporal dependence is unmodified by the normalization constants. The evolution of electric field envelopes $E_{\pm}(z, t)$, normalized such that their square magnitude represents photon densities, $|E_{\pm}(z, t)|^2 = S_{\pm}(z, t)$, propagating forward (+) and backward (−) along the longitudinal z-axis, is governed by the following system of equations (Agrawal and Olsson, 1989):

$$\pm\frac{\partial E_{\pm}}{\partial z} + \frac{1}{v_g}\frac{\partial E_{\pm}}{\partial t} = \frac{1}{2}\left(\Gamma g - \alpha_i\right) E_{\pm}. \tag{21.15}$$

Here, v_g is the group velocity, Γ stands for the optical confinement factor, which determines the portion of the signal propagating through the active region relative to the total photon density, and g is the material gain of the active region, all given at the signal frequency ω. These quantities can be derived from the models given in Section 21.3, and in Figures 21.2 and 21.4, by interpolation at the signal central energy $\hbar\omega$. With $\alpha_i = K_0 + \Gamma K_1 n$ active region loss is denoted (Connelly, 2001), where K_0 and K_1 stand for the carrier-independent and carrier-dependent loss coefficients, respectively, which take into account the intrinsic material loss and the free-carrier and inter-VB absorption (Dailey and Koch, 2009; Schrenk, 2011). It should be noted that, in general, all parameters appearing in Equation 21.15, namely v_g, Γ, g, and α_i, depend on carrier density n, and implicitly on t and z. This dependence notation is omitted from Equation 21.15 and the equations that will follow for compactness and clarity. Nevertheless, it is common to treat v_g and/or Γ as constants (Jin et al., 2003; Totović et al., 2014), which simplifies the model at the expense of accuracy.

Input optical signal, given by its normalized envelope E_0, is assumed to be entering the device at the front facet ($z = 0$), with the amplitude reflection coefficient r_1, and transmission coefficient t_1, as shown in Figure 21.5. The front facet is usually coated with the antireflective layer, and it is a common practice to neglect any residual reflectivity and assume $r_1 = 0$ and $t_1 = 1$, especially if the analytical or semianalytical solution is sought (Antonelli and Mecozzi, 2013; Totović et al., 2014). On the other hand, the rear facet ($z = L$) has the amplitude reflection coefficient r_2, which can be neglected for TW-SOA, but needs to be accounted for in the case of an R-SOA. If the cleaved facet is used as a high-reflecting one, the amplitude reflection coefficient can be determined as $r_2 = (1 - n_r)/(1 + n_r)$ under the assumption of TE polarization,

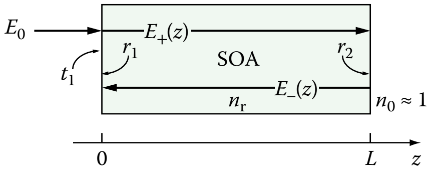


FIGURE 21.5 Semiconductor optical amplifier (SOA) structure with relevant optical signals, refractive indices, and reflection and transmission coefficients denoted.

where n_r is the refractive index of the active region. Boundary conditions according to Figure 21.5 read (Adams et al., 1985, Jin et al., 2003)

$$\begin{aligned} E_+(0) &= t_1 E_0 + r_1 E_-(0), \\ E_-(L) &= r_2 E_+(L). \end{aligned} \tag{21.16}$$

It is common practice to introduce the LEF α in Equation 21.15, which accounts for the phase modulation due to variation of real and imaginary parts of susceptibility χ (Henry, 1982; Agrawal and Olsson, 1989; Dailey and Koch, 2009; Antonelli and Mecozzi, 2013). This variation can be induced by injection of carriers, α_N, and carrier heating (CH), α_CH (Occhi et al., 2003). The total phase change will be the sum of these two contributions since they can be treated as uncorrelated. Carrier-induced LEF is defined as (Osinski and Buus, 1987; Occhi et al., 2003)

$$\alpha_\mathrm{N} = -\frac{\partial Re\{\chi\}/\partial n}{\partial Im\{\chi\}/\partial n} = -2k_0 \frac{\partial n_\mathrm{r}/\partial n}{\partial g/\partial n}, \tag{21.17}$$

where $k_0 = 2\pi\nu/c$ is the wave vector. For a small change in carrier density, which usually is the case, the derivatives in Equation 21.17 can be replaced with finite differences, which leads to (Henry, 1982; Osinski and Buus, 1987)

$$\alpha_\mathrm{N} = -2k_0 \frac{\Delta n_\mathrm{r}}{\Delta g}, \tag{21.18}$$

where Δn_r and Δg are the differential change in refractive index and material gain, respectively, induced by carrier density variation (Henry, 1982; Osinski and Buus, 1987; Agrawal, 1990). The deviation Δn_r is a result of several different mechanisms, including dipole band-to-band transition (Dailey and Koch, 2009; Qin et al., 2012), band-filling (Burstein–Moss effect), bandgap shrinkage, and free-carrier absorption (plasma effect) (Bennett et al., 1990). Band-filling is related to the decrease in absorption for photon energies slightly above nominal bandgap, caused by low density of states in CB that are easily filled, and is mostly pronounced for heavily doped semiconductors (Moss et al., 1973). Bandgap shrinkage is related to large concentrations of electrons at the bottom of CB, which repel one another due to Coulomb force, leading to the lowering of the energy of CB (Bennett et al., 1990). Finally, free-carrier absorption describes transition of free carriers to the higher-energy level due to photon absorption, and the change in refractive index can be described using the Drude model (Moss et al., 1973; Dailey and Koch, 2009; Qin et al., 2012). These four effects are assumed to be independent, and the total refractive index deviation can be found as a sum of their respective contributions.

Another contribution to LEF, caused by carrier temperature variation, is defined as (Occhi et al., 2003)

$$\alpha_\mathrm{CH} = -2k_0 \frac{\partial n_\mathrm{r}/\partial T}{\partial g/\partial T}. \tag{21.19}$$

However, this contribution is often neglected in SOA models, and LEF is assumed to be $\alpha \approx \alpha_\mathrm{N}$. Equation 21.15 now becomes

$$\pm\frac{\partial E_\pm}{\partial z} + \frac{1}{v_\mathrm{g}}\frac{\partial E_\pm}{\partial t} = \frac{1}{2}\left[\Gamma g\,(1 - i\alpha) - \alpha_\mathrm{i}\right] E_\pm. \tag{21.20}$$

It should be stressed that, depending on the sign convention adopted for the α-parameter, there may be variation in the literature in the sign prefixing α in Equation 21.20.

Having defined the equations which govern spatiotemporal electric field envelope distribution of the counterpropagating optical signals, it is possible to devise the stimulated emission rate in Equation 21.14, which is proportional to the total photon density in the amplifier's active region, S_Σ:

$$R_{\mathrm{st}}(n) = v_{\mathrm{g}} g S_\Sigma = v_{\mathrm{g}} g \left|E_+ + E_-\right|^2 . \tag{21.21}$$

It can be seen from Equation 21.21 that the interference between the two counterpropagating optical signals will cause the spatial grating in the carrier density distribution. However, characteristic grating length will be of the order of the signal's wavelength λ, and it will be easily washed out by the diffusion (Yacomotti et al., 2004; Serrat and Masoller, 2006; Totović et al., 2013). This justifies neglecting of all high-frequency terms, keeping only the sum of the squared electric field magnitudes, i.e., photon densities, $S_\Sigma = |E_+|^2 + |E_-|^2 = S_+ + S_-$. Depending on the model complexity, S_Σ can also include ASE noise.

21.4.1.2 Photon Density Model

By expressing the electric field envelope in terms of photon densities, $S_\pm$, and phase, $\varphi_\pm$:

$$E_\pm(z,t) = \sqrt{S_\pm(z,t)} \exp\left[i\varphi_\pm(z,t)\right], \tag{21.22}$$

Equation 21.20 can be reduced to the system of coupled TW equations for forward and backward propagation directions, written with respect to photon densities (Qin et al., 2012):

$$\pm\frac{\partial S_\pm}{\partial z} + \frac{1}{v_{\mathrm{g}}}\frac{\partial S_\pm}{\partial t} = \left(\Gamma g - \alpha_{\mathrm{i}}\right) S_\pm, \tag{21.23}$$

and phase (Agrawal and Olsson, 1989):

$$\pm\frac{\partial \varphi_\pm}{\partial z} + \frac{1}{v_{\mathrm{g}}}\frac{\partial \varphi_\pm}{\partial t} = -\frac{1}{2}\Gamma g \alpha. \tag{21.24}$$

Using the definition of LEF, Equation 21.24 can be written in yet another form that emphasizes the underlying cause of phase variation (Agrawal, 1990; Dailey and Koch, 2007; Qin et al., 2012):

$$\pm\frac{\partial \varphi_\pm}{\partial z} + \frac{1}{v_{\mathrm{g}}}\frac{\partial \varphi_\pm}{\partial t} = k_0 \Gamma \Delta n_{\mathrm{r}}. \tag{21.25}$$

If the signal is assumed to preserve its coherence during propagation, both Equations 21.23 and 21.25 are required; otherwise, it is sufficient to use only Equation 21.23 in modeling an SOA. Appending the phase equation to the model is important especially if the input signal carries information encoded in its phase, which is the case for many advanced modulation formats, such as phase shift keying (PSK), quadrature amplitude modulation (QAM), and quadrature phase shift keying (QPSK), to name a few (Schmuck et al., 2013; Koenig et al., 2014).

Assuming that the signal entering the device at the front facet ($z = 0$) is $E_0 = \sqrt{S_0}\exp(i\varphi_0)$, boundary conditions given by Equation 21.16 can now be rewritten using Equation 21.22. Additionally, amplitude reflection and transmission coefficients can be written as functions of the power reflectivity coefficients R_1 and R_2, corresponding to front and rear facets, respectively. The input signal does not change the propagation direction during transmission through the front facet, so no phase change is present, and $t_1 = \sqrt{(1 - R_1)}$ (Adams et al., 1985). On the other hand, during reflection at any facet, the signal being reflected does change the propagation direction, and a phase change of π needs to be accounted for. This can also be inferred from the definition of the amplitude reflection coefficient r_2 for the cleaved facet, which is negative. Finally, amplitude reflection coefficients can be written as $r_1 = \sqrt{R_1}\exp(i\pi)$ and $r_2 = \sqrt{R_2}\exp(i\pi)$. Substituting amplitude reflection and transmission coefficients in Equation 21.16, and expressing electric fields in terms of photon densities and phase, yields:

$$\begin{aligned}&\sqrt{S_+(0)}\exp[i\varphi_+(0)] = \sqrt{1-R_1}\sqrt{S_0}\exp(i\varphi_0) + \sqrt{R_1}\exp(i\pi)\sqrt{S_-(0)}\exp[i\varphi_-(0)],\\&\sqrt{S_-(L)}\exp[i\varphi_-(L)] = \sqrt{R_2}\exp(i\pi)\sqrt{S_+(L)}\exp[i\varphi_+(L)].\end{aligned} \tag{21.26}$$

By multiplying previous set of equations with their corresponding complex conjugates, the boundary conditions for photon densities can be derived:

$$\begin{aligned}&S_+(0) = (1-R_1)S_0 + R_1S_-(0) - 2\sqrt{(1-R_1)S_0}\sqrt{R_1S_-(0)}\cos[\varphi_0 - \varphi_-(0)],\\&S_-(L) = R_2S_+(L).\end{aligned} \tag{21.27}$$

Equating the arguments of the left-hand side (LHS) and RHS of Equation 21.26, the following boundary conditions for phases can be found (Totović et al., 2013):

$$\begin{aligned}&\varphi_+(0) = \arctan\left[\frac{\sqrt{(1-R_1)S_0}\sin\varphi_0 - \sqrt{R_1S_-(0)}\sin\varphi_-(0)}{\sqrt{(1-R_1)S_0}\cos\varphi_0 - \sqrt{R_1S_-(0)}\cos\varphi_-(0)}\right],\\&\varphi_-(L) = \varphi_+(L) + \pi.\end{aligned} \tag{21.28}$$

Typically, the input optical signal is given in terms of power rather than photon density. For the input optical power P_0, the photon density can be determined as $S_0 = \Gamma P_0/(\hbar\omega v_g WH)$, where $\hbar\omega$ is the signal energy.

21.4.1.3 Signal Spectral Dependence

Most of the parameters figuring in Equations 21.23 and 21.25 do exhibit spectral dependence, as shown in Figures 21.2 and 21.4. When working with photon densities, signal is assumed to be infinitely spectrally narrow and centered at the energy corresponding to the maximum spectral photon density, $\hbar\omega_0$. This assumption justifies the usage of a simplified model where the parameter values are given at the signal central energy, as is the case with Equation 21.23. In practice, signal will always have a finite spectral width, which is not necessarily narrow and cannot always be neglected. This is especially important if SOAs are used in WDM systems, where multiple signals at different wavelengths travel through the active region. In order to correctly model the signal propagation in this case, Equation 21.23 is rewritten in terms of spectral photon density, $s_\pm(\hbar\omega, z, t)$, for forward and backward propagation (Totović et al., 2013):

$$\pm\frac{\partial s_\pm}{\partial z} + \frac{1}{v_g}\frac{\partial s_\pm}{\partial t} = \left(\Gamma g - \alpha_i\right)s_\pm. \tag{21.29}$$

It should be stressed that now v_g, Γ, g, and α_i are spectrally dependent. Additionally, two more changes to the model are required. The first is related to the carriers consumed by stimulated emission in Equation 21.14, which can be determined using

$$R_{st}(n) = \int_0^\infty v_g g s_\Sigma \mathrm{d}(\hbar\omega), \tag{21.30}$$

where $s_\Sigma = s_+ + s_-$. The second is related to the boundary conditions given in Equation 21.27, which are now written with respect to $s_\pm$ instead of $S_\pm$. Finally, if no reflection coating is used for the R-SOA's rear facet, the reflection coefficient R_2 also becomes spectrally dependent through $n_\mathrm{r}(\hbar\omega)$.

21.4.2 Amplified Spontaneous Emission

Spontaneous recombination can usually be viewed as an undesirable side effect of SOA operation, from several viewpoints. First, the carriers that should be reserved for signal amplification are indefinitely lost, meaning that the higher bias currents are required to achieve the same material gain. Second, the random nature of spontaneous recombination induces fluctuations in carrier density, which in turn causes the material gain and refractive index to change. This leads to fluctuations in both intensity and phase of the amplified signal (Ghafouri-Shiraz, 2004). Finally, radiative spontaneous recombination generates wideband signal with random phase, polarization, and spatial orientation. Once these photons are generated, they become impossible to separate from the signal, so they travel jointly through the active region and become amplified. Even if the signal is filtered at the SOA's output, a certain amount of noise will always be present in the narrow band around the signal central energy (Yariv and Yeh, 2007). An additional drawback is that the noise will have higher optical power when the input signal is vague since more carriers will be available for spontaneous recombination (Talli and Adams, 2003; Totović et al., 2013; Koenig et al., 2014). As the signal becomes more amplified, fewer carriers remain, and the ASE contribution becomes less important in terms of power.

The process of spontaneous emission is at its core stochastic, and, within the framework of the semiclassical approach, it can be modeled statistically in terms of its probability, similarly to the shot noise present in electronic devices (Ghafouri-Shiraz, 2004). The nature of ASE noise led to the development of numerous models of different degree of complexity, which can be roughly divided into two categories: deterministic, which do not require any random generators (Talli and Adams, 2003; Totović et al., 2013), and stochastic, which require random sources with certain probability density functions (Marcuse, 1984; Cassioli et al., 2000; Park et al., 2005; Melo and Petermann, 2008; Totović et al., 2011). In the semiclassical framework, ASE is usually treated as white Gaussian noise process, or as Poisson process (Park et al., 2005), whereas in the quantum mechanical formalism the noise is treated using the set of appropriate quantum noise operators, including spontaneous emission process, internal absorption, and vacuum field fluctuations (Shtaif et al., 1998). It has been shown by Shtaif et al. (1998) that the semiclassical results are essentially the same as the ones obtained using quantum description, with the exception of the shot noise term, which needs to be explicitly added to the semiclassical intensity. However, for sufficiently small bandwidths, the shot noise term can be ignored (Shtaif et al., 1998).

21.4.2.1 Deterministic Approach

It has already been pointed out that signal and noise travel together, which would justify modification of the photon density equation, either Equation 21.23 or Equation 21.29, to include another term modeling the rate with which the spontaneous photons are generated (Park et al., 2005; Dailey and Koch, 2007; Zhou et al., 2007; Totović et al., 2015). However, from the standpoint of detailed device analysis, it may be useful to separate noise and signal equations, allowing them to be coupled through the carrier rate equation since these two processes are uncorrelated (Ghafouri-Shiraz, 2004). ASE is known to have wide spectrum; therefore, it is more appropriate to use spectral photon densities, $a_\pm(\hbar\omega, z, t)$, rather than photon densities, for its description. The system of equations for two counterpropagating ASE signals reads (Talli and Adams, 2003; Melo and Petermann, 2008; Totović et al., 2013):

$$\pm\frac{\partial a_\pm}{\partial z} + \frac{1}{v_\mathrm{g}}\frac{\partial a_\pm}{\partial t} = \left(\Gamma g - \alpha_\mathrm{i}\right) a_\pm + \frac{1}{2v_\mathrm{g}}\Gamma\beta_\mathrm{sp} r_\mathrm{sp}, \tag{21.31}$$

where r_sp is the radiative spontaneous recombination rate per unit energy, given by Equation 21.2 or Equation 21.5, depending on the active region type, and β_sp is the spontaneous emission coupling factor given by Equation 21.11. Boundary conditions for noise are $a_+(0) = R_1 a_-(0)$ and $a_-(L) = R_2 a_+(L)$. Phase is not analyzed since it is random and can be averaged to zero. In other words, under the usual operating conditions, i.e., if no resonant cavity is present, ASE does not exhibit coherent behavior.

In order to have the complete model, it is necessary to include ASE together with the signal when calculating total spectral photon density, $s_\Sigma = s_+ + s_- + a_+ + a_-$, which further influences $R_{st}(n)$, given by Equation 21.30.

It should be noted that many other deterministic approaches for ASE analysis exist. The choice depends on the desired model complexity and the tradeoff between accuracy and computational resource consumption. These approaches include the photon statistic master equation method (Mukai and Yamamoto, 1982), field beating (Olsson, 1989), the equivalent circuit model (Berglind and Gillner, 1994; Ghafouri-Shiraz, 2004), the semiclassical wave theory model (Donati and Giuliani, 1997).

21.4.2.2 Resonant Properties of the SOA Cavity

In the previous pages, during the discussion of facet's power reflectivities and boundary conditions, it was assumed that all antireflective facets have zero reflectivity. Although this assumption can be helpful for model simplification, it can also mask some of the processes present in SOAs (Thylén, 1988; Zhou et al., 2007). Residual reflectivity, which always has some finite value, provides a feedback loop that acts as a filter, giving rise to resonances and antiresonances (Adams et al., 1985; Schrenk, 2011). This effect has been reported in both experimental (Olsson, 1989) and theoretical SOA analyses (Zhou et al., 2007; Totović et al., 2013), mostly related to ASE noise. An example of the output spectrum of the signal and noise combined, with visible resonant footprint of the cavity, calculated using the model presented by

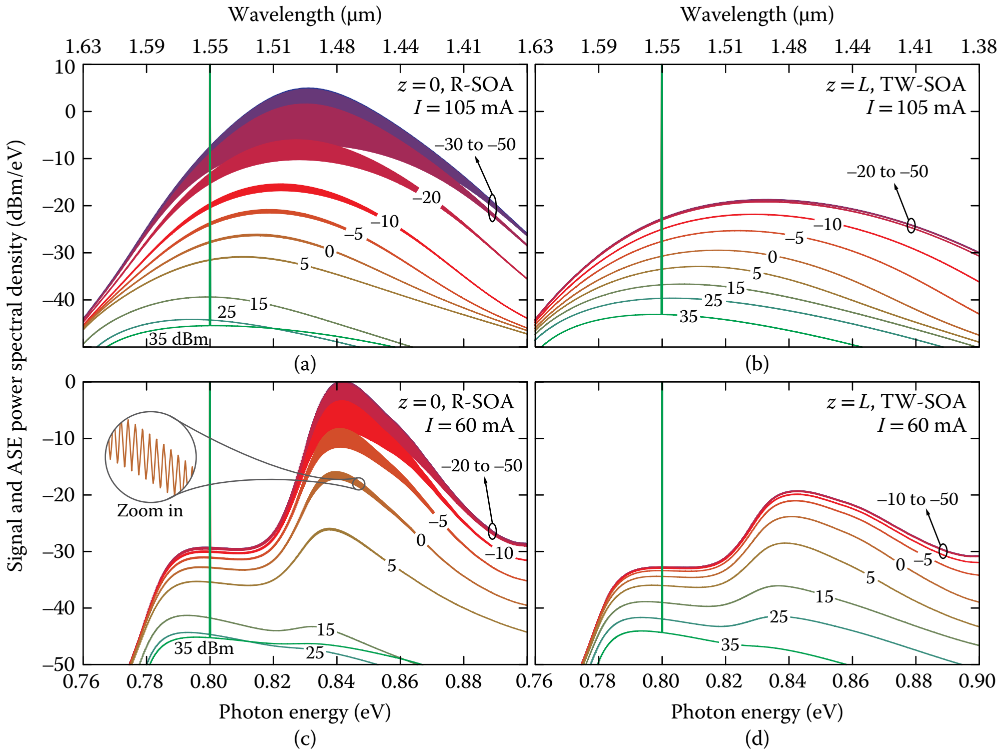


FIGURE 21.6 Total power spectral density at the semiconductor optical amplifier's (SOA's) output for different input optical powers in the case of (a, c) reflective (R) SOA at $z = 0$, and (b, d) traveling wave (TW) SOA at $z = L$, for (a, b) bulk and (c, d) multiple quantum well (MQW) active region.

Totović et al. (2013), is given in Figure 21.6. It can be seen that R-SOA, Figure 21.6a and c, exhibits more pronounced resonances and antiresonances in comparison with TW-SOA, Figure 21.6b and d, which is a consequence of the highly reflective rear facet. The ASE noise spectrum profoundly depends on the optical power of the input signal and resonant pattern washes out as the input optical power increases. This is to be expected since the strong signal easily depletes the carriers and suppresses the noise, consequently reducing the effect of multiple round trips within the cavity on its spectrum. The difference between the output spectrums for distinct active region types, bulk, Figure 21.6a and b, and MQW, Figure 21.6c and d, is a result of different spectrums of radiative spontaneous emission rate and material gain, given in Figure 21.2.

In order to develop the model capable of capturing ripples in the ASE output spectrum, two approaches can be used. One would require an equation for ASE phase evolution in addition to Equation 21.31, which describes ASE spectral photon density. This would enable us to account for interference, and subsequently filtering property of the SOA cavity. However, due to the random nature of the ASE noise phase, this approach would require a random number generator (Park et al., 2005) and could essentially be categorized as a stochastic model. Another approach is based on spectrum slicing and analysis of photon density corresponding to each resonance, or mode, followed by photon redistribution over the energy range between the two antiresonances according to the cavity transfer function. This has been shown to be equivalent to the analysis of the electric field or the spectral photon density and phase (Adams et al., 1985). The phase equation will be included implicitly during the assessment of resonant and antiresonant frequencies of the mth cavity mode, ω_m^{r}, and ω_m^{a}, respectively. These can be derived by equating the phase accumulation during the cavity round trip to $2m\pi$ for resonant, or $(2m-1)\pi$ for antiresonant frequencies, where m is an integer. The single-pass phase shift, Φ, will be the same for either propagation direction and can be found from Equation 21.25 as

$$\Phi = \int_0^L \frac{\partial \varphi_+}{\partial z}\mathrm{d}z = \int_0^L \left(-\frac{1}{v_{\mathrm{g}}}\frac{\partial \varphi_+}{\partial t} + k_0 \Gamma \Delta n_{\mathrm{r}} \right)\mathrm{d}z = \frac{\omega}{v_{\mathrm{g}}}L + k_0 \int_0^L \Gamma \Delta n_{\mathrm{r}}\mathrm{d}z. \tag{21.32}$$

Here, $\omega/v_{\mathrm{g}} = n_{\mathrm{eff},0}k_0 = \beta_0$ is the wave propagation constant for zero carrier density, where $n_{\mathrm{eff},0}$ is the effective index of refraction for zero carrier density or the background index of refraction. Resonant and antiresonant frequencies are calculated from

$$2\frac{\omega}{c}\left(n_{\mathrm{eff},0}L + \int_0^L \Gamma \Delta n_{\mathrm{r}}\mathrm{d}z \right) = \begin{cases} 2m\pi, & \text{for } \omega_m^{\mathrm{r}} \\ (2m-1)\pi, & \text{for } \omega_m^{\mathrm{a}} \end{cases}. \tag{21.33}$$

Since both Γ and Δn_{r} depend on photon energy, $\hbar\omega$, resonant and antiresonant frequencies cannot be expressed in a closed analytical form. Additionally, we can conclude that ω_m^{r} and ω_m^{a} depend on carrier density through the second term in the LHS of Equation 21.33 since both Γ and Δn_{r} depend on n. This implies that the spectral output of either TW- or R-SOA will not be static; rather the frequencies corresponding to resonances and antiresonances will be shifted in the presence of the optical signal (Schrenk, 2011; Totović et al., 2013). However, comparing to the first term in the LHS of Equation 21.33, the contribution of the second term is usually small and can be justifiably neglected.

By integrating Equation 21.31 over the energy between each of the two antiresonances, ω_m^{a} and $\omega_{m+1}^{\mathrm{a}}$, a system of equations can be developed, written with respect to noise photon densities of the mth mode, $A_\pm^m(z,t)$, for forward (+) and backward (−) propagation directions (Schrenk, 2011; Totović et al., 2013):

$$\pm\frac{\partial A_\pm^m}{\partial z} + \frac{1}{v_{\mathrm{g}}}\frac{\partial A_\pm^m}{\partial t} = \left(\Gamma_m g_m - \alpha_{\mathrm{i}}^m\right)A_\pm^m + \frac{1}{2v_{\mathrm{g}}^m}\Gamma_m \beta_{\mathrm{sp}}^m R_{\mathrm{sp}}^m, \tag{21.34}$$

where the index m denotes the parameter values corresponding to the mth cavity resonance, and R_{sp}^{m} stands for the fraction of radiative spontaneous recombination rate injected into the mth mode. This approach, relying on separating amplification and filtering properties of the amplifier cavity, reduces time, memory, and processing resource consumption during simulation.

The filtering function of the amplifier is given in the form of an Airy function corresponding to the FP cavity (Adams et al., 1985):

$$G_m(\hbar\omega) = \frac{\left(1-R_1\right)\left(1-R_2\right)G_{\mathrm{s}}^{m}}{\left(1-G_{\mathrm{s}}^{m}\sqrt{R_1R_2}\right)^2 + 4G_{\mathrm{s}}^{m}\sqrt{R_1R_2}\sin^2\Phi(\hbar\omega)}, \tag{21.35}$$

where Φ is the single-pass phase shift, defined by Equation 21.32, and G_{s}^{m} stands for the single-pass gain at resonant energy $\hbar\omega_m^{\mathrm{r}}$:

$$G_{\mathrm{s}}^{m} = \exp\left[\int_0^L \left(\Gamma_m g_m - \alpha_{\mathrm{i}}^{m}\right)\mathrm{d}z\right]. \tag{21.36}$$

The transmittance given by Equation 21.35 accounts for the signal amplification, so care should be exercised when using it with Equation 21.34. Prior to photon filtering, the transmittance needs to be normalized to unity over the mth mode energy range since the noise amplification has already been accounted for using Equation 21.34, and only photon redistribution is required. The normalized filtering function $T_m(\hbar\omega)$ can be found as

$$T_m(\hbar\omega) = \frac{G_m(\hbar\omega)}{\left(\hbar\omega_{m+1}^{\mathrm{a}} - \hbar\omega_m^{\mathrm{a}}\right)^{-1}\int_{\hbar\omega_m^{\mathrm{a}}}^{\hbar\omega_{m+1}^{\mathrm{a}}} G_m(\hbar\omega)\,\mathrm{d}(\hbar\omega)}, \tag{21.37}$$

which is equivalent to

$$T_m(\hbar\omega) = \frac{G_m(\hbar\omega)}{\pi^{-1}\int_0^{\pi} G_m(\Phi)\,\mathrm{d}\Phi} = \frac{\sqrt{1+\gamma_m}}{1+\gamma_m\sin^2\Phi}, \tag{21.38}$$

where the γ-parameter is defined as follows:

$$\gamma_m = \frac{4G_{\mathrm{s}}^{m}\sqrt{R_1R_2}}{\left(1-G_{\mathrm{s}}^{m}\sqrt{R_1R_2}\right)^2}. \tag{21.39}$$

The validity of the previous model can be easily checked by analysis of the corner case where any of R_1 or R_2 is equal to zero. This would yield $\gamma_m = 0$, and $T_m = 1$, which corresponds to the case of uniformly distributed photons. This result is to be expected since for zero reflectivity no feedback loop exists.

Finally, the spectral photon densities of the noise at the SOA's output can be found using (Totović et al., 2013)

$$\begin{aligned} a_-(0,\hbar\omega) &= \sum_m \left[\left(\hbar\omega_{m+1}^{\mathrm{a}} - \hbar\omega_m^{\mathrm{a}}\right)^{-1} A_-^{m}(0)\,T_m(\hbar\omega)\right], \\ a_+(L,\hbar\omega) &= \sum_m \left[\left(\hbar\omega_{m+1}^{\mathrm{a}} - \hbar\omega_m^{\mathrm{a}}\right)^{-1} A_+^{m}(L)\,T_m(\hbar\omega)\right], \end{aligned} \tag{21.40}$$

where the summation is done over the modes m from 1 to the total number of accounted modes M.

In order to increase the efficiency of numerical calculations related to noise, an approach based on grouping of several adjacent modes into clusters was proposed by Totović et al. (2013). Although this degrades the accuracy to a certain point, the small intermodal space between resonances, which is of the order of $\Delta\nu = 70$ GHz, or $\Delta\lambda = 0.54$ nm for the active region length of 800 μm, ensures that no significant changes for any spectrally dependent parameter exist within the cluster. The proposed optimal number of modes within one cluster is 5, based on the fact that material gain and stimulated emission rate do not change more than 2% for bulk and 3% for MQW active region, within the cluster's range of frequencies, for the most part of the spectrum. In addition, the grouping of modes need not be uniform. In other words, clusters can comprise a lower number of modes in the frequency range close to the signal central frequency and maximum spontaneous emission rate and that number can be increased approaching the edges of spontaneous emission spectrum.

21.4.2.3 Stochastic Approach

Implementing any stochastic process relies on random number generators, or sources, meaning that only numerical analysis is possible in this case. Nonetheless, this approach can be very useful for investigating statistical properties of noise and signals. As both amplitude and phase of noise signal are random, it is suitable to use the equation written with respect to the normalized electric field envelope Equation 21.20, modified such that it includes a noise generator:

$$\pm\frac{\partial E_\pm}{\partial z} + \frac{1}{v_g}\frac{\partial E_\pm}{\partial t} = \frac{1}{2}\left[\Gamma g\left(1 - i\alpha\right) - \alpha_i\right]E_\pm + \mu_\pm. \tag{21.41}$$

Here, the complex term $\mu_\pm(z, t)$ stands for the random noise source, also known as the Langevin noise source (Henry, 1986; Coldren et al., 2012), which can be modeled as the Poisson or Gaussian phase-independent spatially uncorrelated white noise process (D'Ottavi et al., 1995; Cassioli et al., 2000; Park et al., 2005), based on the fluctuation-dissipation theorem (Shtaif et al., 1998). This implies that the mean value of $\mu_\pm(z, t)$ is zero, whereas the autocorrelation function satisfies the following condition:

$$\left\langle\left\langle \mu_\pm(z,t)\,\mu_\pm^*\left(z - z', t - t'\right)\right\rangle\right\rangle = \frac{1}{v_g^2}\Gamma\beta_{sp}R_{sp}L\delta\left(z'\right)\delta\left(t'\right). \tag{21.42}$$

Recalling that Equation 21.20 does not include spectrally dependent parameters, it can be seen that the noise source defined in this manner has infinite bandwidth and consequently infinite power due to neglected spectral dependence of material gain (Cassioli et al., 2000). In practice, material gain will have finite bandwidth and will limit the noise power to a finite value. This effect can be included by passing $\mu_\pm(z, t)$ through the bandpass filter, which will result in

$$\left\langle\left\langle \mu_{F\pm}(z,t)\,\mu_{F\pm}^*\left(z - z', t - t'\right)\right\rangle\right\rangle = \frac{1}{v_g^2}\Gamma\beta_{sp}R_{sp}LB_F\delta\left(z'\right), \tag{21.43}$$

where $\mu_{F\pm}(z, t)$ is the band-limited noise source and B_F is the noise equivalent bandwidth (D'Ottavi et al., 1995; Cassioli et al., 2000). If the segmentation along the spatial coordinate is small enough, such that the carrier density can be considered constant, in numerical simulations the Dirac delta function can be replaced with the inverse step size Δz (Marcuse, 1984):

$$\left\langle\left\langle \mu_{F\pm}(z,t)\,\mu_{F\pm}^*\left(z - z', t - t'\right)\right\rangle\right\rangle = \frac{1}{v_g^2}\Gamma\beta_{sp}R_{sp}LB_F\frac{1}{\Delta z}. \tag{21.44}$$

Now the noise source can be expressed as follows (Marcuse, 1984; Cassioli et al., 2000; Melo and Petermann, 2008):

$$\mu_{\mathrm{F}\pm}(z,t) = \frac{1}{v_{\mathrm{g}}}\sqrt{\Gamma\beta_{\mathrm{sp}}R_{\mathrm{sp}}B_{\mathrm{F}}\frac{L}{\Delta z}}x_{\mathrm{e}}, \tag{21.45}$$

where x_{e} is the complex Gaussian random variable. According to the sampling theorem, the sampling period, which is equivalent to the simulation time-step Δt, needs to be less than or equal to $1/(2B_{\mathrm{F}})$. The complex Gaussian random variable can be expressed in terms of its real and imaginary parts, which finally gives

$$\mu_{\mathrm{F}\pm}(z,t) = \sqrt{\frac{1}{2v_{\mathrm{g}}}\Gamma\beta_{\mathrm{sp}}R_{\mathrm{sp}}\frac{L}{v_{\mathrm{g}}}\frac{1}{\Delta t\Delta z}}\frac{x_1 + ix_2}{\sqrt{2}}, \tag{21.46}$$

where x_1 and x_2 are independent and identically distributed numerically generated Gaussian random variables with zero mean and unit variance (Melo and Petermann, 2008). Similarly, x_{e} can be expressed in terms of magnitude and phase, also distributed according to the Gaussian statistics, with zero mean and unit variance.

21.4.2.4 Noise Figure

The usual method of quantifying the noise is by determining the noise figure (NF). It can be shown (Simon et al., 1989) that the intrinsic NF of an SOA can be obtained using

$$\mathrm{NF}_{\mathrm{dB}} = 10\log_{10}\left[\frac{1}{G} + 2\frac{\left(G_{\mathrm{s}}-1\right)\left(1+R_1G_{\mathrm{s}}\right)}{G_{\mathrm{s}}\left(1-R_1\right)}n_{\mathrm{sp}}\frac{\Gamma g}{\Gamma g - \alpha_{\mathrm{i}}}\right], \tag{21.47}$$

where G stands for the device (transmission) gain and G_{s} for the single pass gain, defined by Equation 21.36.

21.4.3 Extended Rate Equation Model

Previous discussion on SOA modeling covered the most important aspects of the photon–carrier interaction dynamics. Nevertheless, many more upgrades to the model are possible, and the choice of the model complexity mainly depends on the intended SOA usage. Whenever possible, it is good practice to simplify the model and exclude the effects that bear little to no improvement to the results, in order to reduce resource consumption during computation. Some of the upgrades are listed in the following pages, but many more do exist, e.g., separate analysis of polarization components (Pillai et al., 2006; Melo and Petermann, 2008), inclusion of nonlinearities, etc.

21.4.3.1 Carrier Diffusion

Although often omitted from consideration, carrier diffusion plays an important role in SOA modeling. It provides the means for washing out the spatial grating in carrier density distribution caused by the interference between the electric fields of the counterpropagating optical signals (Yacomotti et al., 2004; Serrat and Masoller, 2006; Totović et al., 2013). In this manner, its effects are implicitly accounted for by neglecting the high frequency terms in carrier density spatial distribution and replacing $|E_+ + E_-|^2$ with $|E_+|^2 + |E_-|^2$. This approach, however, has its limitations. It has been shown that only "fast" spatial grating, coming from counterpropagating fields of the same mode, can be neglected, whereas the "slow" one, resulting from interference effects of the different longitudinal modes, remains to a certain degree (Yacomotti et al., 2004; Serrat and Masoller, 2006). In order to encompass all possible cases, the SOA model can be generalized by modifying the carrier density rate Equation 21.14 to include the term describing diffusion (Agrawal and Olsson, 1989; Serrat and Masoller, 2006):

$$\frac{\partial n}{\partial t} = D\frac{\partial^2 n}{\partial z^2} + \frac{I}{qV} - \left[An + R_{\text{sp}}(n) + Cn^3\right] - R_{\text{st}}(n), \tag{21.48}$$

where D stands for the diffusion coefficient and $R_{\text{st}}(n)$ includes the square magnitude of the sum of both signal and noise counterpropagating electric fields.

21.4.3.2 Nonlinear Gain Suppression

The two most important effects that can influence material gain on a short-term scale are spectral hole burning (SHB) (Ahn and Chuang, 1990) and CH (Willatzen et al., 1991). The former represents formation of frequency selective dip in the gain spectrum as a consequence of stimulated emission, whereas the latter describes the fact that the temperature of carriers and lattice may differ. Both can be phenomenologically accounted for by introducing the nonlinear gain suppression factor, ε, which causes gain reduction proportional to the photon density in the active region (Willatzen et al., 1991; Totović et al., 2013). Instead of using the so-called linear gain g, which is dependent only on photon energy and carrier density, equations should be modified such that nonlinear gain is used $g_{\text{NL}} = g/(1 + \varepsilon S_\Sigma)$, which also accounts for gain dependence on photon density. The complete set of equations for wideband SOA modeling can now be written in the following form (Totović et al., 2013):

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{I}{qV} - \left(An + R_{\text{sp}} + Cn^3\right) - \frac{R_{\text{st}}}{1+\varepsilon S_\Sigma}, \tag{21.49}$$

$$\pm\frac{\partial s_\pm}{\partial z} + \frac{1}{v_{\text{g}}}\frac{\partial s_\pm}{\partial t} = \left(\frac{\Gamma g}{1+\varepsilon S_\Sigma} - \alpha_{\text{i}}\right)s_\pm, \tag{21.50}$$

$$\pm\frac{\partial \varphi_\pm}{\partial z} + \frac{1}{v_{\text{g}}}\frac{\partial \varphi_\pm}{\partial t} = \frac{k_0\Gamma\Delta n_{\text{r}}}{1+\varepsilon S_\Sigma}, \tag{21.51}$$

$$\pm\frac{\partial a_\pm}{\partial z} + \frac{1}{v_{\text{g}}}\frac{\partial a_\pm}{\partial t} = \left(\frac{\Gamma g}{1+\varepsilon S_\Sigma} - \alpha_{\text{i}}\right)a_\pm + \frac{1}{2v_{\text{g}}}\Gamma\beta_{\text{sp}}r_{\text{sp}}. \tag{21.52}$$

Presence of the nonlinear gain suppression also modifies the position of resonant and antiresonant frequencies, given by Equation 21.33, and the new condition can be derived from Equation 21.51:

$$2k_0\left(n_{\text{eff},0}L + \int_0^L \frac{\Gamma\Delta n_{\text{r}}}{1+\varepsilon S_\Sigma}\mathrm{d}z\right) = \begin{cases} 2m\pi, & \text{for } \omega_m^{\text{r}} \\ (2m-1)\pi, & \text{for } \omega_m^{\text{a}} \end{cases}. \tag{21.53}$$

21.4.3.3 Carrier Transport Model for MQW and QD SOAs

As previously discussed, materials with higher levels of confinement can outperform bulk SOAs in many aspects. The most common choice of active regions aside from bulk are the ones based on MQWs or QDs. Due to the different carrier dynamics, these devices usually require a more elaborate model of carrier rate equations than the one introduced for bulk SOA.

The MQW region in lasers and SOAs is usually embedded into the separate confinement heterostructure (SCH) region. Injected carriers from the outer edges of the SCH region diffuse across the region, leading to the subsequent capture and emission of carriers by the quantum wells.

Figure 21.7 shows a typical SCH MQW active region. Carrier transport effects in MQW SOA can be modeled by rate equations written with respect to the carrier density in the barrier (continuum) states,

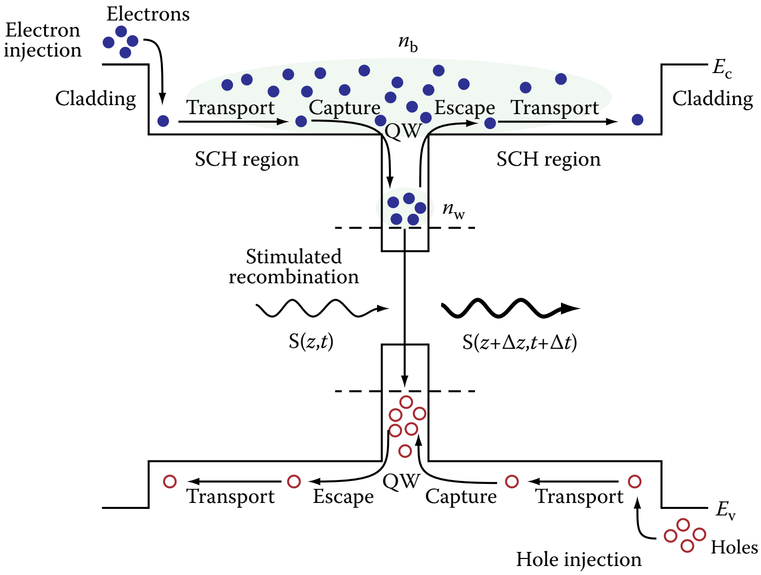


FIGURE 21.7 Band diagram of the multiple quantum well (MQW) semiconductor optical amplifier's (SOA's) active region and schematics showing carrier injection, transport, capture, escape and signal amplification by stimulated emission. (SCH, separate confinement heterostructure; QW, quantum well.)

including the SCH region and active layers, n_b, and the carrier density in the bound states of the well region, n_w, similar to SCH MQW lasers (Nagarajan et al., 1992; Keating et al., 1999; Totović et al., 2012). This model is referred to as the reservoir model (Nagarajan et al., 1992; Keating et al., 1999) and is equivalent to models that incorporate additional effects such as diffusive transport. Coupling of the carrier density in the barrier states above the MQWs to the carrier density in the MQWs is modeled by two terms representing carrier capture and escape into and from the wells, respectively:

$$\frac{dn_b}{dt} = \eta_{inj}\frac{I}{qV_b} - \frac{n_b}{\tau_b} - \frac{n_b}{\tau_{bw}} + \frac{n_w}{\tau_{wb}}\frac{V_w}{V_b}, \tag{21.54}$$

$$\frac{dn_w}{dt} = \frac{n_b}{\tau_{bw}}\frac{V_b}{V_w} - \frac{n_w}{\tau_w} - \frac{n_w}{\tau_{wb}} - \frac{R_{st}}{1+\varepsilon S_\Sigma}. \tag{21.55}$$

Here, η_{inj} is the injection efficiency, V_b is the volume of the SCH and active region, V_w is the volume of the well region, τ_b and τ_w are the carrier recombination lifetimes in the barrier and in the well region, respectively, τ_{bw} is the effective carrier diffusion across the SCH region and capture time by the wells, and τ_{wb} is the thermionic emission and carrier diffusion time from the well to the barrier states.

The TW equations for the counterpropagating optical photon densities retain the same form as in the case of the bulk SOA, with the exception of material gain and internal loss of the active region, which now dominantly depend on carrier density in the bound states of the well, n_w:

$$\pm\frac{\partial S_\pm}{\partial z} + \frac{1}{v_g}\frac{\partial S_\pm}{\partial t} = \left(\frac{\Gamma g}{1+\varepsilon S_\Sigma} - \alpha_i\right) S_\pm. \tag{21.56}$$

A more detailed model of the MQW active region's carrier dynamics may comprise additional rate equations for the reservoir model. For example, in addition to barrier states, well states may be divided into the excited state (ES) and the ground state (GS) (Qin et al., 2012). An even more elaborate approach is required for asymmetric MQW structures, where the tunneling effect of carriers through the barriers needs to be accounted for (Lysak et al., 2005, 2006).

The basic rate equations for QD SOA are essentially the same as for the detailed model of the MQW active region (Qin et al., 2012). The discrete energy levels in the well of QDs include the GS level and the ES level, which is doubly degenerated. The populations of these two levels are described by separate carrier densities, n_G and n_E, respectively, which are normalized with respect to the total dot volume V_D. Dots are interconnected by the wetting layer (WL), described by the carrier density, n_W, which is normalized to the WL volume, V_W. It can be assumed that the carriers are injected directly from the contacts into the WL and the barrier dynamics is thus ignored in the model (Berg et al., 2001; Qasaimesh, 2003). The rate equations describing the carrier dynamics read

$$\frac{\mathrm{d}n_W}{\mathrm{d}t} = \frac{I}{qV_W} + \frac{n_E}{\tau_e^E}\frac{V_D}{V_W}f'_W - \frac{n_W}{\tau_c}f'_E - \frac{n_W}{\tau_{sp}}, \tag{21.57}$$

$$\frac{\mathrm{d}n_E}{\mathrm{d}t} = \frac{n_W}{\tau_c}\frac{V_D}{V_W}f'_E + \frac{n_G}{\tau_e^G}f'_E - \frac{n_E}{\tau_e^E}f'_W - \frac{n_E}{\tau_0}f'_G - \frac{n_E}{\tau_{sp}}, \tag{21.58}$$

$$\frac{\mathrm{d}n_G}{\mathrm{d}t} = \frac{n_E}{\tau_0}f'_G - \frac{n_G}{\tau_e^G}f'_E - \frac{n_G}{\tau_{sp}} - \frac{R_{st}}{1+\varepsilon S_\Sigma}. \tag{21.59}$$

Here, τ_e^E is the escape time of carriers from the ES level to WL, τ_c is the capture time of carriers from the WL to the ES level, τ_{sp} is the spontaneous recombination time, which is assumed to be identical for all levels, τ_e^G is the excitation time of carriers from the GS level to the ES level, and τ_0 is the intradot relaxation time. In Equations 21.57 through 21.59, $f'_{W,E,G} = 1 - f_{W,E,G}$ denote the probabilities of finding an empty carrier state at the WL band edge, the ES and GS levels, respectively, which are closely related to the carrier densities of the corresponding levels (Berg et al., 2001). Finally, material gain is now dependent on carrier density in the GS state. A similar model can be implemented in the case of active regions based on QDashes.

Last, when QDs are embedded into the QW region, it is necessary to extend the model with an additional rate equation that deals with the QW dynamics. Each additional state in QDs also requires a separate rate equation (Kim et al., 2009).

21.4.3.4 Carrier Heating

During the discussion on effects that cause material gain to behave nonlinearly, CH was introduced. Moreover, CH is responsible for phase variation, through Equation 21.19. Depending on the SOA regime of operation, CH might be accounted for phenomenologically, using the nonlinear gain suppression factor, ε, and LEF, α_{CH}, or, in the case of ultrafast applications, a more elaborate model can be used. The temperature dynamics in SOAs can be described using the carrier temperature rate equation (Dailey and Koch, 2007, 2009; Xia and Ghafouri-Shiraz, 2015):

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{1}{\partial U/\partial T}\left(\frac{\mathrm{d}U}{\mathrm{d}t} - \frac{\partial U}{\partial n}\frac{\mathrm{d}n}{\mathrm{d}t}\right) - \frac{T-T_0}{\tau}, \tag{21.60}$$

where T is the carrier temperature, T_0 is the lattice temperature, U is the total carrier plasma energy density, and τ is the electron–phonon interaction time. In this model, electron and hole plasmas are assumed to be equal in both temperature, T, and density, n (Dailey and Koch, 2007). The rate of energy density change can be found to be (Xia and Ghafouri-Shiraz, 2015)

$$\begin{aligned}\frac{\mathrm{d}U}{\mathrm{d}t} &= -v_g\sum_l\left(\hbar\omega_l - E_g\right)g_l\left(S_+^l + S_-^l\right) + v_g\Gamma K_1 n\sum_l \hbar\omega_l\left(S_+^l + S_-^l\right)\\ &\quad - v_g\sum_m\left(\hbar\omega_m - E_g\right)g_m\left(A_+^m + A_-^m\right) + v_g\Gamma K_1 n\sum_m \hbar\omega_m\left(A_+^m + A_-^m\right),\end{aligned} \tag{21.61}$$

where summations are done over each mode l of the two counterpropagating signal photon densities $S^l_{\pm}$, and each mode m of the two counterpropagating ASE noise photon densities $A^m_{\pm}$. The free-carrier and intravalence absorption is captured in ΓK_1. The remaining derivatives, namely, $\partial U/\partial T$ and $\partial U/\partial n$, can be determined from the energy density distribution function (Xia and Ghafouri-Shiraz, 2015), given as (Dailey and Koch, 2009; Coldren et al., 2012)

$$U = \frac{2}{\sqrt{\pi}} k_B T \left(N_c F^c_{3/2} + N_v F^v_{3/2} \right), \tag{21.62}$$

where N_c and N_v stand for the effective densities of states for CB and VB, respectively, and $F^c_{3/2}$ and $F^v_{3/2}$ are the Fermi–Dirac integrals of order 3/2 for CB and VB, respectively. Having determined the rate of temperature change, it is possible to account for its variation in calculating the material gain, radiative spontaneous recombination rate, and variation of index of refraction, which are all temperature dependent through the Fermi–Dirac functions, Equations 21.1 through 21.3, or Equations 21.4 through 21.6, depending on the active region type. The analysis by Xia and Ghafouri-Shiraz (2015) confirms that the difference in carrier density and material gain caused by CH does influence the signal amplification in the picosecond regime of SOA operation.

21.4.3.5 Distributed Bias Current

It is common practice to assume that the driving current of SOA gets instantaneously uniformly distributed along the active region. In practice, this lumped electrode model is not always appropriate. Electric current does require a finite amount of time to travel from the contact to the edges of the electrode, and modeling it as a traveling microwave (TMW) might be more suitable in some cases, especially when the time delay is a significant fraction of the modulation period (Tauber et al., 1994; Mørk et al., 1999; Totović et al., 2015). This distributed nature of the current can have important implications on the results when SOAs are directly modulated at high bitrates (Tauber et al., 1994; Wu et al., 1995; Liljeberg and Bowers, 1997). Moreover, the microwave will be attenuated during the propagation, and the loss can sometimes be very high at typical modulation frequencies (Tauber et al., 1994). An additional point worth noting is that the microwave can be reflected at the end of an electrode, depending on the load impedance. When the microwave reflection coefficient differs from zero, two counterpropagating microwaves will exist (Totović et al., 2015):

$$I(z,t) = \bar{I} + \Delta I_F(z) \exp\left[i\left(2\pi f t - \beta_e z\right)\right] + \Delta I_B(z) \exp\left[i\left(2\pi f t + \beta_e z\right)\right]. \tag{21.63}$$

Here $\bar{I}$ is the stationary value of bias current, ΔI_F and ΔI_B are the small-signal bias current amplitudes for forward (F) and backward (B) propagation with respect to z-axis, which are defined by the spatially variable voltage across the electrode, f is the modulation frequency, $\beta_e = 2\pi f/v_e$ is the microwave propagation constant, $v_e = c/n_e$ is the microwave velocity, and n_e is the effective electric refractive index. Bias current will induce carrier density change through the carrier rate Equation 21.49, which will lead to change in all carrier-dependent parameters.

21.5 Overview of Steady-State and Dynamic TW- and R-SOA Models

Both TW- and R-SOA have found their application niches within the optical networks and photonic circuits. In order to make the optimal choice regarding the SOA material and geometric properties, as well as operating conditions, it is necessary to analyze their performance, in both the steady-state and dynamic regimes. This can be done either experimentally or theoretically, using any of the models presented in this section. Essentially, two borderline approaches can be used depending on required precision, available time, and computational resources. The first one is largely exploited in commercially available software for optical network analysis and relies on a significant number of approximations that either allow for analytical or semianalytical solutions or a relatively simple numerical approach using efficient numerical methods. The second one includes a detailed and comprehensive analysis and is mostly used in SOA design and optimization.

Modeling of steady-state and dynamic properties of SOAs essentially reduces to efficient and accurate self-consistent solving of coupled rate equations written with respect to the counterpropagating signals and noise, and carrier density, with imposed boundary and initial conditions. In the case of TW-SOA, the equation system usually comprises only the forward-propagating TW equation, whereas in the case of R-SOA both forward and backward TW equations are needed. The number of rate equations describing the carrier density depends on the material used for the active region and anticipated model accuracy. A self-consistent approach is required since the carrier depletion in the active region is caused by the stimulated amplification of the input optical signal, and this in turn modifies the material gain and, consequently, further amplification of the signal. If the carrier depletion is weak and the carrier density is approximately uniformly distributed along the active region, the TW equation for the forward-propagating signal (Equation 21.23) can be analytically solved. In the case of the continuous wave (CW), i.e., stationary input optical signal $S_{\text{in}} = S_+(z = 0)$, the output photon density $S_{\text{out}} = S_+(z = L)$ is given by

$$S_{\text{out}} = G_{\text{s}} S_{\text{in}} = S_{\text{in}} \exp\left[\left(\Gamma g - \alpha_{\text{i}}\right) L\right], \tag{21.64}$$

where G_{s} is the single pass gain. An analogous relation can be written for the electric field, where the factor 1/2 should be accounted for in the argument of exponential function: $\exp[(\Gamma g - \alpha_{\text{i}})L/2])$, or $G_{\text{s}}^{1/2}$. In the case of the time-dependent (nonstationary) input optical signal, the solution of Equation 21.23 for the forward-propagating photon density reads

$$S_{\text{out}}(t) = S_{\text{in}}\left(t - \frac{L}{v_{\text{g}}}\right) \exp\left[\left(\Gamma g - \alpha_{\text{i}}\right) L\right]. \tag{21.65}$$

Here, a reference frame traveling with the signal is used, and therefore the temporal dependence of the input signal, $S_{\text{in}}(t)$, is replaced with $S_{\text{in}}(t - L/v_{\text{g}})$. It can be seen that in both cases the output photon density is proportional to G_{s}.

The above relations can be generalized and may also be applied for an SOA section in which the carrier density is uniformly distributed. In this case, S_{in} and S_{out} represent the section's input and output photon densities, ${S_+}^{k-1} = S_+(z_{k-1})$ and ${S_+}^{k} = S_+(z_k)$ for forward propagation, and ${S_-}^{k} = S_-(z_k)$ and ${S_-}^{k-1} = S_-(z_{k-1})$ for backward propagation, respectively, as shown in Figure 21.8. The decrease of the section length makes the assumption of the uniform carrier distribution more justified. This fact represents the basis for the development of various methods, which can provide efficient stationary and time-domain modeling of TW- and R-SOA.

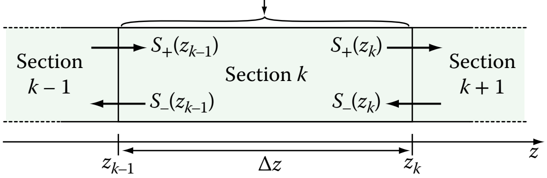


FIGURE 21.8 Schematic representation of the semiconductor optical amplifier (SOA) section with the relevant carrier and photon densities.

21.5.1 Steady-State Models

One of the oldest methods used in modeling of SOAs, optical waveguides, and semiconductor lasers is the transfer matrix method (TMM) (Chu and Ghafouri-Shiraz, 1994), which has long been regarded as an efficient and powerful numerical tool. The robustness and simplicity of its implementation on computer platforms are the major reasons for its success and popularity in using it to analyze complex photonics structures. The method assumes that the entire amplifier is divided into p sections, each with a length of $\Delta z = L/p$, where L is the amplifier length. Each section is labeled such that the kth section extends from $(k-1)\Delta z$ to $k\Delta z$. The value of Δz is small enough such that the carrier density, material gain, and modal gain can be assumed as uniform along each section. If only the forward-propagating wave is of interest, as in TW-SOA, the output photon densities of subsequent sections, S_+^{k-1} and S_+^k, satisfy the relation that is a generalized form of Equation 21.64

$$S_+^k = G_k S_+^{k-1} = S_+^{k-1} \exp\left[\left(\Gamma g_k - \alpha_{\mathrm{i},k}\right)\Delta z\right], \tag{21.66}$$

where g_k and $\alpha_{\mathrm{i},k}$ are the material gain and active region loss corresponding to section k, respectively, and G_k stands for the single pass gain of the kth section. In the case of bidirectional propagation (forward and backward), as in R-SOA or FP-SOA, the matrix relating $(k-1)$th and the kth section may be written in the following form, based on Equation 21.66:

$$\begin{bmatrix} S_+^k \\ S_-^k \end{bmatrix} = \begin{bmatrix} G_k & 0 \\ 0 & G_k^{-1} \end{bmatrix} \begin{bmatrix} S_+^{k-1} \\ S_-^{k-1} \end{bmatrix}. \tag{21.67}$$

Same method can be applied when the TW equation is written with respect to the electric field envelope (Equation 21.15) instead of photon density, intensity, or power. In this case, the transfer matrix may comprise the information about phase (Chu and Ghafouri-Shiraz, 1994):

$$\begin{bmatrix} E_+^k \\ E_-^k \end{bmatrix} = \begin{bmatrix} G_k^{1/2}\exp(i\beta\Delta z) & 0 \\ 0 & G_k^{-1/2}\exp(i\beta\Delta z) \end{bmatrix} \begin{bmatrix} E_+^{k-1} \\ E_-^{k-1} \end{bmatrix}. \tag{21.68}$$

By using input parameters, namely, bias current I and injected optical power P_0, or electric field $E_0 = S_0^{1/2}\exp(i\varphi_0)$, along with boundary conditions and transfer matrices, the carrier density for each section can be found from the Equation 21.14. In the case of R-SOA, the carrier rate equation will be converted into a p-dimensional system of transcendental equations with respect to carrier densities corresponding to all sections (Totović et al., 2014). Once the carrier density in each section is known, the transfer matrices can be used in the evaluation of the photon density or the corresponding electric field envelope in each of the SOA sections.

In this consideration, it is assumed that the photon density, or the corresponding envelope of the electric field, represents the signal. Most of the steady-state models do not account for the noise contribution to the signal, assuming that its phase is uncorrelated to the signal's phase. Therefore, some models neglect the noise (Chu and Ghafouri-Shiraz, 1994), while others use separate TW equations for the signal and the ASE (Jin et al., 2003; de Valicourt et al., 2010), which means that the signal and noise are generally treated separately. However, coupling between the signal and noise is provided by the carrier rate equation that accounts for both contributions, which finally affects the carrier density in the active region and all carrier dependent parameters, and consequently, the photon density. In the case of steady-state analysis, it is possible to solve the TW equation for ASE (Equation 21.31) on domain Δz of section k, and use it in the TMM. The solution is given by

$$a_+^k = G_k a_+^{k-1} + \sigma_k \frac{G_k - 1}{\ln G_k}, \tag{21.69}$$

where a_+^{k-1} and a_+^k represent ASE spectral photon densities at the input and the output of the kth section, respectively, whereas $\sigma_k = 1/(2v_g)\Gamma\beta_{sp}r_{sp}^k\Delta z$ is the contribution of the radiative spontaneous recombination rate spectral density in section k. If the ASE spectrum is uniform over an optical noise bandwidth B_N, the photon density corresponding to ASE can be obtained by multiplying Equation 21.69 by B_N. Otherwise, the total ASE photon density can be found by integrating Equation 21.69 over the ASE bandwidth. The above result can be generalized for both propagation directions (de Valicourt et al., 2010) and written in the following matrix form:

$$\begin{bmatrix} a_+^k \\ a_-^k \end{bmatrix} = \begin{bmatrix} G_k & 0 \\ 0 & G_k^{-1} \end{bmatrix} \begin{bmatrix} a_+^{k-1} \\ a_-^{k-1} \end{bmatrix} + \frac{\sigma_k}{\ln G_k} \begin{bmatrix} G_k - 1 \\ G_k^{-1} - 1 \end{bmatrix}. \tag{21.70}$$

This matrix form does not strictly fit into the standard transfer matrix form due to additional vector representing the ASE contribution of the kth section. However, a considerable number of time-domain multisection methods accounts for ASE by using the form defined by Equation 21.70 (Durhuus et al., 1992; Davis and O'Dowd, 1994; Kim et al., 1999; Occhi et al., 2003; Park et al., 2005; Mathlouthi et al., 2006; Morel and Sharaiha, 2009). Since these methods are not essentially different from TMM, they can be referred to as TMMs.

Although TMM is simple for implementation, it requires a significant number of matrix multiplications, increasing linearly with the number of sections. A large number of multiplications may cause error accumulation and inaccuracy in the evaluation of photon density or the corresponding electric field envelope. The TW equations can be solved by implementation of the FDM rather than the TMM (Connelly, 2001, 2002). Forward differences are used for forward-propagating TWs, and backward differences for backward-propagating TWs. The carrier rate equation in each mesh point is used for determining the carrier density value. Since the self-consistent method implementation is required, the process of calculation of carrier and photon densities relies on iterative procedure, with adjustment of at least one variable in each iteration step, usually the carrier density. However, the carrier density variation along the amplifier is rather small even for a wide range of the signal input powers (Totović et al., 2013). On the other hand, the photon density varies significantly, up to several orders of magnitude. Therefore, the control of the iteration process by the carrier density adjustments, as proposed by Connelly (2001, 2002), makes the process very sensitive since the small variation of the carrier density leads to a large variation of the photon density. It is shown by Jin et al. (2003) that iterations based on photon density may provide faster convergence and increased stability of the iteration process.

21.5.2 Dynamic Models

Due to its simplicity and popularity, TMM has been modified and implemented in the development of the large-signal dynamic model (Davis and O'Dowd, 1994; Kim et al., 1999). The model is bidirectional, which makes it suitable for both TW- and R-SOA time-domain simulations. Similar multisection time-domain models, based on the time-dependent version of Equation 21.67 or Equations 21.68 and 21.70, can also be found in the literature (Occhi et al., 2003; Park et al., 2005; Mathlouthi et al., 2006; Kim et al., 2009). It has been shown by Park et al. (2005) that only the "fullwave" model, which accounts for the intensity and phase of the superimposed electric fields of the signal and noise, provides accurate results. The model based solely on photon densities of signal and noise (or their optical powers) overestimates the influence of the noise on the signal (Park et al., 2005). On the other hand, the model in which the signal is represented by the electric field, and spontaneous emission noise by the photon density, underestimates the effect of noise on the signal (Park et al., 2005).

Another time-domain multisection, wideband, and bidirectional method has been derived using the inverse Fourier transform of the frequency-domain propagation equation (Durhuus et al., 1992). It is essentially a type of TMM, comprising complex time delay. Due to included wideband spectral dependence, it is suitable for simulation of multichannel amplification. A similar method proposed by Morel and Sharaiha

(2009) is implemented in a commercially available software, Advanced Design System from Keysight Technologies (www.keysight.com), after conversion of transfer matrices, including complex time delay, into an equivalent SOA circuit. Another recently developed equivalent circuit model, more precisely transmission line model (TLM), was presented by Xia and Ghafouri-Shiraz (2016). A multisection wideband time-domain method based on the finite-impulse response (FIR) filter scheme has been proposed by Toptchiyski et al. (1999) and Runge et al. (2010). The method is based on two computational steps in each section. The first step includes calculation of local gain with its linear and nonlinear components, as well as the local nonlinear phase change. Then, the propagation equations for the forward- and backward-traveling fields are solved by accounting for the spectral profile of the gain through the implementation of an FIR filter. The FDM can also be applied in time-domain simulations of SOAs. A comprehensive improved finite-difference beam propagation model (IFD-BPM) (Razaghi et al., 2009) has been used for modeling of temporal and spectral properties of copropagating and counterpropagating picosecond optical pulses with different wavelengths. The aforementioned model includes the following effects: interband gain and refractive index dynamics, CH, SHB, two-photon absorption (TPA), ultrafast nonlinear refraction (UNR), gain dispersion, gain peak shift with carrier density variation, and group velocity dispersion (GVD). Due to numerous linear and nonlinear effects, the forward and backward TW equations become essentially nonlinear propagation equations, and are named modified nonlinear Schrödinger equations. Solving these equations is based on coordinate transformation, which can be successfully done by trapezoidal integration and central difference technique. The same method has been also used in the development of the R-SOA pulse propagation model (Connelly, 2012). An interesting method for modeling of SOA time-domain response and four-wave mixing (FWM) was proposed by Mecozzi and Mørk (1997) and Cassioli et al. (2000). The method assumes a relatively small variation of carrier density along the amplifier and consequently a small gain variation. It is based on the analytical integration of material gain over the entire SOA length. This eliminates the spatial coordinate dependence and allows for the amplifier dynamics to be described by solving a set of ordinary differential equations for the complex gain. The ASE noise is modeled by an equivalent noise source with appropriate statistical properties. When the internal loss is included, the method's complexity increases. The SOA simulator is implemented in the Simulink® software, based on the MATLAB® engine and routines for numerical calculations, because of its capability to deal with time-domain analysis of dynamical systems. A similar approach, based on the delayed differential equation, has been used in the derivation of R-SOA's time-domain transmission function, which is then implemented in the semianalytical analysis of FWM in R-SOA (Antonelli and Mecozzi, 2013). The fundamental condition for the method implementation is that the round-trip time of the pulse is small compared to its duration. An improved version of this, the so-called reduced model, is proposed by Dúill and Barry (2015), where the improvement is achieved by the inclusion of the internal losses.

If the dynamic effects of interest occur on a much larger time scale than the time required for the signal to travel along the active region, a simplification to the dynamic model can be introduced by assuming instantaneous propagation of signals across the SOA length (Connelly, 2014, 2015). This assumption significantly simplifies the model by reducing the TW equations to their steady-state form, which can then be solved using the corresponding finite differences for each propagation direction. In this manner, spatial distributions of photon and carrier densities are calculated at each point in time, under the quasi-steady-state conditions. This approach might be helpful as a basis for the simple TW- and R-SOA models; however, a detailed model does require accounting for the finite time needed for wave propagation. In this case, the system of coupled partial differential TW equations may be solved numerically using a first-order upwind scheme based on the FDM (Totović et al., 2011) in order to obtain a full spatiotemporal distribution of all relevant quantities.

21.5.3 Case Study: Steady-State Wideband Self-Consistent Numerical Model

In order to fully understand the mechanisms underlying SOA operation, the model needs to be designed such that it accounts for spectral dependencies of all relevant parameters discussed in Section 21.3 and given in Figures 21.2 through 21.4. These models, typically referred to as wideband, are required to operate not with photon densities, but rather their spectral distributions. Additionally, it is good practice to generalize the model such that it supports any facet reflectivity. This makes the model transparent to the SOA type and it can be further used to analyze TW-, FP-, or R-SOA simply by adjusting the facet reflectivities to the desired value. For TW-SOA, the values can be chosen to be either zero, or, if higher precision is required, some small, but finite value, in the range of 10^{-3}–10^{-5}. In the case of FP- or R-SOA, the reflectivity of a highly reflective facet is usually chosen to be equal to the reflectivity of the interface between cleaved semiconductor and air, $R_2 = (n_r - 1)^2/(n_r + 1)^2$, but it can be set to another value if reflective layers are used.

The full system of equations for a steady-state wideband model can be derived from Equations 21.49 through 21.52 by setting the time derivatives of carrier density and spectral photon densities to zero:

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{I}{qV} - \left(An + R_{\mathrm{sp}} + Cn^3\right) - \frac{R_{\mathrm{st}}}{1 + \varepsilon S_\Sigma} = 0, \tag{21.71}$$

$$\pm\frac{\mathrm{d}s_\pm}{\mathrm{d}z} = \left(\frac{\Gamma g}{1 + \varepsilon S_\Sigma} - \alpha_{\mathrm{i}}\right) s_\pm, \tag{21.72}$$

$$\pm\frac{\mathrm{d}\varphi_\pm}{\mathrm{d}z} = \beta_0 + k_0 \frac{\Gamma \Delta n_r}{1 + \varepsilon S_\Sigma}, \tag{21.73}$$

$$\pm\frac{\mathrm{d}a_\pm}{\mathrm{d}z} = \left(\frac{\Gamma g}{1 + \varepsilon S_\Sigma} - \alpha_{\mathrm{i}}\right) a_\pm + \frac{1}{2v_{\mathrm{g}}}\Gamma \beta_{\mathrm{sp}} r_{\mathrm{sp}}, \tag{21.74}$$

where

$$S_\Sigma(n) = \int_0^\infty \left(s_+ + s_- + a_+ + a_-\right) \mathrm{d}(\hbar\omega), \tag{21.75}$$

$$R_{\mathrm{st}}(n) = \int_0^\infty v_{\mathrm{g}} g \left(s_+ + s_- + a_+ + a_-\right) \mathrm{d}(\hbar\omega). \tag{21.76}$$

Boundary conditions for the signal can be derived from Equations 21.27 and 21.28 and are given by the following set of equations at the front facet:

$$s_+(0) = \left(1 - R_1\right) s_0 + R_1 s_-(0) - 2\sqrt{\left(1 - R_1\right) s_0}\sqrt{R_1 s_-(0)} \cos\left[\varphi_0 - \varphi_-(0)\right], \tag{21.77}$$

$$\varphi_+(0) = \arctan\left[\frac{\sqrt{\left(1 - R_1\right) s_0}\sin\varphi_0 - \sqrt{R_1 s_-(0)}\sin\varphi_-(0)}{\sqrt{\left(1 - R_1\right) s_0}\cos\varphi_0 - \sqrt{R_1 s_-(0)}\cos\varphi_-(0)}\right], \tag{21.78}$$

and by $s_-(L) = R_2 s_+(L)$, and $\varphi_-(L) = \varphi_+(L) + \pi$ at the rear facet. The input signal is assumed to have Gaussian power spectral distribution $\sigma_{\mathrm{G}}(\hbar\omega)$, with FWHM of 0.1 nm, which is a good approximation of the signal generated by the distributed feedback (DFB) laser. Given the input optical power P_0, input signal spectral density can be found by multiplying the input signal photon density, S_0, by the Gaussian function normalized to unity with respect to energy:

$$s_0(\hbar\omega) = \frac{\Gamma(\hbar\omega) P_0}{\hbar\omega v_{\mathrm{g}}(\hbar\omega) WH} \sigma_{\mathrm{G}}(\hbar\omega). \tag{21.79}$$

In analysis of ASE noise, phase is not included explicitly, and therefore no interference exists at either facet, giving $a_+(0) = R_1 a_-(0)$ and $a_-(L) = R_2 a_+(L)$.

In order to account for the resonant properties of the SOA cavity, the procedure described in Section 21.4.2 for deriving Equation 21.34 from Equation 21.31 can be employed, and Equation 21.74 can be transformed into a system of steady-state equations written with respect to the noise photon densities corresponding to each mode m:

$$\pm \frac{\mathrm{d}A_\pm^m}{\mathrm{d}z} = \left(\Gamma_m g_m - \alpha_\mathrm{i}^m\right) A_\pm^m + \frac{1}{2v_\mathrm{g}^m}\Gamma_m \beta_\mathrm{sp}^m R_\mathrm{sp}^m, \tag{21.80}$$

with boundary conditions $A_+^m(0) = R_1 A_-^m(0)$ and $A_-^m(L) = R_2 A_+^m(L)$. The system given by Equation 21.80 represents the steady-state form of Equation 21.34. Introducing the mode photon densities $A_\pm^m$ also implies that modification of Equations 21.75 and 21.76 is required, where the integral over energies for ASE noise now becomes a sum over modes m:

$$S_\Sigma(n) = \int_0^\infty \left(s_+ + s_-\right) \mathrm{d}(\hbar\omega) + \sum_m \left(A_+^m + A_-^m\right), \tag{21.81}$$

$$R_\mathrm{st}(n) = \int_0^\infty v_\mathrm{g} g \left(s_+ + s_-\right) \mathrm{d}(\hbar\omega) + \sum_m v_\mathrm{g}^m g_m \left(A_+^m + A_-^m\right). \tag{21.82}$$

In this manner, signal amplification will be accounted for by TW Equation 21.80 and the photons will be subsequently redistributed within the mode frequency range using the filtering function (Equation 21.38). The resonant and antiresonant frequencies are found by equating the noise phase accumulation to $2m\pi$ and $(2m-1)\pi$, respectively, the same as in Equation 21.33:

$$2k_0\left(n_{\mathrm{eff},0}L + \int_0^L \frac{\Gamma\Delta n_\mathrm{r}}{1+\varepsilon S_\Sigma}\mathrm{d}z\right) = \begin{cases} 2m\pi, & \text{for } \omega_m^\mathrm{r} \\ (2m-1)\pi, & \text{for } \omega_m^\mathrm{a} \end{cases}. \tag{21.83}$$

In the pursuit of increased efficiency, several adjacent modes can be grouped into clusters of $2l+1$ mode, spanning between two antiresonances, from $\hbar\omega_{m-l}^\mathrm{a}$ to $\hbar\omega_{m+l+1}^\mathrm{a}$, where l stands for a small nonnegative integer. All photons corresponding to one cluster are now treated with a single pair of photon densities $A_\pm^m$, centered at the mth mode resonance, $\hbar\omega_m^\mathrm{r}$, where the values of all spectrally dependent parameters in Equations 21.80 through 21.82 are given at $\hbar\omega_m^\mathrm{r}$. In this manner, the number of equation pairs (comprising one equation for each propagation direction) required for modeling the ASE noise is reduced from the number of modes N to the number of clusters $M = N/(2l+1)$. This method effectively reduces the time and resource consumption at the cost of a modest decrease in accuracy. The error in calculating the photon density corresponding to a cluster will be pronounced in the portions of the spectrum where the ASE noise photon count reaches high values. Since the photon count depends on both the radiative spontaneous recombination rate and the material gain among other parameters, this error can be alleviated by nonuniform clustering of modes; the clusters can comprise only one mode in the vicinity of the signal central frequency, where the material gain and stimulated emission rate are high and the number of modes within the cluster can be increased further away in the spectrum.

21.5.3.1 Implementation of the Self-Consistent Method

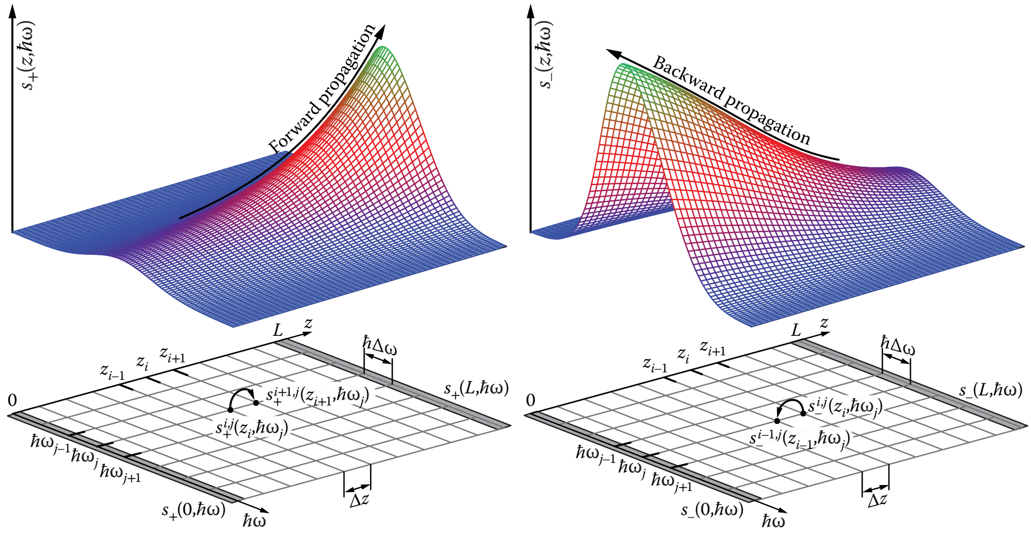
As discussed earlier, the system of equations describing SOA needs to be solved in a self-consistent manner due to the coupling between the signal and noise though the carrier rate equation. Numerical solving of the system requires discretization of the space-spectrum domain into a two-dimensional (2D) mesh, where each point represents a unique pair of position along the longitudinal axis z_i and energy $\hbar\omega_j$. The size of the spectral domain should be chosen such that it covers the region where material gain is positive since in the remaining regions spontaneously emitted photons will be attenuated during propagation and their contribution to the ASE spectrum can be neglected. Step size along the spectral axis should be smaller than the difference between two adjacent resonances of the cavity, which can roughly be estimated from Equation 21.83 as $\Delta(\hbar\omega) = \pi\hbar c/(n_{\mathrm{eff},0}L)$, whereas for the longitudinal axis no particular constraints exist and the smaller step size simply enables a more accurate spatial distribution calculation. The choice of 1001 equidistant points z_i along the longitudinal axis, where i spans from 0 ($z = 0$) to 1000 ($z = L$), provides a quasi-continuous spatial domain, with the step length equal to $\Delta z = L/1000$. An example of such 2D mesh is given in Figure 21.9.

The derivatives over the z-coordinate in Equations 21.72 and 21.73 and Equation 21.80 can now be replaced with finite differences. Depending on the propagation direction of signal and ASE noise, forward or backward, appropriate finite differences are used, forward in the former, and backward in the latter case. This approach allows for successive calculation of the variable values in the next point in space based on the values from the previous one. For forward propagation, the equations can be written in the following generalized form:

$$f_+\left(z_{i+1}\right) = f_+\left(z_i\right) + F_+\left(z_i\right)\Delta z, \tag{21.84}$$

where f_+ stands for any forward-propagating variable, s_+, φ_+, or A_+^m, whereas F_+ denotes the RHS of the corresponding equations, evaluated at the ith point in space, z_i. For backward propagation, using the same generalized notation, the equations can be reduced to

$$f_-\left(z_{i-1}\right) = f_-\left(z_i\right) + F_-\left(z_i\right)\Delta z, \tag{21.85}$$



(a) (b)

FIGURE 21.9 Illustration of the two-dimensional space-spectrum mesh used in the wideband steady-state self-consistent numerical method for modeling semiconductor optical amplifiers (SOAs). The example includes spectral photon density distribution for (a) forward and (b) backward propagations.

where f_- stands for any backward-propagating variable, s_-, φ_-, or A_-^m, whereas F_- denotes the RHS of the corresponding equations, evaluated at the ith point in space, z_i. The coupling between Equations 21.84 and 21.85 is provided by the boundary conditions at $z = 0$ and $z = L$. Carrier density in each point in space can be found as a solution of the transcendental Equation 21.71, which is then followed by calculation of resonant and antiresonant frequencies, using Equation 21.83.

The system given by Equations 21.71 through 21.73, 21.80, and 21.83 represents the basis of one iteration step in the self-consistent numerical method (SCNM). Starting from the arbitrarily chosen carrier density distribution, signal and noise for both propagation directions can be determined and subsequently used for calculating the new carrier distribution. Assuming that the newly calculated carrier density distribution converges to the steady-state distribution, repeating this process should give improved results with each iteration step. The process can be stopped when the difference between the values calculated within two consecutive iteration steps reaches the tolerance point. Although this simplified algorithm is useful for conceptualization of iterative procedure, in practice it suffers from severe instability, and rarely converges. The instability is caused by significant changes in photon spectral densities, up to several orders of magnitude, for subtle changes in carrier density. An approach typically used to alleviate this issue relies on self-consisting tuning of the carrier density (Connelly, 2001) or photon density (Jin et al., 2003) by its averaging over the current and previous iteration step, with proper weight coefficients. Although these approaches have shown to be successful in modeling TW-SOA, in the case of an R-SOA they still diverge for high driving currents and high input signal optical powers. Choosing to simultaneously tune all variables of interest, namely $s_\pm$, $A_\pm^m$, and n, instead of just one, does increase stability, but not sufficiently enough to provide a reliable algorithm. The main issue comes from the interplay between counterpropagating signals and noise, where the backward-propagating ones are the delayed replicas of their forward-propagating counterparts. This "echoing," present whenever either facet reflectivity is nonzero, destabilizes the algorithm and raises the need for restraining the variables through the fading memory of previous iteration steps. Instead of averaging the variables over the current and previous iteration steps, the generalized algorithm assumes inclusion of k previous iteration steps, where $k \geq 1$. In this manner, sudden changes are prevented, and the system steadily converges toward the solution. The contribution of each iteration step can be adjusted using the set of weight coefficients w_i, with a sum of unity, which ponder the results of the ith iteration step during the averaging process. The updated (averaged) iteration variable that is passed to the next iteration step, x_{i+1}, can be found as

$$x_{i+1} = w_i x_i + w_{i-1} x_{i-1} + \ldots + w_{i-k} x_{i-k}. \tag{21.86}$$

The optimal number of iteration steps included will depend on the device's properties, mainly on the front and rear facet reflection coefficients. It has been shown by Totović et al. (2013) that $k = 3$ previous iteration steps, in addition to the current one, suffice to provide a stable algorithm, with a set of weight coefficients $\{w_i, w_{i-1}, w_{i-2}, w_{i-3}\} = \{0.4, 0.3, 0.2, 0.1\}$. The choice of weight coefficients depends on desired robustness of the algorithm and the convergence speed. By using higher pondering values for the current iteration step relative to the previous ones, the iteration process becomes more time efficient at the cost of an increased risk for divergence. Inclusion of several previous iteration steps is required in modeling FP- and R-SOA but can be relaxed to $k = 1$ for TW-SOA.

Prior to entering the iteration procedure, the maximum tolerable relative error $\delta_{\max}$ needs to be defined. The error δ between the two consecutive iteration steps is calculated as a maximum value of relative errors among all monitored variables, for each point in the mesh $(z_i, \hbar\omega_j)$:

$$\delta = \max\left\{\delta(n), \delta\left(s_\pm\right), \delta\left(A_\pm^m\right), \delta\left(\hbar\omega_m^{\mathrm{r}}\right)\right\}. \tag{21.87}$$

The iteration procedure is stopped once the value of the current relative error δ drops below the defined tolerance $\delta_{\max}$. The meticulousness with which the iteration error δ is calculated during each iteration step ensures that all results consistently need to converge in order for the iteration procedure to end. Setting $\delta_{\max}$ to 10^{-3} imposes very strict convergence requirements for all variables.

Aside from the input and control parameters, i.e., I, P_0, weight coefficients, and $\delta_{\max}$, SCNM requires a set of initial guess values (IGVs) for all variables. Since the spectral photon density of the input signal, s_0, depends on the confinement factor as shown in Equation 21.79, its IGV is calculated for an arbitrary carrier density between the transparency value n_{tr} and the maximum allowed value, $n_{\max}$, governed by the bias current. The transparency value of the carrier density is determined by equating material gain to zero, whereas the maximum allowed value can be found from Equation 21.71, under the assumption of zero photon density in the active region:

$$\frac{I}{qV} - \left[An_{\max} + R_{\text{sp}}\left(n_{\max}\right) + Cn_{\max}^3\right] = 0. \tag{21.88}$$

The IGV of spectral photon density for the forward-propagating signal, s_+, is then set to the input signal spectral photon density s_0, whereas the IGV for s_- is set either to zero, for TW-SOA, or to s_0, for FP- and R-SOA. The IGV for phase in both propagation directions, $\varphi_\pm$, is chosen to be zero. Before setting IGVs for ASE noise, it is necessary to determine the number of analyzed modes and the type of clustering. For a device length between 500 and 1000 μm, the intermodal space is of the order of 0.5 nm, and $N = 600$ modes covers the portion of the spectrum where ASE is expected to be pronounced. For the 100 modes around the signal central frequency, clusters of 1 mode have been chosen, whereas in the rest of the spectrum, clusters consist of 5 modes. This gives a total of $M = 200$ clusters to be analyzed. The ASE noise is usually orders of magnitude lower than the signal, therefore IGVs for all clusters $A_\pm^m$ are set to zero. The IGV for carrier density is calculated based on the IGVs for signal and noise photon densities by solving the transcendental Equation 21.71 with respect to n at each point z_i along the discretized longitudinal axis. Using Equation 21.83, IGVs for resonant and antiresonant frequencies are determined. Finally, using the appropriate expressions depending on the active region type, bulk or MQW, all material parameters dependent on carrier density can be determined. These include material gain (Equation 21.1 or Equation 21.4), radiative spontaneous recombination rate (Equation 21.2 or Equation 21.5), refractive index variation (Equation 21.3 or Equation 21.6), confinement factor (Equation 21.9), group velocity (Equation 21.13), and spontaneous emission coupling factor (Equation 21.11). Once all IGVs are defined, the iteration procedure can be initialized.

As discussed earlier, the iteration procedure essentially relies on repeating one iteration step and updating variable values according to the current and k previous iteration steps using Equation 21.86. It should be noted that until the number of completed iteration steps reaches k, corresponding IGVs are used as a replacement for the missing previous iteration values. The iteration step starts with solving Equations 21.72, 21.73, and 21.80, for forward propagation and determining s_+, φ_+, and A_+^m, based on the initial conditions at $z = 0$. This is followed by adjusting the variable values by averaging them over previous iteration steps and applying boundary conditions at $z = L$. Next, using Equations 21.72, 21.73, and 21.80, backward propagation is simulated and s_-, φ_-, and A_-^m are determined, and subsequently adjusted by averaging their values over k previous iteration steps. The total photon density in the active region, S_Σ, and the rate of stimulated emission, R_{st}, are found from Equations 21.81 and 21.82, respectively, which gives the basis for determining carrier density distribution using Equation 21.71. After adjusting n by averaging it over previous iteration steps, all parameters dependent on carrier density can be calculated. Finally, at the end of the iteration step, the maximum relative error between two consecutive iteration steps is calculated using Equation 21.87 and compared with the tolerance $\delta_{\max}$. If the tolerance is satisfied, the iteration process has successfully converged to the solution; otherwise the iteration process continues. The algorithm described for SCNM is summarized in Figure 21.10.

When the iteration process successfully converges to the steady-state solution, normalized filtering function of the cavity $T_m(\hbar\omega)$ can be found from Equation 21.38, and the ASE noise spectral photon

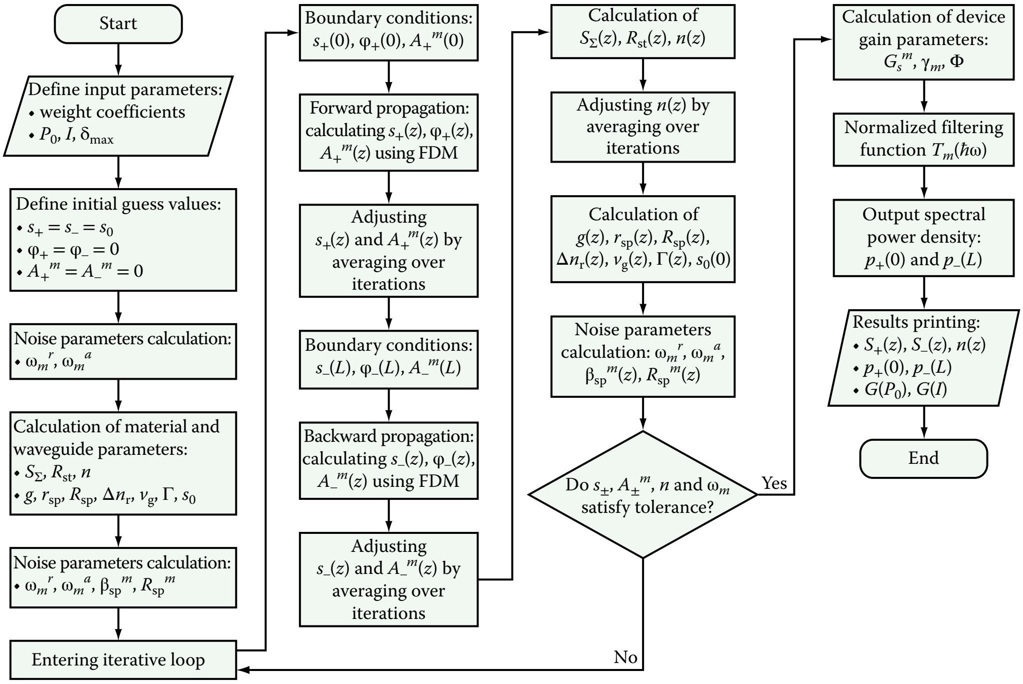

Start
Define input parameters:
• weight coefficients
• P_0, I, $\delta_{\max}$
Define initial guess values:
• $s_+ = s_- = s_0$
• $\varphi_+ = \varphi_- = 0$
• $A_+^m = A_-^m = 0$
Noise parameters calculation:
• ω_m^r, ω_m^a
Calculation of material and waveguide parameters:
• S_Σ, R_{st}, n
• g, r_{sp}, R_{sp}, Δn_r, v_g, Γ, s_0
Noise parameters calculation:
• ω_m^r, ω_m^a, β_{sp}^m, R_{sp}^m
Entering iterative loop
Boundary conditions: $s_+(0)$, $\varphi_+(0)$, $A_+^m(0)$
Forward propagation: calculating $s_+(z)$, $\varphi_+(z)$, $A_+^m(z)$ using FDM
Adjusting $s_+(z)$ and $A_+^m(z)$ by averaging over iterations
Boundary conditions: $s_-(L)$, $\varphi_-(L)$, $A_-^m(L)$
Backward propagation: calculating $s_-(z)$, $\varphi_-(z)$, $A_-^m(z)$ using FDM
Adjusting $s_-(z)$ and $A_-^m(z)$ by averaging over iterations
Calculation of $S_\Sigma(z)$, $R_{st}(z)$, $n(z)$
Adjusting $n(z)$ by averaging over iterations
Calculation of $g(z)$, $r_{sp}(z)$, $R_{sp}(z)$, $\Delta n_r(z)$, $v_g(z)$, $\Gamma(z)$, $s_0(0)$
Noise parameters calculation: ω_m^r, ω_m^a, $\beta_{sp}^m(z)$, $R_{sp}^m(z)$
Do $s_\pm$, $A_\pm^m$, n and ω_m satisfy tolerance?
Yes
No
Calculation of device gain parameters: G_s^m, γ_m, Φ
Normalized filtering function $T_m(\hbar\omega)$
Output spectral power density: $p_+(0)$ and $p_-(L)$
Results printing:
• $S_+(z)$, $S_-(z)$, $n(z)$
• $p_+(0)$, $p_-(L)$
• $G(P_0)$, $G(I)$
End

FIGURE 21.10 Flowchart of the steady-state wideband self-consistent numerical method (SCNM). FDM, finite difference method.

densities determined using

$$a_+(L,\hbar\omega) = \sum_{m=1}^{M}\left[\left(\hbar\omega^{a}_{m+l+1} - \hbar\omega^{a}_{m-l}\right)^{-1} A_+^m(L)\,T_m(\hbar\omega)\right], \tag{21.89}$$

$$a_-(0,\hbar\omega) = \sum_{m=1}^{M}\left[\left(\hbar\omega^{a}_{m+l+1} - \hbar\omega^{a}_{m-l}\right)^{-1} A_-^m(0)\,T_m(\hbar\omega)\right]. \tag{21.90}$$

Finally, the output spectral power densities of signal and noise combined are calculated as

$$p_+(L,\hbar\omega) = \left[1 - R_2(\hbar\omega)\right]\left[s_+(L,\hbar\omega) + a_+(L,\hbar\omega)\right] v_g(\hbar\omega)\, WH\hbar\omega \Big/ \Gamma(\hbar\omega), \tag{21.91}$$

$$p_-(0,\hbar\omega) = \left[1 - R_1(\hbar\omega)\right]\left[s_-(0,\hbar\omega) + a_-(0,\hbar\omega)\right] v_g(\hbar\omega)\, WH\hbar\omega \Big/ \Gamma(\hbar\omega), \tag{21.92}$$

where the signal spectral distribution is calculated using Equations 21.84 and 21.85, whereas the noise spectral distribution is determined from Equations 21.89 and 21.90. It should be noted that, if reflectivities of both facets are chosen to be zero, which is an idealized case of TW-SOA, the previous algorithm can be simplified. After determining and adjusting the forward-propagating variables, the following step, related to backward propagation, can be reduced to calculation and adjusting of only A_-^m since the backward-propagating signal will not exist.

One of the most valuable figures of merit for SOA characterization is the device (transmission) gain G. Strictly speaking, it is defined as the ratio of the signal's output and input photon density, or their corresponding powers, i.e., $G = 10\log_{10}(P_{out}/P_0)$. However, as the signal output photon density cannot be separated from ASE noise in the signal's wavelength range, the output photon density usually includes a small fraction of ASE, which has a negligible effect on G, as long as the input optical power P_0 is sufficiently high (above −40 dBm). For TW-SOA, the output photon density is recorded at the rear facet ($z = L$), $P_{out} = \int p_+(L,\hbar\omega)\mathrm{d}(\hbar\omega)$, whereas for the R-SOA, at the front facet ($z = 0$), $P_{\text{out}} = \int p_-(0, \hbar\omega)\mathrm{d}(\hbar\omega)$. For FP-SOA, either facet can be regarded as the output one, depending on system configuration. In the case when the SOA model bundles up signal and the whole ASE noise spectrum into one quantity, the above definition of device gain G is meaningful for somewhat larger input powers, which can be recognized by the relatively flat dependence of G versus P_0. For insufficient input powers P_0, the device gain tends to increase more rapidly as the input optical power decreases, ultimately reaching the infinite value for the zero-input signal.

21.5.3.2 Results and Discussion

Based on the wideband self-consistent numerical model presented in earlier section, a full analysis of TW- and R-SOA is carried out under the steady-state operation regime. The list of parameters used in the analysis is given in Table 21.3. The remaining parameters, which are both carrier and energy dependent, including g, r_{sp}, Δn_{r}, Γ, v_{g}, and β_{sp}, are determined using the models presented in Section 21.3 and are given in Figures 21.2 and 21.4. Both bulk and MQW-based active regions are optimized to be polarization insensitive, as shown in Figure 21.3, so only TE polarization is analyzed.

21.5.3.2.1 Steady-State Signal and Carrier Densities Spatial Distribution

Figures 21.11 and 21.12 show spatial distribution of carrier densities, $n(z)$, and photon densities of forward and backward-propagating signal and ASE noise combined, $\int[s_+(z, \hbar\omega) + a_+(z, \hbar\omega)]\mathrm{d}(\hbar\omega)$, and $\int[s_-(z, \hbar\omega) + a_-(z, \hbar\omega)]\mathrm{d}(\hbar\omega)$, respectively, for different input optical powers P_0. In Figure 21.11, the distributions are shown for the active region based on the unstrained bulk $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$, with the parameters listed in Table 21.1, whereas in Figure 21.12, the results are given for MQW consisting of six coupled 0.13% tensile-strained $\text{In}_{0.516}\text{Ga}_{0.484}\text{As}$ quantum wells, which are strain-compensated by the $\text{In}_{0.9}\text{Ga}_{0.1}\text{As}_{0.3}\text{P}_{0.7}$ 0.26% compressively strained barriers, with the parameters given in Table 21.2. In both active region types, carrier dynamics is treated using Equation 21.71. Although this is a somewhat simplified model for MQW SOA, its usage is justified when the volumes of the well and barrier regions do not differ significantly. Namely, the carrier recombination time in the barrier, τ_{b}, figuring in the full system of carrier rate equations describing MQW active region, Equations 21.54 and 21.55, is of the order of nanoseconds, unlike the capture time by the wells τ_{bw}, which has the value between 25 and 55 ps (Keating et al., 1999). On the other hand, it has been shown by Tsai et al. (1995) that the escape time from the wells, τ_{wb}, is one to two orders of magnitude higher in comparison with τ_{bw}. Assuming that the volumes of the well and barrier regions, V_{w} and V_{b}, respectively, are similar, the most significant contribution to $\mathrm{d}n_{\text{b}}/\mathrm{d}t$ comes from the first and the third terms on the RHS of Equation 21.54, whereas the second and fourth terms can be neglected. In

TABLE 21.3 SOA Material and Geometric Parameters

Symbol	Quantity	Value
W	Active region width	2 μm
H	Bulk active region height	140 nm
L_{w}	MQW well thickness	19 nm
L_{b}	MQW barrier thickness	10 nm
H	MQW active region height for six QWs	114 nm
L	Active region length	600 μm
K_0	Carrier independent loss coefficient	$62\ \text{cm}^{-1}$
K_1	Carrier dependent loss coefficient	$7.5 \times 10^{-17}\ \text{cm}^2$
A	Shockley–Read–Hall coefficient	$1.1 \times 10^{8}\text{s}^{-1}$
C	Auger coefficient	$5.82 \times 10^{-29}\ \text{cm}^6/\text{s}$
ε	Nonlinear gain suppression factor	$1.5 \times 10^{-17}\ \text{cm}^3$
$\hbar\omega_0$	Signal central energy	0.8 eV
R_1	Antireflective facet power reflectivity	5×10^{-5}

SOA, semiconductor optical amplifier; MQW, multiple quantum well.

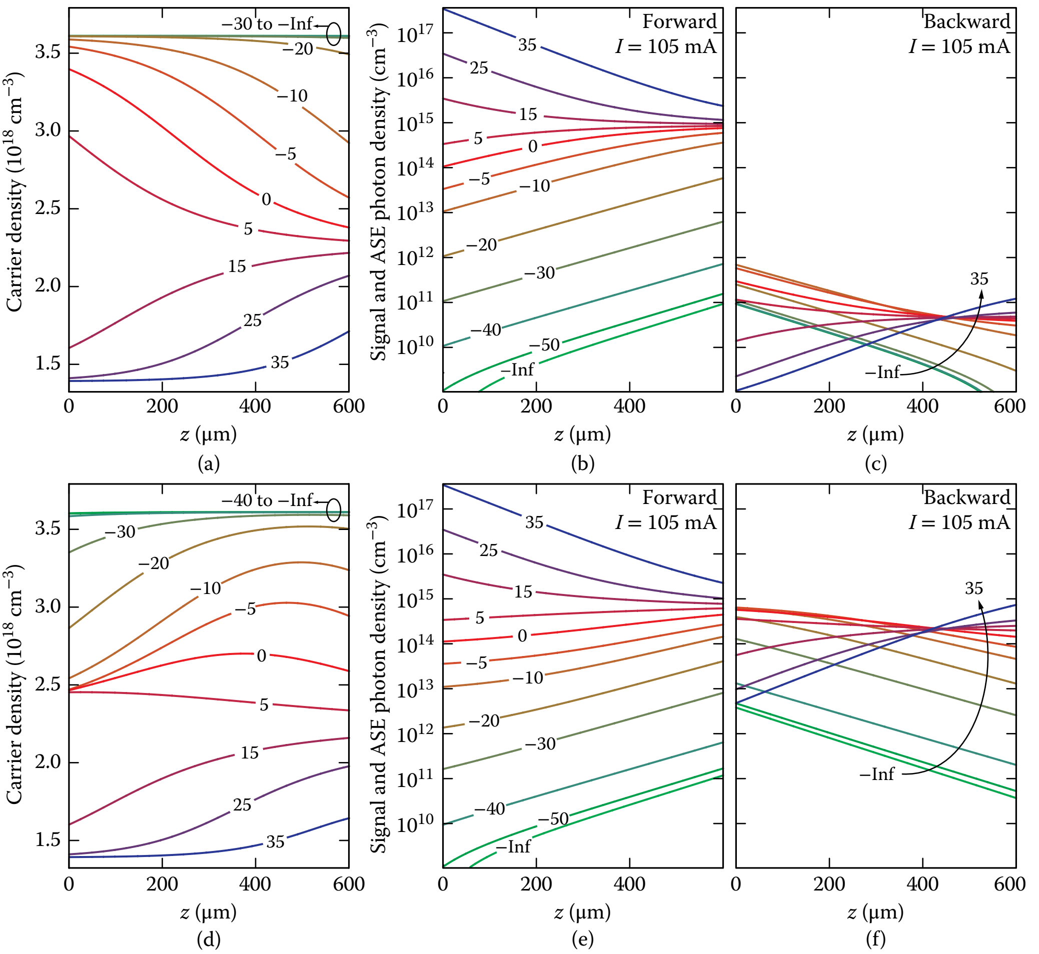


FIGURE 21.11 Spatial distribution of (a), (d) carrier density, and total photon density (signal and amplified spontaneous emission [ASE]) for (b), (e) forward, and (c), (f) backward propagation for different input signal optical powers (given in dBm) in the case of (a) through (c) traveling-wave semiconductor optical amplifier (TW-SOA), and (d) through (f) reflective (R)-SOA, both with the active region based on the unstrained bulk material.

this manner, the generation term $\eta_{\text{inj}}I/(qV_{\text{b}})$ directly influences the carrier density in the barrier states, n_{b}, which is further reflected in the first term of Equation 21.55. This effect becomes even more pronounced if the capture time is short and the steady-state value of n_{b} is quickly reached.

It can be seen that for low input optical powers, up to −30 dBm, the carrier density has values close to the maximum allowed by the available bias current, $n_{\max}$, due to the low consumption by vague optical signals. As P_0 increases, it dumps the carrier density overall and leads to its nonuniform spatial distribution caused by the intensified consumption of carriers. For both SOA types, forward-propagating photon densities have similar values and spatial distributions, as shown in Figure 21.11b and e. The signals are exponentially amplified for low to moderate input optical powers, up to 5 dBm. A further increase of P_0 leads to significant carrier depletion close to the front facet, Figure 21.11a and d, and consequently exponential signal attenuation. The backward-propagating photon densities, Figure 21.11c and f, are orders of magnitude lower for TW-SOA than for R-SOA, which is expected due to very low residual reflectivity of the rear facet. However, spatial distributions are qualitatively similar, and resemble the reversed forward-propagating ones, exponentially amplified for low P_0, and attenuated as P_0 increases. The existence of strong backward-propagating signal in the case of R-SOA leads to a different distribution of $n(z)$

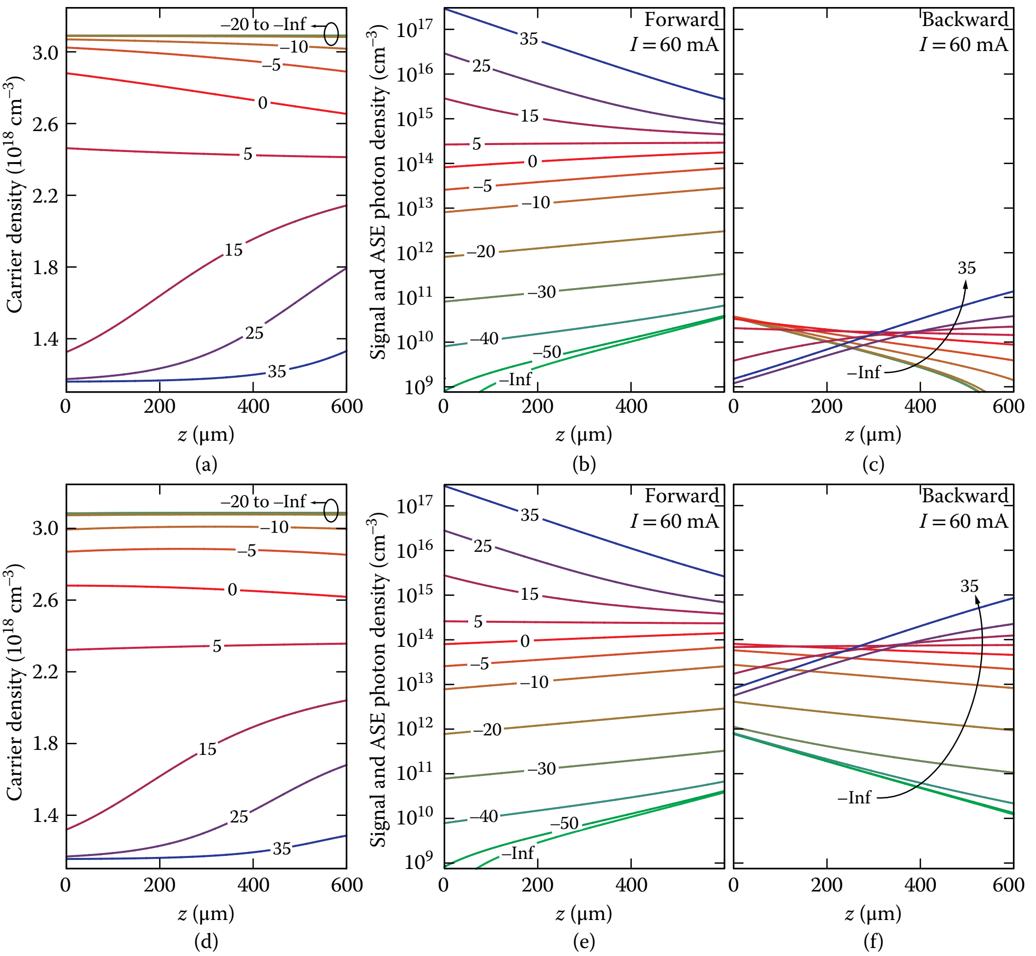


FIGURE 21.12 Spatial distribution of (a, d) carrier density, and total photon density (signal and amplified spontaneous emission [ASE]) for (b, e) forward and (c, f) backward propagation for different input signal optical powers (given in dBm) in the case of (a) through (c) traveling-wave semiconductor optical amplifier (TW-SOA), and (d) through (f) reflective (R)-SOA, both with the active region based on the tensile-strained quantum wells, which are strain-compensated by the compressively strained barriers.

for moderate P_0, between −20 and 5 dBm, in comparison to TW-SOA. The carrier density is more severely depleted closer to the front facet, by the amplified $S_{-}(z)$, as shown in Figure 21.11d, and has a maximum close to the rear facet, caused by the interplay of the counterpropagating signals.

Spatial distributions of carrier and photon densities for MQW-based SOAs, as shown in Figure 21.12, resemble the ones shown for bulk SOAs, as shown in Figure 21.11, and same qualitative analysis can be applied. The main difference comes from lower carrier density, and material gain, due to the polarization-insensitive design and lower bias current. This leads to the more uniform carrier spatial distribution, as shown in Figure 21.12c and d, and somewhat lower device gain, which can be inferred from the values of the output photon densities, at $z = L$ for TW-SOA, and $z = 0$ for R-SOA.

21.5.3.2.2 Steady-State Device Gain

One of the most valuable figures-of-merit for SOA is its steady-state device (transmission) gain. As defined in earlier section, it is the ratio of the output and input optical powers, or photon densities. Figures 21.13 and 21.14 show the dependence of device gain on operating conditions, i.e., bias current I, and input optical power P_0, for bulk, and MQW active region types, respectively, for both TW- and R-SOA.

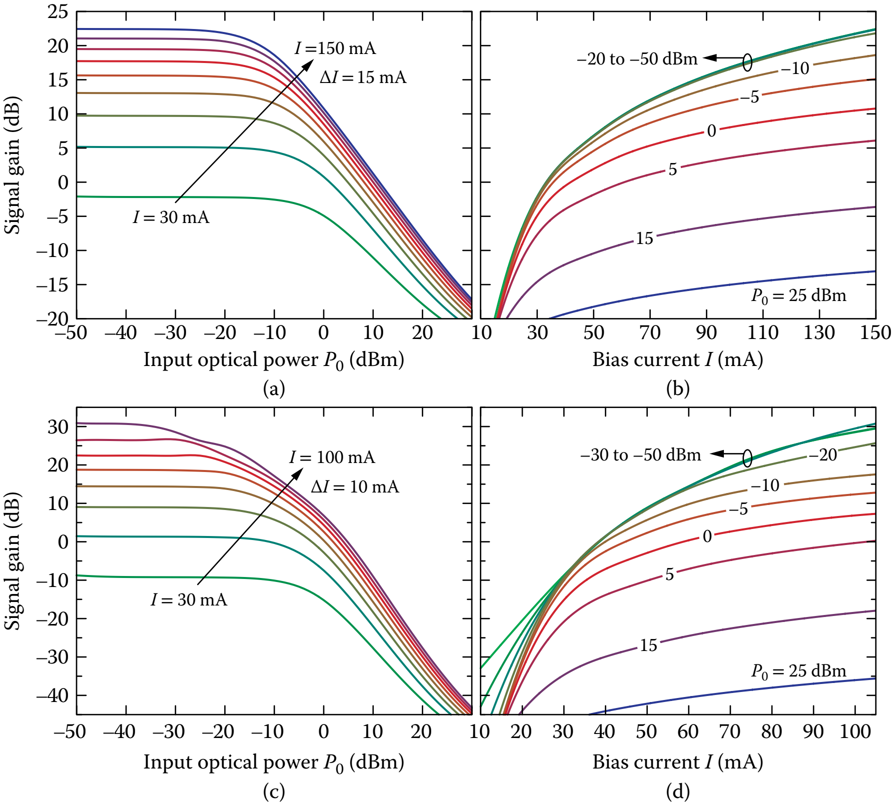


FIGURE 21.13 Device (transmission) gain for bulk active region versus (a), (c) input optical power, and (b), (d) bias current, for (a), (b) traveling-wave semiconductor optical amplifier (TW-SOA), and (c), (d) reflective (R)-SOA.

Figure 21.13 shows that the increase in bias current leads to the higher gain in both SOA types since more carriers are available to provide signal amplification, whereas the increase in P_0 leads to the gain decrease due to excessive carrier consumption. For low to moderate P_0, up to approximately −10 dBm, the device gain does not change significantly with input power and is limited by the length of the active region that is responsible for amplification. Additionally, R-SOA exhibits higher gain, Figure 21.13c and d in comparison with TW-SOA, Figure 21.13a and b, for bias currents beyond 50 mA. In contrast, for the currents below 50 mA, the gain is lower in R-SOA compared with TW-SOA. This comes from double signal propagation through the active region in the case of R-SOA, which emphasizes either amplification, when carrier density is high, or attenuation, when n is low. The maximum operating current of SOA is limited either by maximum bias current density, which provides long-term stable operation, or by the threshold for stimulated emission. In the latter case, the threshold material gain g_{th} can be determined from (Coldren et al., 2012)

$$\Gamma g_{th} - \left(K_0 + \Gamma K_1 n_{th}\right) = \frac{1}{2L} \ln \frac{1}{R_1 R_2}. \tag{21.93}$$

Having the value for g_{th}, it is possible to determine n_{th}, and using Equation 21.71, the corresponding threshold current value. Since the product of the two reflectivities is higher for R-SOA than for TW-SOA, the g_{th} will be lower, and, therefore, the current threshold value will also be lower.

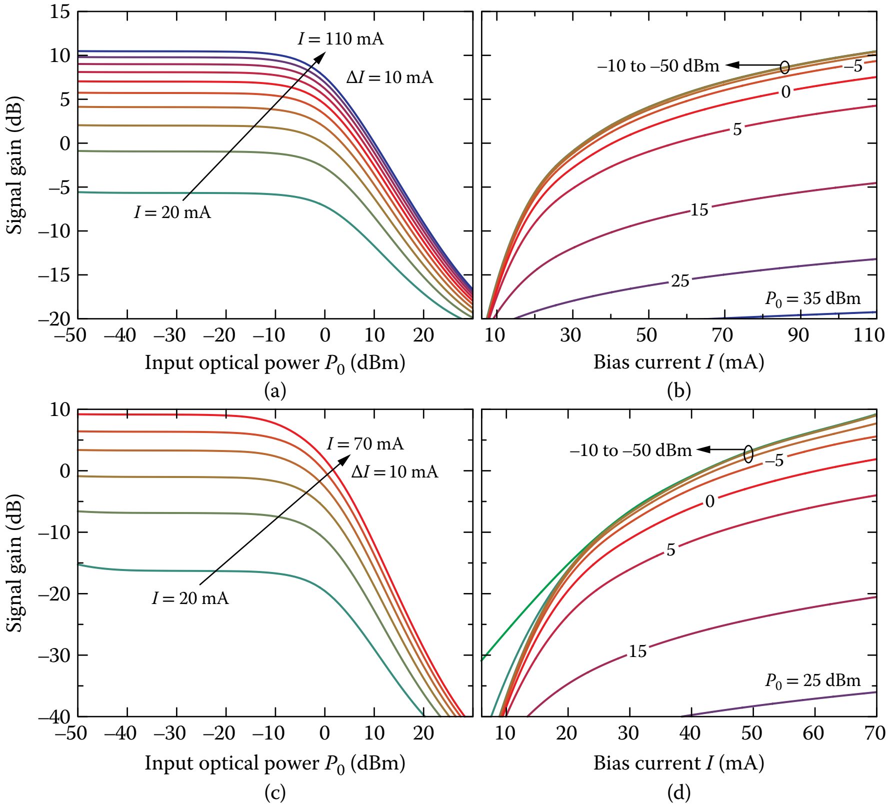


FIGURE 21.14 Device (transmission) gain for the multiple quantum well (MQW) active region versus (a), (c) input optical power, and (b), (d) bias current, for (a), (b) traveling-wave semiconductor optical amplifier (TW-SOA), and (c), (d) reflective (R)-SOA.

An interesting feature in the R-SOA's gain dependence on input optical power can be observed for high bias currents, as shown in Figure 21.13c. As the saturation regime of operation approaches, when the gain begins to decrease, an overshoot in device gain is visible. This effect has already been reported (Schrenk, 2011; Totović et al., 2013) and can be explained by the trade-off between a sudden decrease in the spontaneous emission and ASE with the decrease of carrier density and the corresponding relatively slow decrease in the material gain at the signal wavelength (Totović et al., 2013).

Figure 21.14 shows that MQW SOA provides somewhat lower gain in comparison with bulk, for the same operating conditions. This can be attributed to the lower modal gain, coming from lower confinement factor at the signal central frequency, as shown in Figure 21.4, which is a consequence of the polarization-insensitive design. Nevertheless, qualitatively, the two active regions give similar results, where the device gain increases with bias current increase. For low input optical powers, gain does not vary with P_0, and as the input optical power increases and carrier consumption becomes prominent, the device enters the saturation regime, and gain begins to decrease.

21.5.4 Case Study: Steady-State Semianalytical Model

When simulation time and computational resources are the priority, another approach can be used in SOA analysis. The main underlying idea of the semianalytical approach is to harvest the insightfulness of analytical methods and the versatility of numerical methods to reach a fast and computationally efficient model.

It essentially relies on carefully chosen approximations that can reduce the model complexity but still keep the accuracy at as high a level as possible. Since the carrier density usually does not vary significantly across the active region, it is common practice to model it as a constant, or a piecewise constant function. This approximation is extensively used in TMM, and, as previously discussed, presents the basis for many different SOA models. Choosing a fixed value for $n(z)$ simplifies TW equations to the point where an analytical solution can be found, either in closed form, e.g., Equation 21.64, or in the form of a transcendental equation. The carrier density's spatial distribution can also be approximated by a linear function, but this approach may give an analytical solution for photon densities only when nonlinear gain suppression is neglected (Totović et al., 2014).

Unlike carrier density, signal and noise photon densities can vary up to several orders of magnitude during the propagation from one SOA facet to the other. This implies that small variations in n can build up to significant changes in signal and noise, and special attention is required in selecting the adequate carrier density value. For example, choosing the value at the front or rear facet, $n_0 = n(z = 0)$ or $n_L = n(z = L)$, respectively, can significantly over- or underestimate material gain and spontaneous emission rate and lead to noticeable error in the output photon density, and, consequently, device gain. In order to alleviate this discrepancy, an average value of these two carrier densities can be chosen, $\bar{n} = (n_0 + n_L)/2$, and thus the error caused by neglecting the spatial dependence of n can be partially compensated.

The requirements for increased model efficiency can be met by yet another model approximation, where the signal and noise spectral dependences are neglected and the total spontaneous emission and signal are assigned to a single wavelength corresponding to signal peak, providing model analysis in terms of photon densities rather than spectral photon densities. Since no spectral analysis is present, TW equations written with respect to signal and noise can be wrapped up into a single equation that accounts for both spontaneous and stimulated emission of photons. This equation can be derived as a sum of Equations 21.72 and 21.74, integrated over the whole spectrum, resulting in

$$\pm\frac{\mathrm{d}S_\pm}{\mathrm{d}z} = \left[\frac{\Gamma g}{1+\varepsilon\left(S_+ + S_-\right)} - \alpha_\mathrm{i}\right] S_\pm + \frac{1}{2v_\mathrm{g}}\Gamma\beta_\mathrm{sp}R_\mathrm{sp}, \tag{21.94}$$

where $S_\pm$ now stands for the photon density of the signal and noise combined, at the signal central frequency ω_0. All parameters that are dependent on carrier density and photon energy in Equation 21.94, namely, $\Gamma, g, \alpha_\mathrm{i}, v_\mathrm{g}$, and β_sp, are now replaced with the corresponding numerical dependencies on n, interpolated from the full spectral and carrier dependent model presented in Figures 21.2 and 21.4, for the energy of $\hbar\omega_0$. Additionally, for parameters that do not vary significantly with n, i.e., Γ, α_i, v_g, and β_sp, as shown in Figure 21.4, a fixed value can be used, interpolated for an average carrier density, n_av, between the transparency, n_tr, and maximum allowed value, n_max, given by Equation 21.88. The carrier density at transparency will depend only on material gain, and not on operating conditions, whereas the maximum allowed value for carrier density n_max will depend on the value of bias current. In order to choose the adequate value of n_av for parameter interpolation, it is useful to determine n_max for maximum allowed bias current I_max. The value of I_max can be calculated as $I_\mathrm{max} = J_\mathrm{max}WL$, for a cross-section area defined by the active region's width W and length L, given in Table 21.3, where J_max is the maximum allowed bias current density for the long-term stable operation, which is in the range of 25–35 kA/cm^2. Under the listed conditions, the chosen values for bulk and MQW-SOA are $n_\mathrm{av} = 2.5 \times 10^{18}$ cm^{-3} and $n_\mathrm{av} = 2.1 \times 10^{18}$ cm^{-3}, respectively. The interpolated values of Γ, v_g, β_sp, and additional parameters are listed in Table 21.4 for bulk and MQW at the signal central energy of $\hbar\omega_0 = 0.8$ eV.

If a simpler model is used, particularly if an analytical solution is sought, it proves to be impractical to use numerical dependencies of g and R_sp on n (Figure 21.2). Rather, numerical results can be fitted to analytical functions, which further simplifies calculation. The most common choice for $g(n)$ dependence is a linear one, $g(n) = \alpha(n - n_\mathrm{tr})$, where α is the differential gain (Thylén, 1988; Shtaif et al., 1998; Mørk et al., 1999). This approximation is justified if the bias current is low enough to prevent a significant increase in carrier

TABLE 21.4 Interpolated Parameter Values for Bulk and Multiple Quantum Well (MQW) Active Region at the Signal Central Energy $\hbar\omega_0 = 0.8$ eV and for the Average Carrier Density of $n_{av} = 2.5 \times 10^{18}$ cm^{-3} for Bulk and $n_{av} = 2.1 \times 10^{18}$ cm^{-3} for MQW

Symbol	Quantity	Bulk	MQW
Γ	Optical confinement factor	30.44%	21.62%
v_g	Group velocity	7.76×10^7 m/s	8.73×10^7 m/s
k_0	Wave vector	4.05×10^6 m^{-1}	4.05×10^6 m^{-1}
n_r	Index of refraction	3.660	3.597
Δn_r	Variation of index of refraction	0.071	7.44×10^{-4}
$n_{\text{eff},0}$	Effective index of refraction for zero carrier density	3.405	3.291
n_{eff}	Effective index of refraction	3.490	3.291
β_0	Wave propagation constant for zero carrier density	1.38×10^7 m^{-1}	1.33×10^7 m^{-1}
R_2	Highly reflective facet power reflectivity	0.3258	0.3192
n_{sp}	Inversion factor	1.035	1.055
β_{sp}	Spontaneous emission coupling factor	1.842×10^{-5}	1.869×10^{-5}

All values given are for transverse electric (TE) polarization.

TABLE 21.5 Bulk and Multiple Quantum Well (MQW) Active Region Fitting Parameters at the Signal Central Energy $\hbar\omega_0 = 0.8$ eV for Material Gain and Spontaneous Emission Rate to the two-Parameter Logarithmic and Second-Degree Polynomial Function of Carrier Density, Respectively

Symbol	Quantity	Bulk	MQW
g_0	Material gain fitting parameter	742.182 cm^{-1}	669.205 cm^{-1}
n_{tr}	Carrier density at transparency	1.38×10^{18} cm^{-3}	1.15×10^{18} cm^{-3}
B_0	Spontaneous emission fitting parameter	-8.375×10^{25} cm^{-3}/s	-1.351×10^{26} cm^{-3}/s
B_1	Spontaneous emission fitting parameter	1.344×10^8 s^{-1}	1.975×10^8 s^{-1}
B_2	Spontaneous emission fitting parameter	2.790×10^{-11} cm^3/s	2.146×10^{-11} cm^3/s

density. However, a more suitable model relies on a two-parameter (g_0, n_{tr}) logarithmic dependence (Coldren et al., 2012), which provides gain saturation with carrier density increase, $g(n) = g_0 \ln(n/n_{tr})$, where g_0 is the material gain fitting parameter and n_{tr} is the carrier density at transparency. It should be noted that, although the two-parameter logarithmic gain does provide reasonably good results, for more linear-like dependencies of gain on carrier density, such as for bulk materials, another model may be employed, which includes a third parameter, n_s. This model, $g(n) = g_0 \ln[(n + n_s)/(n_{tr} + n_s)]$, prevents infinite gain value for zero carrier densities, and in the limiting case of $n_s \to 0$ reduces to the purely logarithmic model, whereas for $n_s \to \infty$ reduces to purely linear model (Coldren et al., 2012). The model used in this case study is a two-parameter logarithmic gain since it fits well on numeric results.

The radiative spontaneous recombination rate is usually modeled via the quadratic function (Chu and Ghafouri-Shiraz, 1994; Melo and Petermann, 2008). Although this function fits actual dependence quite well, a more detailed model would include a second-degree polynomial function, $R_{sp}(n) = B_0 + B_1 n + B_2 n^2$, where B_0, B_1, and B_2 are the spontaneous emission rate fitting parameters. Table 21.5 shows a list of fitting parameters at wavelength $\lambda = 1.55$ μm ($\hbar\omega_0 = 0.8$ eV), for active regions based on the unstrained bulk $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$, and the MQW consisting of six coupled 0.13% tensile-strained $\text{In}_{0.516}\text{Ga}_{0.484}\text{As}$ quantum wells, which are strain-compensated by the $\text{In}_{0.9}\text{Ga}_{0.1}\text{As}_{0.3}\text{P}_{0.7}$ 0.26% compressively strained barriers.

In a steady-state analysis, the input signal carries no information encoded in its amplitude or phase, so the phase equation may be omitted from the model, in attempting to reduce the model's complexity. This simplification is additionally supported by the fact that the carrier diffusion will wash out the spatial grating in the carrier density produced by the interference between two counterpropagating optical fields.

For all antireflective facets, residual reflectivity can be neglected and equated with zero. Since the noise power is usually several orders of magnitude lower than the signal power, zero reflectivity implies that in the case of TW-SOA, the signal can be assumed to travel only once through the active region in the forward direction. On the other hand, in the case of R-SOA, the signal will pass through the active region twice, forward and backward, as a consequence of reflection at the rear facet. In other words, no multiple reflections are present, and the cavity does not exhibit resonant behavior. The boundary conditions for the signal at the front and rear facets are $S_+(0) = S_0$ and $S_-(L) = R_2 S_+(L)$, respectively, where S_0 is the input signal photon density. Depending on the SOA type, reflectivity of the rear facet can have either some finite value, for R-SOA, or it can be zero, for TW-SOA, which essentially simplifies the boundary condition to $S_-(L) = 0$.

The steady-state carrier density rate equation can be derived from Equations 21.71, 21.75, and 21.76, under the previously listed assumptions:

$$\frac{I}{qV} - \left(An + R_{\text{sp}} + Cn^3\right) - \frac{v_g g\left(S_+ + S_-\right)}{1 + \varepsilon\left(S_+ + S_-\right)} = 0. \tag{21.95}$$

This equation can be used for R-SOA modeling as is, whereas in the case of TW-SOA, it can be simplified by neglecting the backward-propagating noise contribution and setting the photon density of the backward-propagating signal to zero ($S_- = 0$). The system of coupled equations given by Equations 21.94 and 21.95 cannot be solved analytically unless the approximation regarding the carrier density's spatial distribution is introduced. Different approximations require different analytical treatments and implementations, which are further discussed.

21.5.4.1 The Fundamentals of the Semianalytical Models

The starting point of semianalytical treatment is replacement of the carrier density spatial distribution with a fixed value, equal to the arithmetic mean of the carrier density values at the edges of the SOA structure, or the edges of its sections, namely, $n(z) = \bar{n} = (n_0 + n_L)/2$. These carrier densities are deduced by solving the carrier rate Equation 21.95, where photon density distributions are expressed as functions of the fixed carrier density $\bar{n}$. In other words, once the analytical dependence of $S_\pm(\bar{n}, z) = S_\pm(n_0, n_L, z)$ is determined, it can be substituted in Equation 21.95 at $z = 0$ and $z = L$, which results in a system of two coupled transcendental equations written with respect to n_0 and n_L. The difference between these two carrier density equations comes from the spontaneous emission term $R_{sp}(n)$ and material gain $g(n)$, which for $z = 0$ depend on n_0, and for $z = L$ depend on n_L. Solving these equations provides the numerical value of $\bar{n}$, which is then returned to $S_\pm(\bar{n}, z)$, and the output photon density can be evaluated.

The same method can be applied to an SOA section of length $\Delta z = L/p$, given in Figure 21.15, where p is the number of analyzed sections, and the method can be generalized to the piecewise constant carrier density spatial distribution. The advantage of this approach over the TMM is the choice of the kth section carrier density value, $\bar{n}_k$, as the mean value between the carrier densities at the section interfaces, $\bar{n}_k = (n_{k-1} + n_k)/2$. In this manner, a higher degree of accuracy is provided for a smaller number of sections, compared to the choice of carrier density at the edge of a section.

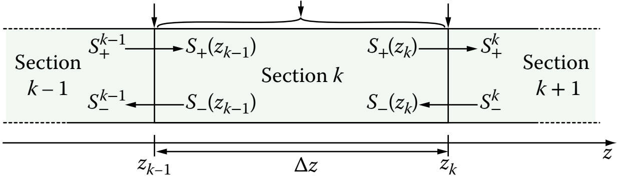


FIGURE 21.15 Schematic representation of the kth section in the segmented semiconductor optical amplifier structure with the relevant carrier and photon densities.

Assuming that the carrier density has a known, fixed value $\bar{n}_k$ at the kth section of an R-SOA, Equation 21.94 needs to be solved analytically for both propagation directions, with initial conditions imposed by the photon densities entering the kth section, which can be determined by solving Equation 21.94 at section $k-1$ for forward, and $k+1$ for backward propagation. However, the analytical solution of Equation 21.94 cannot be obtained for R-SOA without additional approximations, due to the nonlinear coupling of S_+ and S_-. In the case of TW-SOA, only the forward-propagating photon density S_+ is included in Equation 21.94, providing an analytical solution and the corresponding initial conditions for all sections along TW-SOA.

21.5.4.2 Approximate Solutions of Counterpropagating Photon Density Distributions and Implementation of the Semianalytical Algorithm

In order to derive the photon density spatial distributions, S_+ and S_-, corresponding to R-SOA, the system of Equation 21.94 needs to be decoupled, which can be done in two ways.

First, the nonlinear gain suppression can be neglected by setting $\varepsilon = 0$, which results in a linear decoupled system with respect to S_+ and S_-:

$$\pm\frac{\mathrm{d}S_\pm}{\mathrm{d}z} = \left(\Gamma g - \alpha_\mathrm{i}\right) S_\pm + \frac{1}{2v_\mathrm{g}}\Gamma\beta_\mathrm{sp}R_\mathrm{sp}. \tag{21.96}$$

This assumption greatly simplifies the system, and, for the bias currents and input optical powers that are not too high, it does not degrade the model accuracy significantly. The solution of Equation 21.96 for each section can be found in closed analytical form, corresponding to the matrix form (Equation 21.70) used in description of noise by TMM:

$$S_+^k = S_+^{k-1}G_k + \sigma_\mathrm{sp}^k\left(G_k - 1\right)/\ln G_k, \tag{21.97}$$

$$S_-^{k-1} = S_-^k G_k + \sigma_\mathrm{sp}^k\left(G_k - 1\right)/\ln G_k, \tag{21.98}$$

where $G_k = G(\bar{n}_k) = \exp\{[\Gamma g(\bar{n}_k) - \alpha_\mathrm{i}(\bar{n}_k)]\Delta z\}$ is the single-pass gain of the kth section and $\sigma_\mathrm{sp}^k = \sigma_\mathrm{sp}(\bar{n}_k) = 1/(2v_\mathrm{g})\Gamma\beta_\mathrm{sp}R_\mathrm{sp}(\bar{n}_k)\Delta z$ is the single-pass generated spontaneous emission within the kth section. Due to the cascading nature of sections, the output photon density of the $(k-1)$th section, $S_+^{k-1} = S_+(z_{k-1})$, is the input photon density of the kth one for forward propagation, whereas for backward propagation, the output photon density of the $(k+1)$th section, $S_-^k = S_-(z_k)$, is the input photon density of the kth one, as shown in Figure 21.15. This chaining, along with the form of Equations 21.97 and 21.98, suggests that the photon densities at section interfaces are calculated by recursion. The base cases in a recursive method for forward and backward propagation are defined by the boundary conditions at the front $(k=1)$ and rear $(k=p)$ facets, respectively, $S_+^0 = S_0$ and $S_-^p = R_2S_+^p$. The forward-propagating signal exiting the kth section will depend on carrier densities of all preceding sections, $S_+^k = S_+(z_k, \bar{n}_1, \ldots, \bar{n}_k) = S_+(z_k, n_0, \ldots, n_k)$, whereas the backward-propagating signal will depend on carrier densities in all sections, $S_-^k = S_-(z_k, \bar{n}_1, \ldots, \bar{n}_p) = S_-(z_k, n_0, \ldots, n_p)$, due to the imposed initial condition at the rear facet, which assumes reflection of $S_+(z_p, n_0, \ldots, n_p)$. These analytical expressions can now be returned to the carrier rate Equation 21.95 at each section interface located at z_k

$$f_k = 0 = \frac{I}{qV} - \left[An_k + R_\mathrm{sp}\left(n_k\right) + Cn_k^3\right] - v_\mathrm{g}g\left(n_k\right)\left[S_+\left(z_k, n_0, \ldots, n_k\right) + S_-\left(z_k, n_0, \ldots, n_p\right)\right], \tag{21.99}$$

which forms a system of $p+1$ transcendental equations, all of which are dependent on $p+1$ carrier densities, from n_0 to n_p. This system cannot be solved analytically, so a numerical procedure is required. An example of such a procedure, given by Totović et al. (2014), relies on numerically assisted elimination of variables, where, in an iterative approach, the range of carrier densities in which the solution n_k is expected to be found is narrowed simultaneously for all k from 0 to p. The initial range for carrier densities at all interfaces is bounded by the carrier density at transparency, n_{tr}, and the maximum carrier density, $n_{\max}$, limited by the available current (Equation 21.88). In each iteration step, the corresponding range of each n_k is reduced to one fifth of its previous size by the appropriate algorithm, still including the solution n_k. After a predefined number of iteration steps, the values for n_k can be determined by evaluating the RHS of the system (Equation 21.99) for the carrier densities in the narrowed ranges, giving a function $f_k(n)$, and subsequent interpolation of the function $f_k(n)$ in search for n_k which returns 0, i.e., the LHS value of Equation 21.99. Usually, four iteration steps suffice for the error coming from neglected gain suppression to prevail over the error coming from the iteration process. Having the values for carrier densities at all interfaces, it is possible to evaluate the photon densities at each z_k using Equations 21.97 and 21.98, and subsequently determine the device gain. Since this model treats the propagation of the signal and ASE noise combined through Equation 21.96, or more generally Equation 21.94, the device gain is defined as the ratio of the output signal power, along with the total ASE noise, and the input signal power. This definition implies that in the limiting case of zero input signal, the device gain will asymptotically approach infinity due to division by zero.

Another approach that can be used for the decoupling of Equation 21.94 with respect to $S_\pm(z,n)$ is to assume that the opposite-propagating photon density is spatially independent, i.e., $S_\mp(z,n) = S_\mp^{\mathrm{cst}}(n)$. Unlike the carrier density, which varies modestly across the active region, photon density exhibits exponential increase or decrease from one section edge to the other, so a choice of geometric mean is a more suitable one, which gives $S_\mp^{\mathrm{cst},k}(n) = [S_\mp^{k-1}(n)S_\mp^{k}(n)]^{1/2}$. The term $1 + \varepsilon(S_+ + S_-)$ in Equation 21.94, which is responsible for equation coupling, can now be replaced with $1 + \varepsilon(S_\pm + S_\mp^{\mathrm{cst}})$, or, more generally, with a sum of the spatially independent, and spatially dependent functions, $\theta_\pm + \varepsilon S_\pm$, where $\theta_\pm = 1 + \varepsilon S_\mp^{\mathrm{cst}}$. In the case of TW-SOA, no backward-propagating signal exists and $\theta_\pm = 1$. Equation 21.94 now reads

$$\pm\frac{\mathrm{d}S_\pm}{\mathrm{d}z} = \left(\frac{\Gamma g}{\theta_\pm + \varepsilon S_\pm} - \alpha_{\mathrm{i}}\right) S_\pm + \frac{1}{2v_{\mathrm{g}}}\Gamma\beta_{\mathrm{sp}}R_{\mathrm{sp}}, \tag{21.100}$$

and its solution for the kth section can be found in the implicit form from the transcendental equation

$$2\alpha_{\mathrm{i}}^{k}\Delta z = \left(\frac{1}{T_\pm^k} - 1\right)\ln\left|\frac{\theta_\pm^k + \varepsilon S_\pm^{\mathrm{out},k} - \left(1 - T_\pm^k\right)/\mu_\pm^k}{\theta_\pm^k + \varepsilon S_\pm^{\mathrm{in},k} - \left(1 - T_\pm^k\right)/\mu_\pm^k}\right| - \left(\frac{1}{T_\pm^k} + 1\right)\ln\left|\frac{\theta_\pm^k + \varepsilon S_\pm^{\mathrm{out},k} - \left(1 + T_\pm^k\right)/\mu_\pm^k}{\theta_\pm^k + \varepsilon S_\pm^{\mathrm{in},k} - \left(1 + T_\pm^k\right)/\mu_\pm^k}\right|, \tag{21.101}$$

where the auxiliary parameters $\mu_\pm^k$ and $T_\pm^k$ are defined as follows:

$$\mu_\pm^k = \frac{2\alpha_{\mathrm{i}}^k}{\Gamma g_k + \alpha_{\mathrm{i}}^k\theta_\pm^k + \varepsilon\Gamma\beta_{\mathrm{sp}}R_{\mathrm{sp}}^k\Big/\left(2v_{\mathrm{g}}\right)}, \tag{21.102}$$

$$\left(T_\pm^k\right)^2 = 1 - \left(\mu_\pm^k\right)^2\Gamma g_k\theta_\pm^k\Big/\alpha_{\mathrm{i}}^k. \tag{21.103}$$

In Equation 21.101, $S_\pm^{\mathrm{in},k}$ and $S_\pm^{\mathrm{out},k}$ denote the input and output photon densities into and from the kth section, respectively. For forward propagation, $S_+^{\mathrm{in},k} = S_+^{k-1}$ and $S_+^{\mathrm{out},k} = S_+^{k}$, whereas in the case of backward propagation, $S_-^{\mathrm{in},k} = S_-^{k}$ and $S_-^{\mathrm{out},k} = S_-^{k-1}$, according to Figure 21.15. Equation 21.101 is solved with respect to $S_\pm^{\mathrm{out},k}$ recursively, using the previously determined $S_\pm^{\mathrm{in},k}$, with the base cases in a recursive method defined by the boundary conditions in sections $k = 1$ and $k = p$, for forward and backward propagation, respectively. Just as in the previous model, where the nonlinear gain suppression was neglected, the solutions of Equation 21.101 will depend either on carrier densities in all preceding sections for forward propagation, $S_+^{\mathrm{out},k} = S_+(z_k, \bar{n}_1, \dots, \bar{n}_k) = S_+(z_k, n_0, \dots, n_k)$, or on all carrier densities in the case of backward propagation, $S_-^{\mathrm{out},k} = S_-(z_{k-1}, \bar{n}_1, \dots, \bar{n}_p) = S_-(z_{k-1}, n_0, \dots, n_p)$. Substituting these dependencies into the carrier rate Equation 21.95 at each section interface, z_k, gives the system of $p + 1$ transcendental equations:

$$f_k = 0 = \frac{I}{qV} - \left[An_k + R_{\mathrm{sp}}\left(n_k\right) + Cn_k^3\right] - \frac{v_g g\left(n_k\right)\left[S_+\left(z_k, n_0, \ldots, n_k\right) + S_-\left(z_k, n_0, \ldots, n_p\right)\right]}{1 + \varepsilon\left[S_+\left(z_k, n_0, \ldots, n_k\right) + S_-\left(z_k, n_0, \ldots, n_p\right)\right]}. \quad (21.104)$$

Unlike the previous model, for $\varepsilon = 0$, where the photon density dependencies on n_k could be expressed in closed analytical form, the solutions are now given in implicit form (Equation 21.101), which essentially requires a numerical method. Moreover, in order to employ a numerical method, the values for carrier densities in all sections need to be known. This raises the need for a self-consistent numerical approach in solving the system given by Equations 21.101 and 21.104. The numerical method should comprise two distinct steps. First, the IGVs need to be generated for carrier and photon densities at all section interfaces, which is done using the model for $\varepsilon = 0$. Next, an iterative procedure is used, where carrier and photon densities are alternately calculated, and carrier densities are subsequently updated based on the values from the current and previous iteration steps until the predefined tolerance is satisfied. The maximum relative error between two consecutive iteration steps, δ, is calculated as the maximum relative error for all monitored variables, at all interfaces, i.e., $\delta = \max\{\delta(n_k), \delta(S_+^k), \delta(S_-^k)\}$, and the tolerance is chosen to be $\delta_{\max} = 10^{-6}$. The updating of n_k is done using a set of weight coefficients with a sum of unity for the current and previous iteration steps, $\{w_i, w_{i-1}\} = \{0.1, 0.9\}$. The weight coefficients are chosen such that they give balance between the stability and speed of convergence, and usually no more than 100 iteration steps are required for the procedure to reach a self-consistent solution. It should be noted that due to the nature of the algorithm, the system of equations given by Equation 21.104 is not coupled since $S_+(z_k, n_0, \ldots, n_k)$ and $S_-(z_k, n_0, \ldots, n_p)$ are calculated for the carrier values from the previous iteration step and can be treated as parameters. In other words, the current iteration variable n_k is included in Equation 21.104 through the spontaneous emission terms and material gain.

It has been shown by Totović et al. (2014) that depending on the number of chosen sections and the level of approximation, the semianalytical model can be up to two orders of magnitude faster in comparison to the numerical model based on SCNM. Meanwhile, the maximum absolute mismatch in the device gain calculated by the numerical and semianalytical models does not exceed 1.6 dB even for only one section, and neglected nonlinear gain suppression, whereas this mismatch drops to a maximum of 0.11 dB for three sections with included gain suppression for a bulk R-SOA.

21.5.5 Case Study: Dynamic Propagation Model

When the input signal carries information encoded either in its amplitude, or phase, or both, a dynamic approach in SOA modeling is required. The same model needs to be applied for modulation or remodulation purposes, when the bias current is time dependent. The most general dynamic model, suitable for both large- and small-signal modulation, includes spatiotemporal and spectral dependency of both forward- and backward-propagating signals and noise, given by Equations 21.50 through 21.52, and, depending on the active region type, one or several equations describing spatiotemporal carrier density distribution. For bulk SOA, carrier density evolution can be described by Equation 21.49, for the MQW active region by Equations 21.54 and 21.55, and for the QD-SOA by Equations 21.57 through 21.59. Due to the ampleness and complexity of three-dimensional (3D) mesh (time-space-spectrum) implementation in the numerical algorithms, whenever the input signal is spectrally narrow, it is common practice to simplify the model and perform the analysis for photon densities and not their spectral distributions. Similar to Section 21.5.4, material gain and spontaneous emission rate can be approximated with the two-parameter logarithmic and second-order polynomial function, respectively, $g(n) = g_0 \ln(n/n_{\mathrm{tr}})$ and $R_{\mathrm{sp}}(n) = B_0 + B_1 n + B_2 n^2$, where the fitting parameters are given in Table 21.5 for the bulk and MQW active region, at the signal central wavelength of 1.55 μm. All other parameters, which are dependent on photon energy and carrier density, can be approximated by the corresponding fixed values given in Table 21.4, interpolated for the photon energy of $\hbar\omega_0 = 0.8$ eV and an average carrier density n_{av}, which depends on the active region type. Under these assumptions, the signal can be described by the equation similar to Equation 21.23, with the addition of nonlinear gain suppression:

$$\pm\frac{\partial S_\pm}{\partial z} + \frac{1}{v_{\mathrm{g}}}\frac{\partial S_\pm}{\partial t} = \left(\frac{\Gamma g}{1+\varepsilon S_\Sigma} - \alpha_{\mathrm{i}}\right) S_\pm. \tag{21.105}$$

The phase evolution for both propagating directions is given by Equation 21.51:

$$\pm\frac{\partial \varphi_\pm}{\partial z} + \frac{1}{v_{\mathrm{g}}}\frac{\partial \varphi_\pm}{\partial t} = \frac{k_0 \Gamma \Delta n_r}{1+\varepsilon S_\Sigma}. \tag{21.106}$$

ASE noise is treated with a separate equation, resembling Equation 21.105, with the addition of the spontaneous emission rate contribution:

$$\pm\frac{\partial A_\pm}{\partial z} + \frac{1}{v_{\mathrm{g}}}\frac{\partial A_\pm}{\partial t} = \left(\frac{\Gamma g}{1+\varepsilon S_\Sigma} - \alpha_{\mathrm{i}}\right) A_\pm + \frac{1}{2v_{\mathrm{g}}}\Gamma\beta_{\mathrm{sp}}R_{\mathrm{sp}}. \tag{21.107}$$

It should be noted that by using Equation 21.107, spectral dependence of ASE is neglected and all noise photons are treated at the single frequency. Boundary conditions are defined by the input signal, described by its photon density $S_0(t)$ and phase $\varphi_0(t)$, and front and rear facet reflectivities, R_1 and R_2, respectively. For antireflective facets, zero reflectivity can be assumed, whereas for the highly reflective ones, the reflectivity coefficient can be calculated as $R_2 = (n_{\mathrm{r}} - 1)^2/(n_{\mathrm{r}} + 1)^2$, if no reflective layer is placed. At the front facet, which is antireflective, the boundary conditions read $S_+(0) = S_0, \varphi_+(0) = \varphi_0, A_+(0) = 0$, and for the rear facet, which can be either highly- or antireflective, $S_-(L) = R_2 S_+(L), \varphi_-(L) = \varphi_+(L) + \pi$, and $A_-(L) = R_2 A_+(L)$. In the case of TW-SOA, rear facet reflectivity R_2 will be zero, and no backward-propagating signal will exist. However, since the noise is generated within the SOA, with equal probability of photon traveling in any direction, backward-propagating noise will exist, both for TW- and R-SOA.

Carrier density distribution is described by Equation 21.49, where the current I can be time dependent:

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{I}{qV} - \left(An + R_{\mathrm{sp}} + Cn^3\right) - \frac{v_{\mathrm{g}} g S_\Sigma}{1+\varepsilon S_\Sigma}. \tag{21.108}$$

In Equations 21.105 through 21.108, S_Σ stands for the total photon density in the SOA's active region, including both signal and noise, $S_\Sigma = S_+ + S_- + A_+ + A_-$.

The system given by Equations 21.105 through 21.108 cannot be analytically solved, and several numerical approaches are described in Section 21.5.2. Although TMM is widely used due to its simplicity, coarse segmentation along the longitudinal axis might lead to an error in calculating the photon density distribution. Moreover, this error can be accumulated during matrix multiplication and can be even more pronounced when two counterpropagating signals exist.

21.5.5.1 Upwind Scheme Numerical Implementation

In order to mitigate potential errors, a first-order upwind scheme numerical method based on FDM can be developed, where the space and time axis are treated as quasi-continuous through fine segmentation. After choosing the number of points along the longitudinal axis, P, which define the section length, $\Delta z = L/(P-1)$, segmentation of the temporal axis is done, such that $\Delta t \leq \Delta z/v_{\mathrm{g}}$, according to the Courant–Friedrichs–Lewy stability condition. Depending on the length of the analyzed time interval, T, the number of points along the temporal axis is determined as $Q = \lceil T/\Delta t \rceil + 1$. All derivatives in Equations 21.105 through 21.108 are replaced with finite differences, such that for temporal derivatives forward differences are used:

$$\frac{\mathrm{d}f}{\mathrm{d}t} \approx \frac{f\left(t_{j+1}\right) - f\left(t_j\right)}{\Delta t}, \tag{21.109}$$

where f is any time-dependent variable, i.e., $S_\pm, \varphi_\pm, A_\pm, n$, and for spatial derivatives, the choice is made based on the propagation direction. For forward propagation

$$\frac{\mathrm{d}f_+}{\mathrm{d}z} \approx \frac{f_+\left(z_i\right) - f_+\left(z_{i-1}\right)}{\Delta z}, \tag{21.110}$$

where f_+ denotes any of the following variables S_+, φ_+, A_+, whereas for backward propagation

$$\frac{\mathrm{d}f_-}{\mathrm{d}z} \approx \frac{f_-\left(z_{i+1}\right) - f_-\left(z_i\right)}{\Delta z}, \tag{21.111}$$

where f_- denotes any of the following variables: S_-, φ_-, A_-. Using Equations 21.109 through 21.111, a set of coupled partial differential equations describing signal and noise propagation along the SOA, Equations 21.105 through 21.107, can be transformed into the set of coupled linear algebraic equations, for forward

$$f_+\left(z_i, t_{j+1}\right) = f_+\left(z_i, t_j\right) + \Delta t v_\mathrm{g}\left[F_+\left(z_i, t_j\right) - \frac{f_+\left(z_i, t_j\right) - f_+\left(z_{i-1}, t_j\right)}{\Delta z}\right], \tag{21.112}$$

and backward propagation

$$f_-\left(z_i, t_{j+1}\right) = f_-\left(z_i, t_j\right) + \Delta t v_\mathrm{g}\left[F_-\left(z_i, t_j\right) + \frac{f_-\left(z_{i+1}, t_j\right) - f_-\left(z_i, t_j\right)}{\Delta z}\right], \tag{21.113}$$

where $f_\pm$ stands for any variable $S_\pm, \varphi_\pm$, or $A_\pm$, whereas $F_\pm$ stands for the RHS of the corresponding equation evaluated at point (z_i, t_j) in the 2D spatiotemporal mesh. For carrier density equation, the transformation yields

$$n\left(z_i, t_{j+1}\right) = n\left(z_i, t_j\right) + \Delta t F\left(z_i, t_j\right), \tag{21.114}$$

where F stands for the RHS of Equation 21.108, evaluated at point (z_i, t_j) in the 2D spatiotemporal mesh. Forward differences for temporal derivatives enable calculation of all variables in the next point in time, t_{j+1}, based on the values from the current one t_j, as shown in Equations 21.112 through 21.114, and illustrated in Figure 21.16.

The numerical implementation of the discretized system (Equations 21.112 through 21.114) essentially requires two distinctive iteration processes: one for the temporal and another for spatial evolution. For every point j along the temporal axis, spatial distribution of carrier density for the next temporal point $(j + 1)$ is determined using Equation 21.114, which is followed by simulation of forward and subsequently

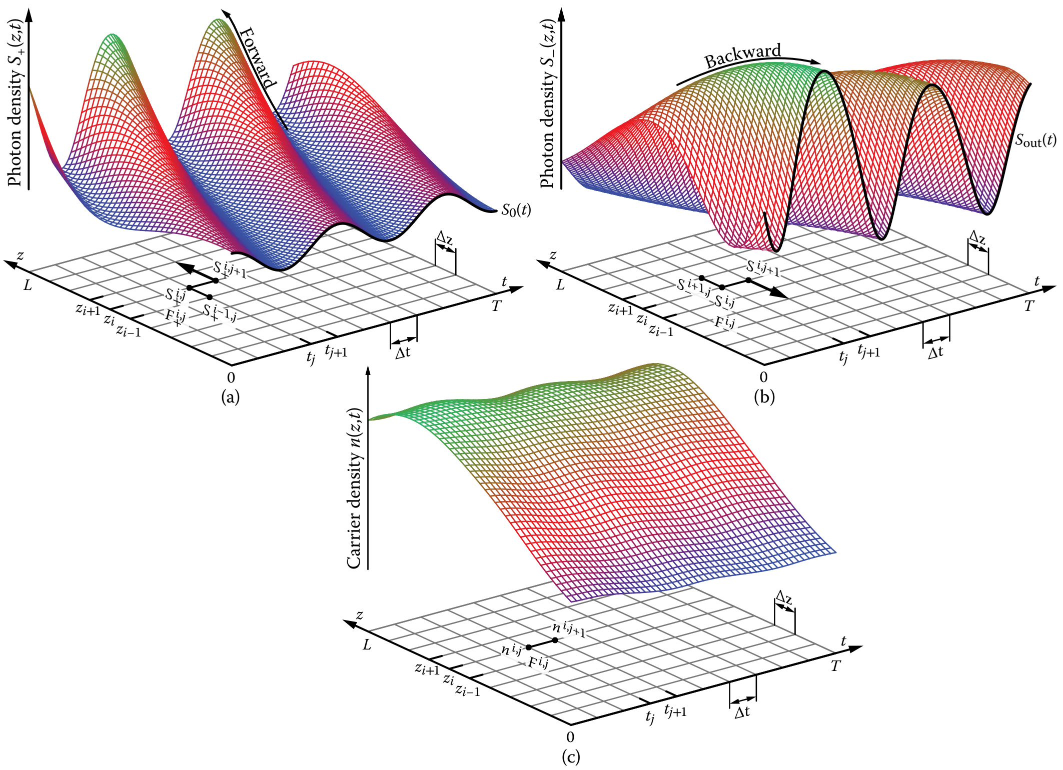


FIGURE 21.16 Illustration of two-dimensional (2D) spatiotemporal mesh with denoted relevant points for calculation of (a) forward- and (b) backward-propagating signal photon densities by upwind finite differences scheme, and (c) carrier densities by the forward finite differences method. The same upwind scheme can be applied to the phase, or noise photon density calculation. The example includes spatiotemporal distribution for (a) forward signal, (b) backward signal, and (c) carrier density.

backward propagation of signal and noise along the SOA's active region using Equations 21.112 and 21.113. This process is repeated until the end of the temporal axis is reached, incrementing j by one, up to Q. Since the results in the $(j + 1)$th point in time are dependent on the values from the previous jth point, initial conditions for spatial distributions of all variables are required (for $j = 0$), in order to start the iteration process over time by calculating the values in the point $j = 1$. Initial values will depend on the assumed SOA state before the initialization of the simulation. If no input signal is present before $t = 0$, it can be assumed that both $S_\pm(z, 0)$ and $\varphi_\pm(z, 0)$ are zero. The noise will exist regardless of the signal's presence, so stationary distributions can be assumed for $A_\pm(z, 0)$. Based on these values, using Equation 21.108, stationary distribution for carrier density can also be determined and associated with $n(z, 0)$. Within the nested loop, for a fixed j, system (Equations 21.112 through 21.114) is solved for every point z_i, from 1 to P for forward-propagation, and P − 1 to 0 for backward propagation. The algorithm is summarized in the flowchart given in Figure 21.17.

21.5.6 Case Study: TMW Small-Signal Model

A very valuable figure-of-merit in using SOA for modulation purposes is its electro-optical (E/O) small-signal −3dB modulation bandwidth Ω_{3dB}. It is defined as the frequency of the modulation signal, i.e., electric current, for which the output SOA small signal, i.e., photon density or optical power, drops

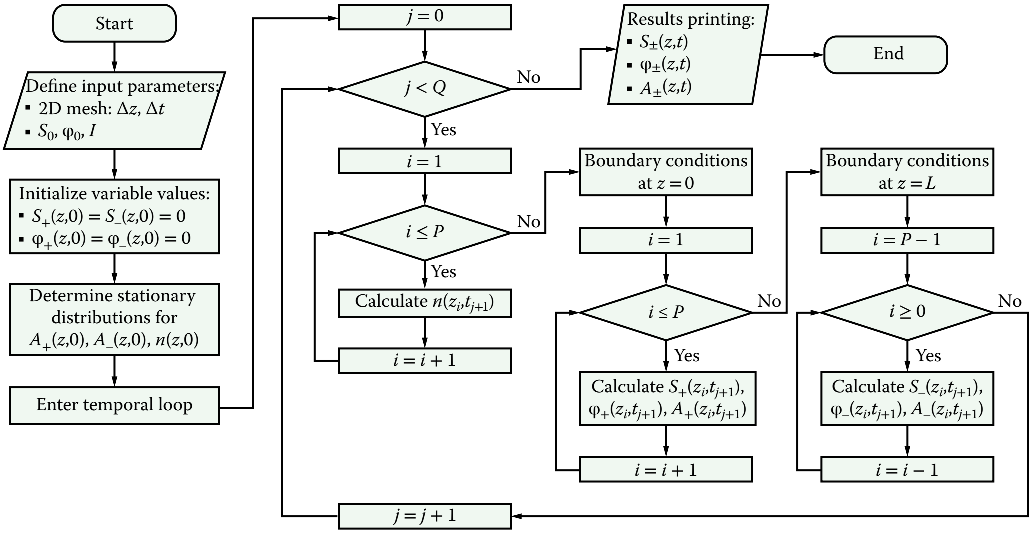

Start
Define input parameters:
▪ 2D mesh: Δz, Δt
▪ S_0, φ_0, I
Initialize variable values:
▪ $S_+(z,0) = S_-(z,0) = 0$
▪ $\varphi_+(z,0) = \varphi_-(z,0) = 0$
Determine stationary distributions for $A_+(z,0)$, $A_-(z,0)$, $n(z,0)$
Enter temporal loop
$j = 0$
$j < Q$
No
Yes
$i = 1$
$i \leq P$
No
Yes
Calculate $n(z_i,t_{j+1})$
$i = i + 1$
$j = j + 1$
Results printing:
▪ $S_\pm(z,t)$
▪ $\varphi_\pm(z,t)$
▪ $A_\pm(z,t)$
End
Boundary conditions at $z = 0$
$i = 1$
$i \leq P$
No
Yes
Calculate $S_+(z_i,t_{j+1})$, $\varphi_+(z_i,t_{j+1})$, $A_+(z_i,t_{j+1})$
$i = i + 1$
Boundary conditions at $z = L$
$i = P - 1$
$i \geq 0$
No
Yes
Calculate $S_-(z_i,t_{j+1})$, $\varphi_-(z_i,t_{j+1})$, $A_-(z_i,t_{j+1})$
$i = i - 1$

FIGURE 21.17 Flowchart of the upwind finite differences algorithm for the calculation of carrier and photon density spatiotemporal distributions, for both signal and noise.

to the half of its zero-frequency value. Knowing the influence of different structural and material parameters, as well as operating conditions, on $\Omega_{3\text{dB}}$, provides the means for optimization of the modulation response and increasing the bandwidth. Under the assumption that the input signal is spectrally narrow, the analysis is based on photon densities for forward and backward propagation, which include both signal and noise. The TW equations have the form similar to Equation 21.105, with the addition of the term describing spontaneous emission, as in Equation 21.107:

$$\pm\frac{\partial S_\pm}{\partial z} + \frac{1}{v_\text{g}}\frac{\partial S_\pm}{\partial t} = \left(\frac{\Gamma g}{1+\varepsilon S_\Sigma} - \alpha_\text{i}\right) S_\pm + \frac{1}{2v_\text{g}}\Gamma\beta_\text{sp}R_\text{sp}, \tag{21.115}$$

where $S_\Sigma = S_+ + S_-$. Since the intensity modulation is of interest, the analysis of the signal phase is not necessary. The carrier rate equation has the same form as Equation 21.108:

$$\frac{\text{d}n}{\text{d}t} = \frac{J}{qH} - \left(An + R_\text{sp} + Cn^3\right) - \frac{v_\text{g} g S_\Sigma}{1+\varepsilon S_\Sigma}, \tag{21.116}$$

where current density $J = I/(WL)$ is used instead of current I. This choice stems from the fact that modulation current I depends on device length, and the current density proves to be a more suitable parameter for benchmarking the performance of SOAs with different active region lengths. Generally, bias current density will consist of the stationary component and a small signal, which will, through Equation 21.116, lead to the carrier density modulation, and consequently to modulation of all carrier-dependent parameters, namely, g, R_sp, α_i, and $S_\pm$. From this point onward, all stationary values will be denoted by an overline, whereas all small-signals will be prefixed with Δ.

For the stationary value of bias current density, $\bar{J}$, all temporal derivatives will diminish from Equations 21.115 and 21.116, and the system will reduce to its steady-state form, given by Equations 21.94 and 21.95. In order to determine the steady-state values of $\bar{n}$ and $\bar{S}_\pm$, any of the previously discussed methods for steady-state analysis, given in Sections 21.5.1, 21.5.3, and 21.5.4, can be used. It is assumed that the CW input signal enters the device at the front facet, giving $S_+(z = 0, t) = \bar{S}_+(z = 0) = S_0$. The boundary condition for the backward-propagating signal at the rear facet is given by $S_-(z = L, t) = R_2 S_+(z = L, t)$, where R_2 can have either a finite value, for R-SOA, or it can be zero, for TW-SOA.

The typical approach in small-signal analysis is to assume that the small-signal values are approximately lower by an order of magnitude or more, in comparison to the stationary values, and can be treated as perturbation. This assumption enables linearization of the system given by Equations 21.115 and 21.116 with respect to the small-signal quantities. As discussed in Section 21.4.3.5, bias current cannot instantaneously reach a uniform spatial distribution across the electrode and should be treated as a harmonic TMW with finite velocity, v_e. Moreover, due to reflection at the end of microstrip electrode, the microwave will propagate in both directions with respect to the longitudinal axis (Totović et al., 2015). This model becomes very important for the frequencies exceeding $v_e/(2\pi L)$, when an entire electric pulse is accommodated in the SOA's active region. A general model of the small-signal bias current density with sinusoidal waveform, defined by the spatially variable voltage across the electrode, can be written in the following form based on Equation 21.63:

$$\begin{aligned} J(z,t) &= \bar{J} + \Delta J(z)\exp(i\Omega t) \\ &= \bar{J} + \Delta J_{\mathrm{F}}(z)\exp\left[i\left(\Omega t - \beta_e z\right)\right] + \Delta J_{\mathrm{B}}(z)\exp\left[i\left(\Omega t + \beta_e z\right)\right], \end{aligned} \tag{21.117}$$

where $\Delta J(z)$ is the total small-signal bias current density spatial distribution, comprising both forward (F)- and backward (B)-propagating microwaves, $\Omega = 2\pi f$ is the angular frequency, with f being the frequency of the TMW, and β_e is the microwave propagation constant.

21.5.6.1 Derivation of the Small-Signal Model

Modulation of the bias current density leads to the modulation of carrier density n, and all carrier-dependent parameters, g, R_{sp}, α_i, and $S_\pm$, commonly denoted ξ, which will also have sinusoidal form

$$\begin{aligned} \xi(z,t) &= \bar{\xi}(z) + \Delta\xi(z)\exp(i\Omega t) \\ &= \bar{\xi}(z) + \Delta\xi_{\mathrm{F}}(z)\exp\left[i\left(\Omega t - \beta_e z\right)\right] + \Delta\xi_{\mathrm{B}}(z)\exp\left[i\left(\Omega t + \beta_e z\right)\right], \end{aligned} \tag{21.118}$$

where $\Delta\xi(z)$ is the total small-signal spatial distribution of the corresponding quantities. For known dependencies of material gain, and spontaneous emission rate, on carrier density, $g(n) = g_0 \ln(n/n_{tr})$, and $R_{sp}(n) = B_0 + B_1 n + B_2 n^2$, respectively, it is possible to express small-signal values of $\Delta g^{\mathrm{F/B}}$ and $\Delta R_{sp}^{\mathrm{F/B}}$ via $\Delta n^{\mathrm{F/B}}$ using the first derivatives of the corresponding functions, evaluated for the steady-state carrier density. This gives

$$\Delta g^{\mathrm{F/B}} = \left(\mathrm{d}g/\mathrm{d}n\right)\Big|_{n=\bar{n}} \Delta n^{\mathrm{F/B}} = \left(g_0/\bar{n}\right)\Delta n^{\mathrm{F/B}}, \tag{21.119}$$

$$\Delta R_{sp}^{\mathrm{F/B}} = \left(\mathrm{d}R_{sp}/\mathrm{d}n\right)\Big|_{n=\bar{n}} \Delta n^{\mathrm{F/B}} = \left(B_1 + 2B_2\bar{n}\right)\Delta n^{\mathrm{F/B}}. \tag{21.120}$$

Substituting Equations 21.117 through 21.120 into Equation 21.116, and separating the stationary, and small signals, gives a system comprising a steady-state rate equation, given by Equation 21.95, and small-signal equations for forward and backward TMW propagation. In order to linearize the small-signal equations, two approximations are required. The first one is related to the term describing the stimulated emission in Equation 21.116, namely, $v_g g S_\Sigma/(1 + \varepsilon S_\Sigma)$, which is nonlinear due to the presence of the small-signal photon densities both in numerator and denominator. The function

$$\left(1 + \varepsilon S_\Sigma\right)^{-1} = \left[1 + \varepsilon\bar{S}_\Sigma + \varepsilon\left(\Delta S_+^{\mathrm{F}} + \Delta S_-^{\mathrm{F}}\right)e^{i(\Omega t - \beta_e z)} + \varepsilon\left(\Delta S_+^{\mathrm{B}} + \Delta S_-^{\mathrm{B}}\right)e^{i(\Omega t + \beta_e z)}\right]^{-1} \tag{21.121}$$

can be rephrased as

$$\left(1+\varepsilon S_\Sigma\right)^{-1}=\frac{1}{1+\varepsilon\bar{S}_\Sigma}\left[1+\frac{\varepsilon\left(\Delta S_+^{\mathrm{F}}+\Delta S_-^{\mathrm{F}}\right)}{1+\varepsilon\bar{S}_\Sigma}e^{i(\Omega t-\beta_{\mathrm{e}}z)}+\frac{\varepsilon\left(\Delta S_+^{\mathrm{B}}+\Delta S_-^{\mathrm{B}}\right)}{1+\varepsilon\bar{S}_\Sigma}e^{i(\Omega t+\beta_{\mathrm{e}}z)}\right]^{-1}. \tag{21.122}$$

The second term in the product of the RHS of Equation 21.122 can be treated as $(1+x)^{-1}$, where $x \ll 1$, and therefore approximated with $1-x$, which leads to

$$\left(1+\varepsilon S_\Sigma\right)^{-1}\approx\frac{1}{1+\varepsilon\bar{S}_\Sigma}-\frac{\varepsilon\left(\Delta S_+^{\mathrm{F}}+\Delta S_-^{\mathrm{F}}\right)}{\left(1+\varepsilon\bar{S}_\Sigma\right)^2}e^{i(\Omega t-\beta_{\mathrm{e}}z)}-\frac{\varepsilon\left(\Delta S_+^{\mathrm{B}}+\Delta S_-^{\mathrm{B}}\right)}{\left(1+\varepsilon\bar{S}_\Sigma\right)^2}e^{i(\Omega t+\beta_{\mathrm{e}}z)}. \tag{21.123}$$

The following step in linearization is applying the first-order approximation, i.e., neglecting of all small-signal terms of the order higher than one. This gives the following system of small-signal rate equations for forward and backward TMW propagation

$$i\Omega\Delta n^{\mathrm{F/B}}=\frac{\Delta J^{\mathrm{F/B}}}{qH}-\left(A+B_1+2B_2\bar{n}+3C\bar{n}^2\right)\Delta n^{\mathrm{F/B}}-\left[\frac{g_0}{\bar{n}}\frac{v_{\mathrm{g}}\bar{S}_\Sigma}{1+\varepsilon\bar{S}_\Sigma}\Delta n^{\mathrm{F/B}}+v_{\mathrm{g}}\frac{g_{\mathrm{eff}}}{\Gamma}\left(\Delta S_+^{\mathrm{F/B}}+\Delta S_-^{\mathrm{F/B}}\right)\right], \tag{21.124}$$

where $g_{\mathrm{eff}}=\Gamma\bar{g}/(1+\varepsilon\bar{S}_\Sigma)^2$. The previous system can be solved with respect to the small-signal carrier densities resulting from forward (F)- and backward (B)-propagating microwaves:

$$\Delta n^{\mathrm{F/B}}=\frac{\Delta J^{\mathrm{F/B}}/(qH)-v_{\mathrm{g}}g_{\mathrm{eff}}\left(\Delta S_+^{\mathrm{F/B}}+\Delta S_-^{\mathrm{F/B}}\right)/\Gamma}{A+B_1+2B_2\bar{n}+3C\bar{n}^2+\dfrac{g_0}{\bar{n}}\dfrac{v_{\mathrm{g}}\bar{S}_\Sigma}{1+\varepsilon\bar{S}_\Sigma}+i\Omega}. \tag{21.125}$$

Substituting Equations 21.118 through 21.120 and Equation 21.125 into Equation 21.115, followed by the linearization process, leads to the system of two steady-state TW equations given by Equation 21.94, and four small-signal TW equations written with respect to the small-signal photon densities propagating in any of the two directions (denoted by subscript sign $\Delta S_\pm$) and resulting from any of the microwave propagation directions (denoted by the superscript $\Delta S^{\mathrm{F/B}}$):

$$\pm\frac{\mathrm{d}\Delta S_\pm^{\mathrm{F}}}{\mathrm{d}z}=\frac{\Gamma\gamma_\pm}{qHv_{\mathrm{g}}}\Delta J_{\mathrm{F}}-\left(\varepsilon\bar{S}_\pm+\gamma_\pm\right)g_{\mathrm{eff}}\Delta S_\mp^{\mathrm{F}}+\left[g_{\mathrm{eff}}\left(1+\varepsilon\bar{S}_\mp-\gamma_\pm\right)-\bar{\alpha}_{\mathrm{i}}-i\left(\frac{\Omega}{v_{\mathrm{g}}}\mp\beta_{\mathrm{e}}\right)\right]\Delta S_\pm^{\mathrm{F}}, \tag{21.126}$$

$$\pm\frac{\mathrm{d}\Delta S_\pm^{\mathrm{B}}}{\mathrm{d}z}=\frac{\Gamma\gamma_\pm}{qHv_{\mathrm{g}}}\Delta J_{\mathrm{B}}-\left(\varepsilon\bar{S}_\pm+\gamma_\pm\right)g_{\mathrm{eff}}\Delta S_\mp^{\mathrm{B}}+\left[g_{\mathrm{eff}}\left(1+\varepsilon\bar{S}_\mp-\gamma_\pm\right)-\bar{\alpha}_{\mathrm{i}}-i\left(\frac{\Omega}{v_{\mathrm{g}}}\pm\beta_{\mathrm{e}}\right)\right]\Delta S_\pm^{\mathrm{B}}, \tag{21.127}$$

where

$$\gamma_\pm=\frac{v_{\mathrm{g}}\left(\dfrac{g_0/\bar{n}}{1+\varepsilon\bar{S}_\Sigma}-K_1\right)\bar{S}_\pm+\dfrac{1}{2}\beta_{\mathrm{sp}}\left(B_1+2B_2\bar{n}\right)}{A+B_1+2B_2\bar{n}+3C\bar{n}^2+\dfrac{g_0}{\bar{n}}\dfrac{v_{\mathrm{g}}\bar{S}_\Sigma}{1+\varepsilon\bar{S}_\Sigma}+i\Omega}. \tag{21.128}$$

The γ-parameter is dimensionless, complex, and spatially and frequency dependent. It is responsible for frequency dependence of small-signal photon density distributions and their sensitivity to the frequency change for all other parameters fixed. In the case of zero frequency, $\gamma_\pm$ becomes purely real, whereas for high frequencies, such as the ones close to the modulation bandwidth and beyond, its imaginary part becomes dominant over the real one.

Since Equations 21.126 and 21.127 are derived from Equation 21.115, the boundary conditions for small-signal equations are inherited from the ones corresponding to Equation 21.115. The input signal is assumed to be CW, so $\Delta S_+^{\mathrm{F/B}}(0) = 0$. At the rear facet, the boundary conditions are conditioned by the power reflectivity coefficient, $\Delta S_-^{\mathrm{F/B}}(L) = R_2 \Delta S_+^{\mathrm{F/B}}(L)$.

21.5.6.2 Numerical Implementation

In order to determine the −3dB bandwidth, $f_{3\mathrm{dB}} = \Omega_{3\mathrm{dB}}/(2\pi)$, it is necessary to calculate the output small-signal photon density for a range of modulation frequencies, $\Delta S_{\mathrm{out}}(\Omega)$, and determine the one for which the condition $|\Delta S_{\mathrm{out}}(\Omega_{3\mathrm{dB}})/\Delta S_{\mathrm{out}}(0)| = 1/2$ is satisfied. In the case of TW-SOA, the output signal will be recorded at the rear facet, $\Delta S_{\mathrm{out}} = \Delta S_+(L) = \Delta S_+^{\mathrm{F}}(L)\exp(-i\beta_e L) + \Delta S_+^{\mathrm{B}}(L)\exp(i\beta_e L)$, whereas for R-SOA, at the front facet, $\Delta S_{\mathrm{out}} = \Delta S_-(0) = \Delta S_-^{\mathrm{F}}(0) + \Delta S_-^{\mathrm{B}}(0)$. These signals can be found by solving the boundary value problem given by Equations 21.126 and 21.127. In the case of an R-SOA, this system comprises four coupled differential equations of the first order, with functional coefficients, and can be solved using any of the available numerical algorithms for boundary value problems, e.g., the FDM that implements the three-stage Lobatto IIIA formula (Hairer and Wanner, 1996). For a TW-SOA, backward-propagating signals can be neglected due to zero power reflectivity, and very low contribution of the spontaneous emission, so the system reduces to two noncoupled equations which can be solved with numerical integration.

Prior to solving system of Equations 21.126 and 21.127, it is necessary to calculate spatial distributions of the steady-state variables, $\bar{n}$ and $\bar{S}_\pm$, and auxiliary parameters, g_{eff} and $\gamma_\pm$. As previously discussed, steady-state values can be calculated by any of the models presented in earlier sections.

After the output small-signal variables are calculated, the −3dB bandwidth can be determined by interpolating the function $|\Delta S_{\mathrm{out}}(\Omega)/\Delta S_{\mathrm{out}}(0)|$ in search of the value $\Omega_{3\mathrm{dB}}$ which gives 1/2 as a solution. For rare, special cases, $\Omega_{3\mathrm{dB}}$ might be found in analytical form. In order to do this, usually a significant number of approximations is required, and the resulting analytical expression might be very complex (Antonelli et al., 2015), or the expression has a simpler form, but is valid under limited conditions.

21.5.6.3 Results and Discussion

Implementing the algorithm described in the previous pages provides the dependence of −3dB bandwidth, $f_{3\mathrm{dB}}$, on the operating conditions, given in Figure 21.18.

The analysis is done for the microwave velocity of $v_e = 4.6\times10^7$ m/s (Tauber et al., 1994), and under the assumption that the load and characteristic impedances of the transmission line, Z_{L} and Z_{C}, respectively, are matched, which results in the microwave reflection coefficient $\Gamma_{\mathrm{L}} = (Z_{\mathrm{L}} - Z_{\mathrm{C}})/(Z_{\mathrm{L}} + Z_{\mathrm{C}}) = 0$. The small signal bias current density, ΔJ, is assumed to be equal to $0.1\bar{J}$.

Comparing the two SOA types, TW-SOA, Figure 21.18a and b, and R-SOA, Figure 21.18c and d, it can be concluded that, not only does the R-SOA provide higher bandwidths, but it does so for a much wider range of input optical powers. Both SOA types exhibit an increase in $f_{3\mathrm{dB}}$ with the increase of steady-state bias current density, but the behavior with P_0 variation differs significantly. In the case of TW-SOA, the bandwidth generally increases with the increase of P_0, reaching its maximum for deep saturation, and input optical powers between 20 and 25 dBm. Contrarily, R-SOA shows two maxima, providing the choice between two operation regimes, one for the low input optical powers and high device gain, and one for the deep saturation. It should be noted that −3dB bandwidth significantly depends on device length (Totović et al., 2015), which can be optimized to achieve maximum values of $f_{3\mathrm{dB}}$. Aside from changes in values of $f_{3\mathrm{dB}}$ with the length variation, the prominence of low- and high-power maxima in R-SOA can also be

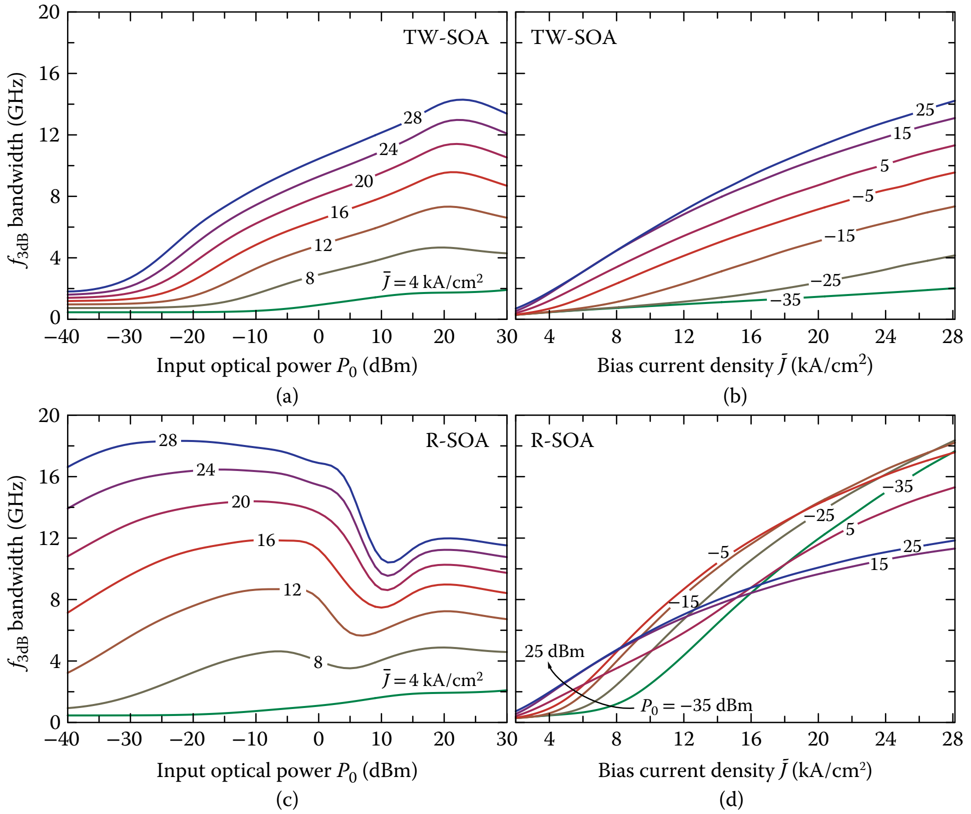


FIGURE 21.18 Bandwidth dependence on (a, c) the continuous wave (CW) input optical power P_0 (dBm) and (b, d) the bias current density $\bar{J}$ (kA/cm^2) for (a), (b) traveling-wave semiconductor optical amplifier (TW-SOA), and (c), (d) reflective (R)-SOA in the case of L = 800 μm.

interchanged. It should be noted that calculated bandwidth is intrinsic, and that inclusion of parasitic effects of the supporting circuit may modify the modulation response.

21.5.7 Case Study: Small-Signal Model for Transparent SOA

Transparent SOA has been repeatedly analyzed as a characteristic example of operating conditions under which the analytical expression for −3dB E/O bandwidth may be obtained (Mørk et al., 1999; Antonelli et al., 2015; Totović et al., 2016). By transparency, it is assumed that the steady-state modal gain is equal to the loss of the amplifier cavity, $\Gamma\bar{g} = \bar{\alpha}_i$, and therefore the steady-state photon density remains unchanged during its propagation. However, even under these circumstances, an analytical expression cannot be obtained without a number of approximations which simplify the system.

21.5.7.1 Framework for the Small-Signal Analysis

In this case study, the expressions will be derived under the following assumptions (Totović et al., 2016):

1. Nonlinear gain suppression is neglected, $\varepsilon\bar{S}_\Sigma \ll 1$.
2. Contribution of ASE noise to the signal is neglected, $\beta_{sp} = 0$.
3. Carrier loss due to the spontaneous emission is modeled via carrier lifetime, τ_s.

4. Current is assumed to reach uniform distribution along the electrode instantaneously, i.e., small-signal current density is assumed to be spatially independent, $\Delta J(z) = \Delta J$.

Within this framework, the system of equations governing signal and carrier dynamics can be derived from Equations 21.115 and 21.116, and reads

$$\pm\frac{\partial S_\pm}{\partial z} + \frac{1}{v_\mathrm{g}}\frac{\partial S_\pm}{\partial t} = \left(\Gamma g - \alpha_\mathrm{i}\right) S_\pm, \tag{21.129}$$

$$\frac{\mathrm{d}n}{\mathrm{d}t} = \frac{J}{qH} - \frac{n}{\tau_\mathrm{s}} - v_\mathrm{g} g S_\Sigma. \tag{21.130}$$

Assuming that the bias current has the general form given by Equation 21.117, $J(z,t) = J(t) = \bar{J} + \Delta J \exp(i\Omega t)$, the system (Equations 21.129 and 21.130) can be decoupled to the steady-state and small-signal equations. After linearization, small-signal carrier density has the form similar to Equation 21.125:

$$\Delta n = \frac{\Delta J/(qH) - v_\mathrm{g}\bar{g}\left(\Delta S_+ + \Delta S_-\right)}{1/\tau_\mathrm{s} + v_\mathrm{g} g_0/\bar{n} \cdot \bar{S}_\Sigma + i\Omega}, \tag{21.131}$$

and the system of equations written with respect to small-signal photon densities reads

$$\pm\frac{\mathrm{d}\Delta S_\pm}{\mathrm{d}z} = \frac{\Gamma\gamma_\pm}{qHv_\mathrm{g}}\Delta J + \left(\Gamma\bar{g} - \bar{\alpha}_\mathrm{i} - i\frac{\Omega}{v_\mathrm{g}}\right)\Delta S_\pm - \Gamma\bar{g}\gamma_\pm\left(\Delta S_\pm + \Delta S_\mp\right), \tag{21.132}$$

where

$$\gamma_\pm = \frac{v_\mathrm{g}\left(g_0/\bar{n} - K_1\right)}{1/\tau_s + v_\mathrm{g} g_0/\bar{n} \cdot \bar{S}_\Sigma + i\Omega}\bar{S}_\pm \tag{21.133}$$

has a meaning similar to the γ-parameter defined by Equation 21.128. Although the system given by Equation 21.132 is analytically solvable, the complexity of the solution undermines its benefit, and prevents us from determining the analytical solution for $\Omega_{3\mathrm{dB}}$.

21.5.7.2 Employing the Transparency Condition

Choosing the transparency as the operating regime defines the steady-state values of carrier and photon densities. From the condition $\Gamma\bar{g} = \bar{\alpha}_\mathrm{i}$, the value for carrier density can be determined:

$$\bar{n} = -\frac{g_0}{K_1} W_L\left[-\frac{K_1 n_\mathrm{tr}}{g_0}\exp\left(\frac{K_0}{\Gamma g_0}\right)\right], \tag{21.134}$$

where $W_\mathrm{L}(x)$ stands for the Lambert W (product-logarithm) function. Introducing the transparency condition in the steady-state form of Equation 21.129 results in $\mathrm{d}S_\pm/\mathrm{d}z = 0$, meaning that the steady-state photon density is not spatially dependent and is defined by the boundary conditions. For both TW- and R-SOA, the forward-propagating steady-state photon density $\bar{S}_+$ will be equal to the input photon density S_0, whereas the backward-propagating photon density will either be zero, for TW-SOA, or $\bar{S}_- = R_2\bar{S}_+ = R_2 S_0$, for R-SOA. Substituting $\bar{S}_\pm$ in the steady-state form of Equation 21.130 gives correlation between the stationary bias current density and the input photon density, for which the transparency condition is met:

$$S_0 = \frac{1}{v_\mathrm{g} g_0 \ln\left(\bar{n}/n_\mathrm{tr}\right)\left(1 + R_2\right)}\left(\frac{\bar{J}}{qH} - \frac{\bar{n}}{\tau_\mathrm{s}}\right). \tag{21.135}$$

Aside from the maximum value of bias current density, dictated by the long-term stable SOA operation, Equation 21.135 gives a limit for the minimal value of $\bar{J}_{\mathrm{min}} = qH\bar{n}/\tau_{\mathrm{s}}$ since the input photon density cannot be negative.

Introducing the transparency condition in Equation 21.132 gives

$$\pm\frac{\mathrm{d}\Delta S_{\pm}}{\mathrm{d}z} = \frac{\Gamma\gamma_{\pm}}{qHv_{\mathrm{g}}}\Delta J - i\frac{\Omega}{v_{\mathrm{g}}}\Delta S_{\pm} - \Gamma\bar{g}\gamma_{\pm}\left(\Delta S_{\pm} + \Delta S_{\mp}\right). \tag{21.136}$$

Moreover, having a fixed value for $\bar{n}$ reduces the γ-parameter to a spatially independent value, meaning that Equation 21.136 is the system of differential equations of the first order with fixed coefficients. The output small-signal photon density of TW-SOA has the following form

$$\Delta S_{+}(L) = \frac{\Gamma}{qHv_{\mathrm{g}}}\Delta J\frac{\gamma_{+}}{\Gamma\bar{g}\gamma_{+} + i\Omega/v_{\mathrm{g}}}\left\{1 - \exp\left[-\left(\Gamma\bar{g}\gamma_{+} + i\frac{\Omega}{v_{\mathrm{g}}}\right)L\right]\right\}, \tag{21.137}$$

whereas the output small-signal photon density of an R-SOA reads

$$\Delta S_{-}(0) = 2R_2\frac{\Gamma\gamma_{+}}{qHv_{\mathrm{g}}}\Delta J \times \left\{\frac{1+3R_2}{2}\Gamma\bar{g}\gamma_{+} + i\frac{\Omega}{v_{\mathrm{g}}} + \sqrt{\left(\frac{1-R_2}{2}\Gamma\bar{g}\gamma_{+}\right)^2 + \left[\left(1+R_2\right)\Gamma\bar{g}\gamma_{+} + i\frac{\Omega}{v_{\mathrm{g}}}\right] i\frac{\Omega}{v_{\mathrm{g}}}}\right.$$
$$\left.\times \coth\left[L\sqrt{\left(\frac{1-R_2}{2}\Gamma\bar{g}\gamma_{+}\right)^2 + \left[\left(1+R_2\right)\Gamma\bar{g}\gamma_{+} + i\frac{\Omega}{v_{\mathrm{g}}}\right] i\frac{\Omega}{v_{\mathrm{g}}}}\right]\right\}^{-1}. \tag{21.138}$$

21.5.7.3 Analytical Formulae for Modulation Bandwidth

As previously discussed, the γ-parameter influences the SOA's small-signal modulation response through its dependence on Ω. For zero frequency, $\gamma_{\pm}$ becomes purely real, whereas for high frequencies, close to $\Omega_{3\mathrm{dB}}$ and beyond, it can be approximated with purely imaginary parameter

$$\gamma_{\pm}\big|_{\Omega\to\Omega_{3\mathrm{dB}}} \approx v_{\mathrm{g}}\left(g_0/\bar{n} - K_1\right)\bar{S}_{\pm}\Big/(i\Omega). \tag{21.139}$$

In addition, during the derivation of the analytical expression for $\Omega_{3\mathrm{dB}}$, all trigonometric functions that are encountered, namely $\sin(x)$ for $x \to 0$ in case of TW-SOA and $\coth(x)$ for $x \to 0$ in the case of R-SOA, are approximated by the first term of their respective Maclaurin series expansion, $\sin(x) \approx x$ and $\coth(x) \approx 1/x$, respectively. This leads to the simpler form of Equations 21.137 and 21.138 in the vicinity of $\Omega_{3\mathrm{dB}}$. For TW-SOA, the output small-signal photon density is

$$\left|\Delta S_{+}(L)\right|\big|_{\Omega\to\Omega_{3\mathrm{dB}}} \approx \frac{\Gamma\Delta J}{qH}\left(\frac{g_0}{\bar{n}} - K_1\right)L\frac{S_0}{\Omega}, \tag{21.140}$$

and for R-SOA

$$\left|\Delta S_{-}(0)\right|\big|_{\Omega\to\Omega_{3\mathrm{dB}}} \approx 2R_2\frac{\Gamma\Delta J}{qH}\left(\frac{g_0}{\bar{n}} - K_1\right)L\frac{S_0}{\Omega}. \tag{21.141}$$

Finally, from the condition $|\Delta S_{\mathrm{out}}(\Omega_{3\mathrm{dB}})/\Delta S_{\mathrm{out}}(0)| = 1/2$, the −3dB bandwidth can be found, under the transparency regime. For TW-SOA, it is given by

$$\Omega_{3\mathrm{dB}} = \Gamma g_0 \ln \frac{\bar{n}}{n_{\mathrm{tr}}} L v_{\mathrm{g}} \left(\frac{g_0}{\bar{n}} - K_1 \right) S_0 \left\{ 1 + \coth \left[\Gamma g_0 \ln \frac{\bar{n}}{n_{\mathrm{tr}}} \frac{L}{2} \frac{v_{\mathrm{g}} \left(g_0/\bar{n} - K_1 \right) S_0}{1/\tau_{\mathrm{s}} + v_{\mathrm{g}} g_0/\bar{n} \cdot S_0} \right] \right\}, \qquad (21.142)$$

and for R-SOA it has the form

$$\Omega_{3\mathrm{dB}} = 2 \left[\frac{1}{\tau_{\mathrm{s}}} + v_{\mathrm{g}} \frac{g_0}{\bar{n}} \left(1 + R_2 \right) S_0 \right] + \left(1 + 3R_2 \right) \Gamma g_0 \ln \frac{\bar{n}}{n_{\mathrm{tr}}} L v_{\mathrm{g}} \left(\frac{g_0}{\bar{n}} - K_1 \right) S_0. \qquad (21.143)$$

21.6 Conclusion

SOAs have been extensively studied ever since WDM networks became widely adopted. In order to enable massive deployment of optical access networks, optical technologies, including SOAs, have expanded toward the lower levels of network hierarchy, aiming to relieve first/last mile bottleneck. Due to their versatility, SOAs are the key components in many of these levels. R-SOA has recently become one of the most promising candidates for the next-generation wavelength division multiplexed passive optical networks (WDM-PONs), representing the new Fiber-to-the-Home paradigm. Moreover, it is on its way to become a crucial component in Worldwide Interoperability for Microwave Access (WiMAX) and Wireless-Fidelity (Wi-Fi) Radio-over-Fiber (RoF) access network architecture. TW- and R-SOAs are devices of choice for many nonlinear applications, including those related to FWM, wavelength conversion, and all optical signal processing in general.

The choice of the SOA type for a specific purpose does not necessarily require better overall performance but usually better performance in one or several aspects. Therefore, both TW- and R-SOA have found their applications. However, optimization of the devices themselves, as well as the system as a whole, should always be performed in order for the SOAs to reach their full potential. The optimization of the device includes the choice of material, device geometry, and operating conditions, whereas the system analysis involves many more aspects that need to be addressed. This process requires an elaborate, but efficient SOA model, which accounts for all relevant effects. In Section 21.5, various models have been presented for both the steady-state and dynamic operating regimes. All models are built starting from the rate equations describing SOA dynamics, which can include multiple levels of approximation, as discussed in Section 21.4. The basic model recognizes only the essential processes present in the SOAs, i.e., carrier injection, stimulated and spontaneous recombination, and signal amplification. More realistic models comprise additional effects, such as wideband ASE noise, carrier transport, nonlinear gain suppression, temperature effects, and distributed bias current, all of which are discussed in detail. Moreover, in order for the model to accurately represent the SOA's performance, a detailed analysis of the optical properties of materials used for the active region and the waveguide geometry should be performed, which can be done using the models presented in Section 21.3. This includes calculation of the full spectral profile of material gain, radiative spontaneous recombination rate, refractive index variation, and optical confinement factor, among other parameters.

For SOA optimization in the steady-state operating regime, the wideband model, based on the self-consistent iterative method, given in Section 21.5.3, can provide detailed results, but does consume a significant amount of computational resources. This method is mainly designed for the characterization and optimization of the device itself. On the other hand, for quick assessment of the SOA's performance, especially in more complex network architectures, the semi-analytical model, presented in Section 21.5.4, proves to be time and resource efficient, at the expense of somewhat decreased precision. In the dynamic operating regime, the upwind scheme finite difference model can be used, as shown in Section 21.5.5, which is transparent to the modulation type, and designed to support even advanced modulation formats. For SOAs used in direct E/O modulation, −3dB bandwidth is a good indicator of the SOAs' performance, and can be determined and optimized based on the detailed numerical model given in Section 21.5.6. When resources are limited, and the SOA operates in the transparency regime, simple analytical solutions, derived in Section 21.5.7, can be used for the same purpose.

The overview of various steady-state and dynamic models, as well as the case studies presented, provides insight into the methods for numerically accurate and efficient simulations of standalone and system-embedded SOAs. These methods can help in the development of computational tools which can be used for the optimization of SOA design based on the dependence of its performance on technological parameters and operating conditions. It should be noted that advanced SOA models, which deal with nonlinear effects, are usually built upon the models which simulate standard SOA amplification regime.

References

Adachi S (1989) Optical dispersion relations for GaP, GaAs, GaSb, InP, InAs, InSb, $Al_xGa_{1-x}As$, and $In_{1-x}Ga_xAs_yP_{1-y}$. *Journal of Applied Physics*. 66:6030–6040.

Adams MJ, Collins JV, Henning ID (1985) Analysis of semiconductor laser optical amplifiers. *IEEE Proceedings Journal—Optoelectronics*. 132:58–63.

Agrawal GP (1990) Effect of gain and index nonlinearities on single-mode dynamics in semiconductor lasers. *IEEE Journal of Quantum Electronics*. 26:1901–1909.

Agrawal GP, Olsson NA (1989) Self-phase modulation and spectral broadening of optical pulses in semiconductor laser amplifiers. *IEEE Journal of Quantum Electronics*. 25:2297–2306.

Ahn D, Chuang SL (1990) Optical gain and gain suppression of quantum-well lasers with valence band mixing. *IEEE Journal of Quantum Electronics*. 26:13–24.

Antonelli C, Mecozzi A (2013) Reduced model for the nonlinear response of reflective semiconductor optical amplifiers. *IEEE Photonics Technology Letters*. 25:2243–2246.

Antonelli C, Mecozzi A, Zhefeng H, Santagiustina M (2015) Analytic study of the modulation response of reflective semiconductor optical amplifiers. *Journal of Lightwave Technology*. 33:4367–4376.

Bennett BR, Soref RA, del Alamo JA (1990) Carrier-induced change in refractive index of InP, GaAs and InGaAsP. *IEEE Journal of Quantum Electronics*. 26:113–122.

Berg TW, Bischoff S, Magnusdottir I, Mørk J (2001) Ultrafast gain recovery and modulation limitations in self-assembled quantum-dot devices. *IEEE Photonics Technology Letters*. 13:541–543.

Berglind E, Gillner L (1994) Optical quantum noise treated with classical electrical network theory. *IEEE Journal of Quantum Electronics*. 30:846–853.

Carapellese N, Tornatore M, Pattavina A (2014) Energy-efficient baseband unit placement in a fixed/mobile converged WDM aggregation network. *IEEE Journal on Selected Areas in Communications*. 32:1542–1551.

Carlo AD, Reale A, Tocca L, Lugli P (1998) Polarization-independent-strained semiconductor optical amplifiers: A tight-binding study. *IEEE Journal of Quantum Electronics*. 34:1730–1739.

Carrère H, Truong VG, Marie X, Brenot R, De Valicourt G, Lelarge F, Amand T (2010) Large optical bandwidth and polarization insensitive semiconductor optical amplifiers using strained InGaAsP quantum wells. *Applied Physics Letters*. 97:121101.

Cartoixa X, Ting DZY, McGill TC (2003) Numerical spurious solutions in the effective mass approximation. *Journal of Applied Physics*. 93:3974–3981.

Cassioli D, Scotti S, Mecozzi A (2000) A time-domain computer simulator of the nonlinear response of semiconductor optical amplifiers. *IEEE Journal of Quantum Electronics*. 36:1072–1080.

Chao CYP, Chuang SL (1992) Spin-orbit-coupling effects on the valence-band structure of strained semiconductor quantum wells. *Physical Review B*. 46:4110–4122.

Cho YS, Choi WY (2001) Analysis and optimization of polarization-insensitive semiconductor optical amplifiers with delta-strained quantum wells. *IEEE Journal of Quantum Electronics*. 37:574–579.

Chu CYJ, Ghafouri-Shiraz H (1994) Analysis of gain and saturation characteristics of a semiconductor laser optical amplifier using transfer matrices. *Journal of Lightwave Technology*. 12:1378–1386.

Chuang SL (1991) Efficient band-structure calculations of strained quantum wells. *Physical Review B*. 43:9649–9661.

Chuang SL (1995) *Physics of Optoelectronic Devices*. New York, NY: Wiley-Interscience.

Coldren LA, Corzine SW, Mašanović ML (2012) *Diode Lasers and Photonic Integrated Circuits*. New York, NY: Wiley-Interscience.

Connelly MJ (2001) Wideband semiconductor optical amplifier steady-state numerical model. *IEEE Journal of Quantum Electronics*. 37:439–447.

Connelly MJ (2002) *Semiconductor Optical Amplifiers*. Norwell, MA: Kluwer.

Connelly MJ (2007) Wide-band steady-state numerical model and parameter extraction of a tensile-strained bulk semiconductor optical amplifier. *IEEE Journal of Quantum Electronics*. 43:47–56.

Connelly MJ (2012) Reflective semiconductor optical amplifier pulse propagation model. *IEEE Photonics Technology Letters*. 24:95–97.

Connelly MJ (2014) Reflective semiconductor optical amplifier modulator dynamic model. *14th International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD)*. 2014:41–42.

Connelly MJ (2015) Reflective semiconductor optical amplifier electrode voltage based phase shifter model. *International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD)*. 2015:39–40.

D'Ottavi A, Iannone E, Mecozzi A, Scotti S, Spano P, Dall'Ara R, Eckner J, Guekos G (1995) Efficiency and noise performance of wavelength converters based on FWM in semiconductor optical amplifiers. *IEEE Photonics Technology Letters*. 7:357–359.

Dailey JM, Koch TL (2007) Impact of carrier heating on SOA transmission dynamics for wavelength conversion. *IEEE Photonics Technology Letters*. 19:1078–1080.

Dailey JM, Koch TL (2009) Simple rules for optimizing asymmetries in SOA-based Mach–Zehnder wavelength converters. *Journal of Lightwave Technology*. 27:1480–1488.

Davis MG, O'Dowd RF (1994) A transfer matrix method based large-signal dynamic model for multielectrode DFB lasers. *IEEE Journal of Quantum Electronics*. 30:2458–2466.

de Oliveira Ribeiro R, Pontes MJ, Giraldi MTMR, Carvalho MCR (2005) Characterisation of all-optical wavelength conversion by cross-gain modulation of ASE on a SOA. International Conference on Microwave and Optoelectronics (SBMO/IEEE MTT-S) 2005:218–221.

de Valicourt G (2012) Next generation of optical access network based on reflective-SOA, Ch. 1. In *Selected Topics on Optical Amplifiers in Present Scenario*, edited by S. K. Garai. Rijeka, Croatia: InTech.

de Valicourt G, Violas MA, Wake D, Van Dijk F, Ware C, Enard A, Make D, Liu Z, Lamponi M, Duan GH, Brenot R (2010) Radio-over-fiber access network architecture based on new optimized RSOA devices with large modulation bandwidth and high linearity. *IEEE Transactions on Microwave Theory and Techniques*. 58:3248–3258.

Donati S, Giuliani G (1997) Noise in an optical amplifier: Formulation of a new semiclassical model. *IEEE Journal of Quantum Electronics*. 33:1481–1488.

Dúill SPÓ, Barry LP (2015) Improved reduced models for single-pass and reflective semiconductor optical amplifiers. *Optics Communications*. 334:170–173.

Durhuus T, Mikkelsen B, Stubkjaer KE (1992) Detailed dynamic model for semiconductor optical amplifiers and their crosstalk and intermodulation distortion. *Journal of Lightwave Technology*. 10:1056–1065.

Figueiredo RC, Ribeiro NS, Gallep CM, Conforti E (2015) Frequency- and time-domain simulations of semiconductor optical amplifiers using equivalent circuit modeling. *Optical Engineering*. 54:114107.

Foreman BA (1993) Effective-mass Hamiltonian and boundary conditions for the valence bands of semiconductor microstructures. *Physical Review B*. 48:4964–4967.

Foreman BA (1997) Elimination of spurious solutions from eight-band k.p theory. *Physical Review B*. 56:12748–12751.

Gebrewold SA, Bonjour R, Barbet S, Maho A, Brenot R, Chanclou P, Brunero M, Marazzi L, Parolari P, Totovic A, Gvozdic D, Hillerkuss D, Hafner C, Leuthold J (2015) Self-seeded RSOA-fiber cavity lasers vs. ASE spectrum-sliced or externally seeded transmitters—A comparative study. *Applied Sciences*. 5:1922–1941.

Ghafouri-Shiraz H (2004) *The Principles of Semiconductor Laser Diodes and Amplifiers*. London Imperial College Press.

Gioannini M (2004) Numerical modeling of the emission characteristics of semiconductor quantum dash materials for lasers and optical amplifiers. *IEEE Journal of Quantum Electronics*. 40:364–373.

Gladisch A, Braun RP, Breuer D, Ehrhardt A, Foisel HM, Jaeger M, Leppla R, Schneiders M, Vorbeck S, Weiershausen W, Westphal F (2006) Evolution of terrestrial optical system and core network architecture. *Proceedings of the IEEE*. 94:869–891.

Godefroy A, Le Corre A, Clerot F, Salaan S, Loualiche S, Simon JC, Henry L, Vaudry C, Keromnes JC, Joulie G, Lamouler P (1995) 1.55-μm polarization-insensitive optical amplifier with strain-balanced superlattice active layer. *IEEE Photonics Technology Letters*. 7:473–475.

Grundmann M, Stier O, Bimberg D (1995) InAs/GaAs pyramidal quantum dots: Strain distribution, optical phonons, and electronic structure. *Physical Review B*. 52:11969–11981.

Gvozdić DM, Ekenberg U (2006) Superefficient electric-field-induced spin-orbit splitting in strained p-type quantum wells. *Europhysics Letters*. 73:927–933.

Hairer E, Wanner G (1996) *Solving Ordinary Differential Equations II: Stiff and Differential-Algebraic Problems*. Berlin: Springer-Verlag.

Henry CH (1982) Theory of the linewidth of semiconductor lasers. *IEEE Journal of Quantum Electronics*. 18:259–264.

Henry CH (1986) Theory of spontaneous emission noise in open resonators and its application to lasers and optical amplifiers. *Journal of Lightwave Technology*. 4:288–297.

Ito T, Yoshimoto N, Magari K, Sugiura H (1998) Wide-band polarization-independent tensile-strained InGaAs MQW-SOA gate. *IEEE Photonics Technology Letters*. 10:657–659.

Jin CY, Guo WH, Huang YZ, Yu LJ (2003) Photon iterative numerical technique for steady-state simulation of gain-clamped semiconductor optical amplifiers. *IEEE Proceedings—Optoelectronics*. 150:503–507.

Joergensen C, Danielsen SL, Stubkjaer KE, Schilling M, Daub K, Doussiere P, Pommerau F, Hansen PB, Poulsen HN, Kloch A, Vaa M, Mikkelsen B, Lach E, Laube G, Idler W, Wunstel K (1997) All-optical wavelength conversion at bit rates above 10 Gb/s using semiconductor optical amplifiers. *IEEE Journal of Selected Topics in Quantum Electronics*. 3:1168–1180.

Joma M, Horikawa H, Xu CQ, Yamada K, Katoh Y, Kamijoh T (1993) Polarization insensitive semiconductor laser amplifier with tensile strained InGaAsP/InGaAsP multiple quantum well structure. *Applied Physics Letters*. 62:121–122.

Kang JM, Lee SH, Kwon HC, Han SK (2006) WDM-PON with broadcasting function using direct ASE modulation of reflective SOA. *European Conference on Optical Communications (ECOC)*. 2006: 24–28.

Kani J-I (2010) Enabling technologies for future scalable and flexible WDM-PON and WDM/TDM-PON systems. *IEEE Journal of Selected Topics in Quantum Electronics*. 16:1290–1297.

Kawano K, Kitoh T (2001) *Introduction to Optical Waveguide Analysis: Solving Maxwell's Equations and the Schrödinger Equation*. New York, NY: Wiley-Interscience.

Keating T, Jin X, Chuang SL, Hess K (1999) Temperature dependence of electrical and optical modulation responses of quantum-well lasers. *IEEE Journal of Quantum Electronics*. 35:1526–1534.

Kim J, Chuang SL (2006) Theoretical and experimental study of optical gain, refractive index change, and linewidth enhancement factor of p-doped quantum-dot lasers. *IEEE Journal of Quantum Electronics*. 42:942–952.

Kim J, Laemmlin M, Meuer C, Bimberg D, Eisenstein G (2008) Static gain saturation model of quantum-dot semiconductor optical amplifiers. *IEEE Journal of Quantum Electronics*. 44:658–665.

Kim J, Laemmlin M, Meuer C, Bimberg D, Eisenstein G (2009) Theoretical and experimental study of high-speed small-signal cross-gain modulation of quantum-dot semiconductor optical amplifiers. *IEEE Journal of Quantum Electronics*. 45:1658–1666.

Kim Y, Lee H, Kim S, Ko J, Jeong J (1999) Analysis of frequency chirping and extinction ratio of optical phase conjugate signals by four-wave mixing in SOA's. *IEEE Journal of Selected Topics in Quantum Electronics*. 5:873–879.

Koenig S, Bonk R, Schmuck H, Poehlmann W, Pfeiffer Th, Koos C, Freude W, Leuthold J (2014) Amplification of advanced modulation formats with a semiconductor optical amplifier cascade. *Optics Express*. 22:17854–17871.

Kolokolov KI, Li J, Ning CZ (2003) k·p Hamiltonian without spurious-state solution. *Physical Review B*. 68:161308.

Labukhin D, Li X (2006) Polarization insensitive asymmetric ridge waveguide design for semiconductor optical amplifiers and super luminescent light-emitting diodes. *IEEE Journal of Quantum Electronics*. 42:1137–1143.

Liljeberg T, Bowers JE (1997) Velocity mismatch limits in semiconductor lasers and amplifiers. *IEEE Conference Proceedings: 10th Annual Meeting Lasers and Electro-Optics Society Annual Meeting (LEOS '97)*. 1:341–342.

Liu G, Chuang SL (2002) Modeling of Sb-based type-II quantum cascade lasers. *Physical Review B*. 65:165220.

Liu G, Jin X, Chuang SL (2001) Measurement of linewidth enhancement factor of semiconductor lasers using an injection-locking technique. *IEEE Photonics Technology Letters*. 13:430–432.

Loudon R, Ramoo D, Adams MJ (2005) Theory of spontaneous emission noise in multisection semiconductor lasers. *Journal of Lightwave Technology*. 23:2491–2504.

Lysak VV, Kawaguchi H, Sukhoivanov IA, Katayama T, Shulika AV (2005) Ultrafast gain dynamics in asymmetrical multiple quantum-well semiconductor optical amplifiers. *IEEE Journal of Quantum Electronics*. 41:797–807.

Lysak VV, Sukhoivanov IA, Shulika OV, Safonov IM, Lee YT (2006) Carrier tunneling in complex asymmetrical multiple-quantum-well semiconductor optical amplifiers. *IEEE Photonics Technology Letters*. 18:1362–1364.

Magari K, Okamoto M, Noguchi Y (1991) 1.55-μm polarization insensitive high-gain tensile-strained-barrier MQW optical amplifier. *IEEE Photonics Technology Letters*. 3:998–1000.

Magari K, Okamoto M, Suzuki Y, Sato K, Noguchi Y, Mikami O (1994) Polarization-insensitive optical amplifier with tensile-strained-barrier MQW structure. *IEEE Journal of Quantum Electronics*. 30:695–701.

Marcuse D (1984) Computer simulation of laser photon fluctuations: Theory of single-cavity laser. *IEEE Journal of Quantum Electronics*. 20:1139–1148.

Mathlouthi W, Lemieux P, Salsi M, Vannucci A, Bononi A, Rusch LA (2006) Fast and efficient dynamic WDM semiconductor optical amplifier model. *Journal of Lightwave Technology*. 24:4353–4365.

Mathur A, Dapkus PD (1992) Polarization insensitive strained quantum well gain medium for lasers and optical amplifiers. *Applied Physics Letters*. 61:2845–2847.

Mazzucato S, Carrère H, Marie X, Amand T, Achouche M, Caillaud C, Brenot R (2015) Gain, amplified spontaneous emission and noise figure of bulk InGaAs/InGaAsP/InP semiconductor optical amplifiers. *IET Optoelectronics*. 9:52–60.

Mecozzi A, Mørk J (1997) Saturation effects in nondegenerate four-wave mixing between short optical pulses in semiconductor laser amplifiers. *IEEE Journal of Selected Topics in Quantum Electronics*. 3:1190–1207.

Mecozzi A, Wiesenfeld JM (2001) The roles of semiconductor optical amplifiers in optical networks. *Optics & Photonics News*. 12:36–42.

Melo AM, Petermann K (2008) On the amplified spontaneous emission noise modeling of semiconductor optical amplifiers. *Optics Communications*. 281:4598–4605.

Meney AT, Gonul B, O'Reilly EP (1994) Evaluation of various approximations used in the envelope-function method. *Physical Review B*. 50:10893–10904.

Michie C, Kelly AE, McGeough J, Armstrong I, Andonovic I, Tombling C (2006) Polarization-insensitive SOAs using strained bulk active regions. *Journal of Lightwave Technology*. 24:3920–3927.

Morel P, Sharaiha A (2009) Wideband time-domain transfer matrix model equivalent circuit for short pulse propagation in semiconductor optical amplifiers. *IEEE Journal of Quantum Electronics*. 45:103–116.

Mørk J, Mecozzi A, Eisenstein G (1999) The modulation response of a semiconductor laser amplifier. *IEEE Journal of Selected Topics in Quantum Electronics*. 5:851–860.

Moss TS, Burrell GJ, Ellis B (1973) *Semiconductor Opto-Electronics*. London: Butterworth & Co.

Mukai T, Yamamoto Y (1982) Noise in an AlGaAs semiconductor laser amplifier. *IEEE Transactions on Microwave Theory and Techniques*. 30:410–421.

Nagarajan R, Ishikawa M, Fukushima T, Geels RS, Bowers JE (1992) High speed quantum-well lasers and carrier transport effects. *IEEE Journal of Quantum Electronics*. 28:1990–2008.

Occhi L, Schares L, Guekos G (2003) Phase modeling based on the α-factor in bulk semiconductor optical amplifiers. *IEEE Journal of Selected Topics in Quantum Electronics*. 9:788–797.

Olsson NA (1989) Lightwave systems with optical amplifiers. *Journal of Lightwave Technology*. 7:1071–1082.

Osinski M, Buus J (1987) Linewidth broadening factor in semiconductor lasers—An overview. *IEEE Journal of Quantum Electronics*. 23:9–29.

Park J, Li X, Huang WP (2005) Comparative study of mixed frequency-time-domain models of semiconductor laser optical amplifiers. *IEE Proceedings—Optoelectronics*. 152:151–159.

Pham QT (2014) Highly effective crosstalk mitigation method using counter-propagation in semiconductor optical amplifier for remodulation WDM-PONs. *Journal of Photonics*. 2014:610967.

Pillai BSG, Premaratne M, Abramson D, Lee KL, Nirmalathas A, Lim C, Shinada S, Wada N, Miyazaki T. (2006) Analytical characterization of optical pulse propagation in polarization-sensitive semiconductor optical amplifiers. *IEEE Journal of Quantum Electronics*. 42:1062–1077.

Pitris S, Vagionas C, Kanellos GT, Pleros N, Kisacik R, Tekin T, Broeke R (2015) Monolithically integrated all-optical SOA-based SR Flip-Flop on InP platform. *International Conference on Photonics in Switching (PS)*. 2015:208–210.

Qasaimesh O (2003) Optical gain and saturation characteristics of quantum-dot semiconductor optical amplifier. *IEEE Journal of Quantum Electronics*. 39:794–800.

Qasaimesh O (2013) Broadband gain-clamped linear quantum dash optical amplifiers. *Optical and Quantum Electronics*. 45:1277–1286.

Qin C, Huang X, Zhang X (2012) Theoretical investigation on gain recovery dynamics in step quantum well semiconductor optical amplifiers. *Journal of the Optical Society of America B*. 29:607–613.

Ramaswami R (2002) Optical fiber communication: From transmission to networking. *IEEE Communications Magazine*. 40:138–147.

Razaghi M, Ahmadi V, Connelly MJ (2009) Comprehensive finite-difference time-dependent beam propagation model of counterpropagating picosecond pulses in a semiconductor optical amplifier. *Journal of Lightwave Technology*. 27:3162–3174.

Reithmaier JP, Somers A, Deubert S, Schwertberger R, Kaiser W, Forchel A, Calligaro M, Resneau P, Parillaud O, Bansropun S, Krakowski M, Alizon R, Hadass D, Bilenca A, Dery H, Mikhelashvili V, Eisenstein G, Gioannini M, Montrosset I, Berg TW, van der Poel M, Mørk, J, Tromborg B (2005) InP based lasers and optical amplifiers with wire-/dot-like active regions. *Journal of Physics D: Applied Physics*. 38:2088–2102.

Runge P, Elschner R, Petermann K (2010) Time-domain modeling of ultralong semiconductor optical amplifiers. *IEEE Journal of Quantum Electronics*. 46:484–491.

Schmuck H, Bonk R, Poehlmann W, Haslach C, Kuebart W, Karnick D, Meyer J, Fritzsche D, Weis E, Becker J, Freude W, Pfeiffer T (2013) Demonstration of SOA-assisted open metro-access infrastructure for heterogeneous services. *39th European Conference and Exhibition on Optical Communication (ECOC)*. 2013:1–3.

Schrenk B (2011) Characterization and Design of Multifunction Photonic Devices for Next Generation Fiber-to-the-Home Optical Network Units. PhD diss., Universitat Politècnica de Catalunya.

Serrat C, Masoller C (2006) Modeling spatial effects in multi-longitudinal-mode semiconductor lasers. *Physical Review A*. 73:043812.

Shtaif M, Tromborg B, Eisenstein G (1998) Noise spectra of semiconductor optical amplifiers: Relation between semiclassical and quantum descriptions. *IEEE Journal of Quantum Electronics*. 34:869–878.

Silver M, Phillips AF, Adams AR, Greene PD, Collar AJ (2000) Design and ASE characteristics of 1550-nm polarization-insensitive semiconductor optical amplifiers containing tensile and compressive wells. *IEEE Journal of Quantum Electronics*. 36:118–122.

Simon JC, Doussiere P, Pophillat L, Fernier B (1989) Gain and noise characteristics of a 1.5 μm near-travelling-wave semiconductor laser amplifier. *Electronics Letters*. 25:434–436.

Stabile R, Williams KA (2011) Photonic integrated semiconductor optical amplifier switch circuits. In *Advances in Optical Amplifiers*, edited by P. Urquhart, 205–230. Rijeka, Croatia: InTech.

Stier O, Grundmann M, Bimberg D (1999) Electronic and optical properties of strained quantum dots modeled by 8-band k.p theory. *Physical Review B*. 59:5688–5701.

Sygletos S, Tomkos I, Leuthold J (2008) Technological challenges on the road toward transparent networking. *Journal of Optical Networking*. 7:321–350.

Talli G, Adams MJ (2003) Amplified spontaneous emission in semiconductor optical amplifiers: Modelling and experiments. *Optics Communications*. 218:161–166.

Tauber DA, Spickermann R, Nagarajan R, Reynolds T, Holmes AL, Bowers JE (1994) Inherent bandwidth limits in semiconductor lasers due to distributed microwave effects. *Applied Physics Letters*. 64:1610–1612.

Thylén L (1988) Amplified spontaneous emission and gain characteristics of Fabry-Perot and traveling wave type semiconductor laser amplifiers. *IEEE Journal of Quantum Electronics*. 24:1532–1537.

Tiemeijer LF, Thijs PJA, Van Dongen T, Slootweg RWM, van der Heijden JMM, Binsma JJM, Krijn MPCM (1993) Polarization insensitive multiple quantum well laser amplifiers for the 1300 nm window. *Applied Physics Letters*. 62:826–828.

Tishinin D, Uppal K, Kim I, Dapkus PD (1997) 1.3-μm polarization insensitive amplifiers with integrated-mode transformers. *IEEE Photonics Technology Letters*. 9:1337–1339.

Toptchiyski G, Kindt S, Petermann K, Hilliger E, Diez S, Weber HG (1999) Time-domain modeling of semiconductor optical amplifiers for OTDM applications. *Journal of Lightwave Technology*. 17:2577–2583.

Totović AR, Crnjanski JV, Krstić MM, Gvozdić DM (2011) Application of multi-quantum well RSOA in remodulation of 100 Gb/s downstream RZ signal for 10 Gb/s upstream transmission. *19th Telecommunications Forum (TELFOR)*. 2011:840–843.

Totović AR, Crnjanski JV, Krstić MM, Gvozdić DM (2012) Modelling of carrier dynamics in multi-quantum well semiconductor optical amplifiers. *Physica Scripta*. 2012:014032.

Totović AR, Crnjanski JV, Krstić MM, Gvozdić DM (2014) An analytical solution for stationary distribution of photon density in traveling-wave and reflective SOAs. *Physica Scripta*. 2014:014013.

Totović AR, Crnjanski JV, Krstić MM, Gvozdić DM (2014) An efficient semi-analytical method for modeling of traveling-wave and reflective SOAs. *Journal of Lightwave Technology*. 32:2106–2112.

Totović AR, Crnjanski JV, Krstić MM, Gvozdić DM (2015) Numerical study of the small-signal modulation bandwidth of reflective and traveling-wave SOAs. *Journal of Lightwave Technology*. 33:2758–2764.

Totović AR, Crnjanski JV, Krstić MM, Mašanović ML, Gvozdić DM (2013) A self-consistent numerical method for calculation of steady-state characteristics of traveling-wave and reflective SOAs. *IEEE Journal of Selected Topics in Quantum Electronics*. 19:3000411.

Totović AR, Levajac VG, Gvozdić DM (2016) Electro-optical modulation bandwidth analysis for traveling-wave and reflective semiconductor optical amplifiers in transparency operating regime. *Optical and Quantum Electronics*. 48:262.

Tsai CY, Tsai CY, Lo YH, Spencer RM, Eastman LF (1995) Nonlinear gain coefficients in semiconductor quantum-well lasers: Effects of carrier diffusion, capture, and escape. *IEEE Journal of Selected Topics in Quantum Electronics*. 1:316–330.

Valiente I, Lablonde L, Simon JC, Billès L (1996) Effects of amplified spontaneous emission on gain recovery dynamics of semiconductor optical amplifiers. In *Optical Amplifiers and Their Applications*, edited by R. Jopson, K. Stubkjaer, and M. Suyama, Vol. 5, *OSA Trends in Optics and Photonics Series*. Washington, DC: Optical Society of America.

Veprek RG, Steiger S, Witzigmann B (2007) Ellipticity and the spurious solution problem of $k \cdot p$ envelope equations. *Physical Review B*. 76:165320.

Visser TD, Blok H, Demeulenaere B, Lenstra D (1997) Confinement factors and gain in optical amplifiers. *IEEE Journal of Quantum Electronics*. 33:1763–1766.

Vyrsokinos K, Vagionas C, Fitsios D, Miliou A (2014) Frequency and time domain analysis of all optical memories based on SOA and SOA-MZI switches. *IEEE Optical Interconnects Conference*. 2014: 57–58.

Wang J, Maitra A, Poulton CG, Freude W, Leuthold J (2007) Temporal dynamics of the alpha factor in semiconductor optical amplifiers. *Journal of Lightwave Technology*. 25:891–900.

Willatzen M, Uskov A, Mørk J, Olesen H, Tromborg B, Jauho AP (1991) Nonlinear gain suppression in semiconductor lasers due to carrier heating. *IEEE Photonics Technology Letters*. 3:606–609.

Wolfson D. All-optical signal processing and regeneration. PhD diss., Research Center COM, Technical University of Denmark, 2001.

Wu B, Georges JB, Cutrer DM, Lau KY (1995) On distributed microwave effects in semiconductor lasers and their practical implications. *Applied Physics Letters*. 67:467–469.

Wünstel K, Laube G, Idler W, Daub K, Lach E, Dütting K, Klenk M, Schilling M (1996) High speed and polarization insensitive wavelength converters by MQW optimization, In *Conference Proceedings of IPR'96*, paper IMG6, Boston, MA.

Xia M, Ghafouri-Shiraz H (2015) Theoretical analysis of carrier heating effect in semiconductor optical amplifiers. *Optical and Quantum Electronics*. 47:2141–2153.

Xia M, Ghafouri-Shiraz H (2016) A novel transmission line model for quantum well semiconductor optical amplifiers. *Optical and Quantum Electronics*. 48:52.

Xing W, Yikai S, Xiang L, Leuthold J, Chandrasekhar S (2004) 10-Gb/s RZ-DPSK transmitter using a saturated SOA as a power booster and limiting amplifier. *IEEE Photonics Technology Letters*. 16:1582–1584.

Yacomotti AM, Furfaro L, Hachair X, Pedaci F, Giudici M, Tredicce J, Javaloyes J, Balle S, Viktorov EA, Mandel P (2004) Dynamics of multimode semiconductor lasers. *Physical Review A*. 69:053816.

Yamatoya T, Koyama F (2004) Optical preamplifier using optical modulation of amplified spontaneous emission in saturated semiconductor optical amplifier. *Journal of Lightwave Technology*. 22:1290–1295.

Yariv A, Yeh P. *Photonics: Optical Electronics in Modern Communications*. New York, NY: Oxford University Press, 2007.

Zhang Y, Ruden PP (1999) 1.3 μm polarization-insensitive optical amplifier structure based on coupled quantum wells. *IEEE Journal of Quantum Electronics*. 35:1509–1514.

Zhou E, Zhang X, Huang D (2007) Analysis on dynamic characteristics of semiconductor optical amplifiers with certain facet reflection based on detailed wideband model. *Optics Express*. 15:9096–9106.

Zilkie AJ, Meier J, Mojahedi M, Poole PJ, Barrios P, Poitras D, Rotter TJ, Yang C, Stintz A, Malloy KJ, Smith PWE, Aitchison JS (2007) Carrier dynamics of quantum-dot, quantum-dash, and quantum-well semiconductor optical amplifiers operating at 1.55 μm. *IEEE Journal of Quantum Electronics*. 43:982–991.

Zimmerman DR, Spiekman LH (2004) Amplifiers for the masses: EDFA, EDWA, and SOA amplets for metro and access applications. *Journal of Lightwave Technology*. 22:63–70.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

22

Tapered Semiconductor Optical Amplifiers

22.1 Introduction 697
22.2 Modeling Approaches 698
Carrier Transport Models • Optical Models • Thermal Models • Material Properties
22.3 Model Implementation 703
Overview of Solution Procedure • Solver for Electrical Equations • Solver for Optical Equations • Solver for Thermal Equations
22.4 Simulation Example: 1.5-μm InGaAsP/InP Tapered Amplifier 706
Device Geometry and Simulation Parameters • Analysis of the Reference SOA • Effect of the Confinement Factor and Taper Angle
22.5 Summary 712

José-Manuel G. Tijero

Antonio Pérez-Serrano

Gonzalo del Pozo

and

Ignacio Esquivias

22.1 Introduction

In recent years, the pervasive extension of semiconductor-based laser sources has been reaching application domains, such as lidar, material processing, metrology, and frequency doubling, where other laser sources are still prevalent because of the concurrent requirements of high power, high spectral purity, and high-beam quality. Two slightly different semiconductor devices are already in use for some of these applications: tapered laser diodes (Walpole, 1996; Wenzel et al., 2003; Sumpf et al., 2009) and master-oscillator power-amplifiers (MOPAs) (O'Brien et al., 1993; Spreemann et al., 2009). A tapered MOPA consists of either a distributed Bragg reflector laser or a distributed feedback laser acting as the master oscillator (MO) and a tapered semiconductor optical amplifier (SOA) acting as the power amplifier. Both hybrid and integrated MOPAs are of great interest. In a hybrid MOPA, the MO and the SOA are coupled through additional optics (Schwertfeger et al., 2011) with the advantage of separate fabrication of the components but at the expense of a more complex set-up. The integrated MOPAs have the advantages of reduced total size but the drawback of possible optical coupling between sections (Vilera et al., 2015).

The schematic of a typical tapered SOA is shown in Figure 22.1a. It is composed of a straight and narrow index-guided section used to filter the input beam and convert it into a single spatial mode, and a gain-guided tapered section where the beam is amplified while preserving its shape. The tapered SOA of an integrated MOPA does not include an index-guided section since the MO provides a single spatial mode to the tapered section. Cross-sectional views of the index and gain-guided sections are shown in Figure 22.1b. The index-guided section is typically a ridge waveguide (RW) structure in which the etched regions (filled with an isolator) provide lateral optical confinement. In the tapered gain-guided section, the injection area is defined by a proton-implanted region.

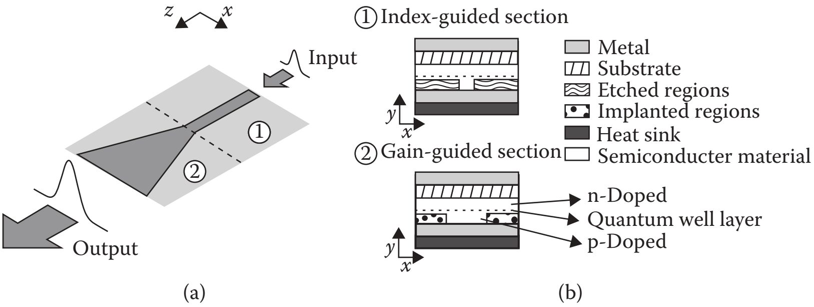


FIGURE 22.1 (a) Schematic planar view of a typical tapered SOA, consisting of a straight and narrow index-guided section and a gain-guided tapered section. (b) Cross-sectional views of the index-guided section (1) and the gain-guided section (2).

The characteristics of the output beam of a tapered SOA are very similar to those of tapered lasers which have been studied in detail by simulations and experiments (Williams et al., 1999; Sujecki et al., 2003; Borruel et al., 2004) and are reviewed in Volume 2 of this book (Chapter 28) (Esquivias et al., 2017). In brief, the single spatial mode of the RW section is launched into the tapered region where it suffers simultaneously diffraction, gain guiding, and index guiding (or antiguiding) due to the carrier and temperature dependencies of the semiconductor refractive index. The full taper angle α_{tap} is usually designed to match the free diffraction angle at $1/e^2$ (Borruel et al., 2008), and therefore it depends on the RW width and the lateral index step. At low power, the beam is strongly astigmatic: the virtual source in the lateral axis (x-axis) is located approximately at a distance from the output facet given by the taper section length divided by the effective index n_{eff}, while the virtual source in the vertical axis (y-axis) is at the output facet. The far-field (FF) pattern is almost Gaussian in both axes and much narrower in the lateral axis. However, when the output power increases, the beam quality degrades due to the combined effects of carrier and thermal lensing, as shown in Section 22.4.

The optimization of the tapered SOA's geometry and epitaxial design is a key point to improve the performance of these devices. The complex interaction between the semiconductor media and the optical field makes it extremely difficult to advance meaningful predictions of the behavior of a specific design by means of simple analytical calculations. Such difficulties together with the complex and costly fabrication process make necessary the use of complete numerical simulation tools to help the analysis and design of tapered SOAs to be used in hybrid or integrated MOPAs. Different numerical models, with different degrees of complexity, have been proposed and used to predict the behavior of tapered SOAs. In the case of the straight SOAs used in communications, the main emphasis has been placed on analyzing the dynamic response and the role of the amplified spontaneous emission in the noise characteristics (Razaghi et al., 2009; Connelly, 2001, and references therein). In contrast, in the case of high-power tapered SOAs the main performance characteristics are the maximum output power and the beam properties, and therefore steady-state models including thermal equations have been employed (Lang et al., 1993; Lai and Lin, 1998; Tijero et al., 2015).

In Section 22.2, we present a review of the different modeling approaches that can be applied to tapered SOAs; in Section 22.3, we explain the implementation of our numerical model; and in Section 22.4 we present as an example the results of the simulations of a 1.5-μm tapered SOA, including the analysis of the influence of some selected design parameters on the device performance. The Section 22.5 summary concludes the chapter.

22.2 Modeling Approaches

There are many options for the modeling approaches needed to simulate tapered SOAs. The proper choice depends on the degree of numerical complexity and on the performance characteristics of interest. In this section, we review the most common approaches, explain the main differences between them, and refer the interested reader to the most important references for a detailed accounts. We will also use references corresponding to tapered lasers and laser diodes in general, as most of the modeling approaches for high-power SOAs are equally appropriate for high-power semiconductor lasers.

One of the main characteristics of tapered SOAs and lasers compared with other semiconductor devices is the nonuniformity of the photon, carrier, and temperature profiles along the longitudinal (z-axis) and the lateral directions. This leads to the need to solve the coupled equations in at least two dimensions and to use and using some approximation to solve the equations in the vertical direction separately.

22.2.1 Carrier Transport Models

The main role of a carrier transport model is to provide the 2D carrier density distribution along the active region, which is needed to calculate the main material parameters affecting the optical wave propagation (optical gain and losses and refractive index variation), as well as to calculate the local heat sources required by the thermal model. But, in turn, the knowledge of the local photon density and the temperature distribution is needed for solving the electrical equations, and therefore iterative schemes are usually applied (see Section 22.3). The external input is the applied voltage (V_0), rather than the total current (I), which is the usual input in experiments. In the RW section, the current flow spreads out of the nominal contact width. Furthermore, the local current density depends not only on the applied voltage but also on the local temperature and stimulated recombination, and hence on the local photon density, which is strongly dependent on the lateral and longitudinal position. In consequence, the current density J is not uniform, but depends on the position in the x–z plane, $J = J(x, z)$.

Most of the published tapered SOA models make use of the unipolar approximation assuming that the electron (n) and hole (p) densities in the active layer are identical (Lang et al., 1993; Lai and Lin, 1998; Spreemann et al., 2009). Under this approximation, which is also very common in laser diode models, the carrier density in the active layer $N(x, z)$ can be calculated using the simple rate equation:

$$\frac{\mathrm{d}N(x,z)}{\mathrm{d}t} = \frac{J(x,z)}{qd_{\mathrm{act}}} - R(N) - R_{\mathrm{stim}}(N,S) + D_{\mathrm{eff}}\nabla^2_{x,z}N(x,z), \tag{22.1}$$

where q is the elementary charge, d_{act} the active layer thickness, $R(N)$ the carrier recombination rate including nonradiative and spontaneous recombination, and $R_{\mathrm{stim}}(N, S)$ the stimulated recombination rate, which depends on the local photon density $S(x, z)$. Finally, finally, $D_{\mathrm{eff}}\nabla^2_{x,z}N(x, z)$ accounts for carrier diffusion in the longitudinal and lateral directions and is perhaps the most important consideration. In this expression, D_{eff} is an effective diffusion coefficient accounting for both the carrier diffusion and the current spreading. The local current density $J(x, z)$ can be expressed in terms of the applied voltage, the contact (and/or p-layer) resistivity, and the junction voltage $V_J(x, z)$, which in turn depends on the carrier density. Details on this formulation can be found in Lai and Lin (1998).

The unipolar approximation is a simple way to reduce the complexity of the carrier transport model and provides very reasonable results, but the p–n junction is essentially a bipolar device, thus requiring the solution of the complete semiconductor drift–diffusion equations. These equations can be expressed in terms of three unknown variables, the electrostatic potential (ϕ) and the electron (n) and hole (p) concentrations. In the steady state, the equations are (Selberherr, 1984)

$$\nabla\left\{\varepsilon_S\nabla\varphi\right) + q\left\{p - n + C_i\right) = 0 \tag{22.2}$$

$$\nabla\mathbf{j_n} - q\left\{R(n,p) + R_{\mathrm{stim}}(n,p,S)\right) = 0 \tag{22.3}$$

$$\nabla\mathbf{j_p} + q\left\{R(n,p) + R_{\mathrm{stim}}(n,p,S)\right) = 0, \tag{22.4}$$

where ε_S is the static dielectric constant, C_i is the charged impurity density (ionized donor density minus ionized acceptor density), and j_n (j_p) is the electron (hole) current density. The current densities can be expressed as a function of the local variables ϕ, n, p, and temperature (T) using as local parameters the electron and hole mobilities and thermal diffusivities.

The carrier transport models often consider the capture and escape of carriers into and out of the quantum wells (QWs) (Nagarajan et al., 1992). This effect is well known in high-speed lasers, and it also influences the steady-state emission properties. When the carrier capture time is not negligible, there is an accumulation of carriers in the confinement layers that can result in a reduction of the efficiency (Borruel et al., 2003). In the case of unipolar models using rate equations, the carrier capture/escape processes are modeled by solving a second equation for the unconfined carriers. In the case of bipolar models of laser diodes using the complete semiconductor equations, different approaches have been proposed (Borruel et al., 2003; Grote et al., 2005), but as far as we know the importance of these effects in the modeling of SOAs has not been analyzed.

Bipolar carrier transport models require also appropriate boundary conditions at the metal–semiconductor contacts (Selberherr, 1984) and at the abrupt heterojunctions, where thermionic emission or tunneling-assisted thermionic emission should be considered instead of the standard drift–diffusion equations.

22.2.2 Optical Models

All optical models start from Maxwell equations and a description of the material susceptibility in the frequency or time domain, leading to optical wave equations, which have to be solved numerically. There are two main types of models: those solving the equations in the time domain, usually known as traveling-wave models (TWM), and frequency domain models. TWM provide dynamic solutions as well as spectral properties, and are appropriate for pulse propagation, beam filamentation, transient phenomena, and multifrequency devices. Frequency domain models are much simpler to implement and are useful for steady-state studies where the overall behavior of the device can be described in terms of a single frequency.

The beam propagation method (BPM) (Van Roey et al., 1981) is the most commonly used frequency domain method for the field propagation in waveguide optoelectronic and fiber devices, thanks to its low computational time demand. BPM calculates the propagation along the longitudinal direction of the steady-state transverse field, verifying the Helmholtz equation:

$$\nabla^2 \vec{E}(x,y,z) + \tilde{n}^2(x,y,z) k_0^2 \vec{E}(x,y,z) = 0, \tag{22.5}$$

where $\vec{E}(x,y,z)$ is the complex optical field vector, $\tilde{n}(x,y,z)$ is the complex refractive index, and k_0 is the wavenumber in vacuum.

In order to obtain a solution for the Helmholtz equation, the scalar approximation is commonly used in the study of tapered lasers and amplifiers (Sujecki et al., 2003; Lang et al., 1993). This approximation ignores the vectorial nature of the field, assuming continuity of the field and its derivatives through the dielectric interfaces. The approximation implies that the optical polarization remains unchanged along the device. The effective refractive index approximation is used to decrease the computational complexity of the BPM (Buus, 1982). It consists of solving Equation 22.5 for the vertical axis at every lateral position, taking into account the existence of etched regions and considering only a real index vertical profile (passive approximation). This leads to an effective index lateral profile and a vertical distribution of the optical mode $f(y)$. The effective index profile is used for the beam propagation and the function $f(y)$ is usually assumed to be independent of the longitudinal and lateral positions. In this way, the original 3D problem becomes a 2D problem in the x–z plane.

Dynamical TWMs naturally incorporate the spatial effects and the multifrequency character of the field, with the geometry of the device being incorporated through the boundary conditions for the waves. Although TWMs have long been applied to study laser physics (Lugiato and Narducci, 1985), application to semiconductor media requires a proper description in the time domain of the spectral dependence of the gain and the index changes due to the injected carriers. TWMs are based on making the slowly varying amplitude approximation or paraxial approximation of the electric fields. Under this approximation, the equations for the electric field of the optical wave can be written as

$$-\frac{i}{2k_0}\frac{\partial^2 E_\pm}{\partial x^2} \pm \frac{\partial E_\pm}{\partial z} + \frac{1}{v_g}\frac{\partial E_\pm}{\partial t} = f(E_+, E_-, P_+, P_-), \tag{22.6}$$

where $E_\pm$ are the complex counter propagating electric fields, v_g is the group velocity, and the function f contains different terms describing the relationship of the field with the dielectric polarization ($P_\pm$) that includes gain, internal losses, Bragg couplings, etc. Several approaches have been used to solve Equation 22.6, including finite-difference time-domain (Fischer et al., 1996) and Fourier split-step methods (Spreemann et al., 2009; Pérez-Serrano et al., 2013). Although TWMs have been applied mainly for describing laser dynamics, they have also been applied to tapered SOAs (Balsamo et al., 1996) and for the analysis of coupled cavity effects in tapered MOPAs (Spreemann et al., 2009; Pérez-Serrano et al., 2013).

22.2.3 Thermal Models

The temperature distribution along the complete device is a relevant variable in the simulation of a tapered SOA due to the dependence of the gain and refractive index on the local temperature. The 3D steady-state heat equation is expressed as

$$\nabla(\kappa\nabla T) + w(x, y, z) = 0, \tag{22.7}$$

where κ is the temperature-dependent thermal conductivity and $w(x, y, z)$ the local density of heat sources. This term usually includes Joule, nonradiative recombination, and free-carrier absorption heat sources. A complete description, based on a rigorous thermodynamic approach (Wachutka, 1990), can be found in Bandelow et al. (2005). Two heat sources are usually not included in most of the laser and amplifier models due to the difficulties in the quantification of their local distribution. These are the sources of heat arising from the energy lost by the scattered stimulated photons and by the spontaneously emitted photons that do not leave the device. After a photon recycling process, their energy ends up in heat transferred to the lattice somewhere in the device. To account for this effect, as well as for other uncertainties in the expressions and parameters for the local heat sources, an additional excess power P_{exc} shall be included (Borruel et al., 2002). The amount of this power is calculated from the energy conservation expression:

$$P_{\mathrm{exc}} = IV_0 - (P_{\mathrm{out}} - P_{\mathrm{in}}) - \int w(x, y, z)\mathrm{d}V, \tag{22.8}$$

where P_{out} and P_{in} are the output and input powers, respectively, and the integration extends over the entire device volume. After determining P_{exc}, the corresponding heat source w_{exc} is either distributed uniformly over the whole device or weighted according to the different absorption properties of the epitaxial layers.

Thermal equations should be solved taking into account the boundary conditions imposed by the thermal properties and the geometry of the mount and submount and the heatsink temperature.

22.2.4 Material Properties

22.2.4.1 Material Susceptibility

There are different options to describe the interaction between the gain medium and the optical field. The simplest approach, which is usually employed in rate equation models, is to consider a logarithmic or linear dependence of the material gain on the carrier density, together with a constant value for the linewidth enhancement factor (LEF) (Lai and Lin, 1998). The gain spectra can be taken as Lorentzian (Spreemann et al., 2009), although the most sophisticated models include a complete solution of the band dispersion equations to calculate the material gain as a function of the local carrier densities, the wavelength, and the temperature. A complete microscopic 8×8 $\mathbf{k}{\cdot}\mathbf{p}$ model for the semiconductor band structure implies a large computational effort and even a more simplified 4×4 valence band mixing model results in a high computation time when repeated locally. Additionally, most of the input parameters for the complete models are not well known. For these reasons, some modeling approaches perform initial complete or semicomplete band structure calculations and either (1) use a parabolic band model previously fitted to the results of the complete model (Borruel et al., 2004), (2) use curve fits to a more sophisticated model (Mariojouls et al., 2000), or (3) use precalculated look-up tables (Koch et al., 2005).

The refractive index's dependence on the wavelength, carrier density, and temperature can also be calculated using complete microscopic approaches (Koch et al., 2005) and then again using look-up tables to decrease the computational cost. A simpler approach is to neglect the wavelength dependence of the index and to use simple expressions for the carrier and temperature dependencies, i.e.,

$$\delta\tilde{n} = \delta\tilde{n}_T + \delta\tilde{n}_N, \tag{22.9}$$

where $\delta\tilde{n}$ is the variation of the active layer index with respect to its reference value at the heatsink temperature and without injection, $\delta\tilde{n}_T$ is the thermally induced variation, and $\delta\tilde{n}_N$ is the carrier contribution to the variation. $\delta\tilde{n}_T$ is usually considered a linear function of the local temperature increase while $\delta\tilde{n}_N$ can be considered either as dependent on the carrier density (Connelly, 2001; Borruel et al., 2004) or it is calculated from the gain variations using a constant LEF (Lang et al., 1993; Mariojouls et al., 2000).

Full space-time models require a good description of the material complex susceptibility, which has been done in different ways: at microscopic level (Gehrig and Hess, 2001), by using multiple Lorentzians fitted either to experiments or to microscopic calculations (Moloney et al., 1997), and by a convolution integral (Javaloyes and Balle, 2010).

22.2.4.2 Carrier Recombination

The unipolar models typically use a polynomial approximation for the carrier recombination rate in Equation 22.1, which can be given by

$$R(n) = An + Bn^2 + Cn^3, \tag{22.10}$$

where A, B, and C are the coefficients corresponding to nonradiative trap recombination, spontaneous recombination, and Auger recombination, respectively. It is important to remark that even in this simple description, the dependence of the coefficients on temperature, especially the dependence of C, should be taken into account. Some other polynomial approaches have also been employed (Connelly, 2001). The more complete bipolar models use the complete Shockley–Read–Hall (SRH) recombination expression for the nonradiative trap recombination and the electron and hole Auger recombination rates and calculate the spontaneous recombination rate from the complete energy band description by using the same approach used for the optical gain.

22.2.4.3 Internal Losses Coefficient

The internal absorption losses have a dramatic effect on the amplifier performance, and therefore the coefficient accounting for them, α_{in}, is an extremely important parameter for the simulation. It can be taken as a constant value to be estimated from experiments in broad area lasers, or it can be considered the result of two contributions: the scattering losses α_{scat}, and the losses due to different mechanism of photon absorption by free carriers, the so-called free-carrier absorption losses α_{fc}. In the approaches that solve the electrical equations in the vertical direction (Borruel et al., 2004; Grote et al., 2005), α_{fc} is locally calculated taking into account the carrier and photon distribution by

$$\alpha_{\mathrm{fc}}(x,z) = \int\limits_{L_y} \left[\kappa_n n(x,y,z) + \kappa_p p(x,y,z)\right] S(x,y,z)\mathrm{d}y, \tag{22.11}$$

where L_y is the thickness of the epilayer structure, and $\kappa_n(\kappa_p)$ is the electron (hole) free carrier absorption cross section. In most materials, the main contribution to α_{fc} arises from the hole term due to the mechanism of intervalence band absorption.

The most sophisticated amplifier models, especially those solving the complete electro-thermal-optical semiconductor equations, require many additional material parameters such as carrier mobilities, bandgap narrowing coefficients, energy band parameters, thermoelectric powers, thermal conductivities, etc.

22.3 Model Implementation

22.3.1 Overview of Solution Procedure

In this section, we describe the implementation of HAROLD 4.0, the model and code that we have developed for the simulation of tapered SOAs. Our approach is based on a previous model for tapered lasers (Borruel et al., 2004; Sujecki et al., 2003). The original algorithm has been modified slightly and a part of it has been rewritten in order to consider a single pass amplifier instead of a resonant cavity. HAROLD 4.0 solves self-consistently the steady-state partial differential equations describing the electrical, thermal, and optical behavior of a tapered SOA. The simulator includes a 3D electrical solver for the Poisson and continuity equations (Equations 22.2 through 22.4) coupled to a 3D thermal solver for the heat-flow equation (Equation 22.7). The thermal solver uses the local heat sources provided by the electrical solution. On the optical side, a wide-angle finite-difference beam propagation method (WA-FDBPM) (Hadley, 1992), under the effective index approximation, is used to propagate the optical field in the x–z plane. The optical solution is not fully 3D, as the vertical direction and the x–z plane are solved separately, and therefore the complete model is considered as quasi-3D.

Due to the large number of variables and iteration loops involved in the model, an important point is how to initialize the solution. In order to provide a good initial guess to the quasi-3D algorithm, a 1D laser simulator (Harold, 2001) is used for the initialization of the electrical, thermal, and optical variables. Although the 1D laser simulator was not designed for amplifiers, it provides useful initial guess values for the electrothermal variables as well as some inputs for the quasi-3D algorithm such as the applied voltage V_0 and the optical field modal profile in the vertical direction $f(y)$. The 1D simulator requires the electrical, thermal, and optical properties of all the device materials. The user introduces a target bias current and the 1D simulator solves the electrothermal and optical equations for an equivalent broad area laser having the same electrical injection area as the tapered SOA. The 1D simulator provides V_0, as well as an initialization for the vertical distribution of the electrostatic potential, the electron and hole concentrations, and the temperature. The power, wavelength, and lateral profile of the input optical field are provided by the user as inputs for the algorithm. The input optical field is symmetric and therefore only one half of the device is considered in the simulations.

The main flow of the quasi-3D algorithm is shown in Figure 22.2. It includes the following steps: (1) the 1D simulator provides V_0 and $f(y)$, and initializes all unknown 2D variables according to the initial 1D solution for the target current; (2) the initial lateral optical field profile at the first longitudinal slice ($z_i = 1$) is either set by the user or calculated as the fundamental mode of the waveguide in the lateral direction; (3) A 2D electrical solver is applied to the first slice in the $x - y$ plane at $z_i = 1$. It provides a 2D map of the electrical variables (n, p, and φ) and of the Joule (w_{joule}), the nonradiative (w_{nr}), and the free-carrier (w_{fc}) heat sources. It also calculates the complex effective refractive index along the lateral axis

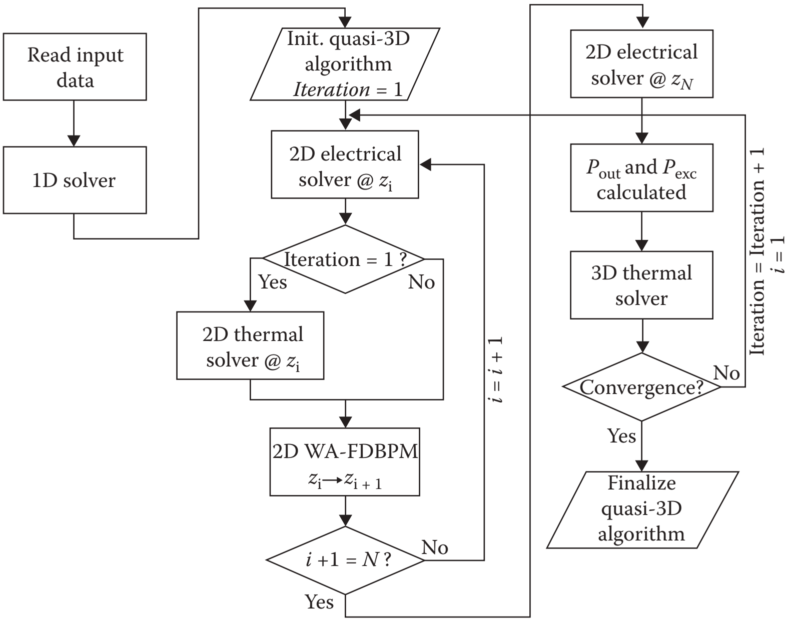

Read input data
1D solver
Init. quasi-3D algorithm
Iteration = 1
2D electrical solver @ z_i
Iteration = 1 ?
Yes
No
2D thermal solver @ z_i
2D WA-FDBPM
$z_i \rightarrow z_{i+1}$
$i+1 = N$?
No
Yes
$i = i + 1$
2D electrical solver @ z_N
P_{out} and P_{exc} calculated
3D thermal solver
Convergence?
No
Yes
Iteration = Iteration + 1
$i = 1$
Finalize quasi-3D algorithm

FIGURE 22.2 Main flow of the quasi-3D algorithm HAROLD 4.0 (see text for details).

at $z_i = 1$, thus providing the gain and refractive index variation due to the carriers. (4) The calculated heat sources are used as inputs for a 2D thermal solver applied to the x–y plane at $z_i = 1$. It should be noted that the thermal and electrical meshes are different and therefore some interpolation between the meshes is used. The 2D thermal solver provides a map of the temperature in the x–y plane and the thermally induced refractive index variation. This 2D thermal solver is only applied in the first iteration in order to initialize the thermal solution. (5) The WA-FDBPM is used to propagate the optical field from $z_i = 1$ to $z_i = 2$ using a fine mesh in the x–z plane. It uses the gain and refractive index changes given by the electrothermal solvers in the previous step and takes into account the geometry of the device. (6) Steps (3)–(5) are repeated until the algorithm arrives to the output facet at $z_i = N$. At this point, the output power P_{out} is known and the excess power P_{exc} can be calculated. Then, a 3D thermal solver is applied allowing heat flow between the z-slices. (7) The process described in (3), (5), and (6) is repeated until convergence is reached and a stable optical field, as well as stable temperature, carrier, and current distributions are found.

In the following subsections, the solvers for the electrical, thermal, and optical equations are described in more detail.

22.3.2 Solver for Electrical Equations

The electrical equations are solved in the volume defined by the total thickness of the epitaxial layers, the amplifier length, and the chip width. The input for the equations is taken from material databases according to the epilayer structure. Etched and implanted regions are simulated by assuming negligible carrier mobility. The electrical equations are solved together with two capture–escape continuity equations for electrons and holes at each QW, taking into account the nonequilibrium condition between the confined carriers in the QWs and the unconfined carriers in the barriers. Nonradiative (SRH and Auger) and spontaneous recombination terms are included along the complete device, in addition to the stimulated recombination in the active region. The model also includes free carrier absorption and band gap renormalization with standard dependencies and appropriate parameters.

The material gain and spontaneous recombination rate in the QWs are calculated using the expressions in Coldren and Corzine (1995), considering simple parabolic bands and Lorentzian broadening to decrease the computational effort. The parameters for the gain calculations (energy levels, effective masses, and intraband relaxation time) were previously fitted to yield similar results to those of a more sophisticated gain model that takes into account the valence band mixing. The local material gain (or absorption) and carrier-induced refractive index changes are calculated at the wavelength of the input power for each point of the mesh, taking into account the local temperature and carrier concentrations. They are subsequently used as inputs for the optical solver.

The equations are discretized using a finite difference approach. Nonuniform, separate meshes are taken for the vertical and lateral directions. The denser regions in the vertical and lateral directions correspond to the QW and to the border of the implanted/etched areas, respectively. The lateral mesh depends on the longitudinal position and typically contains around 40 mesh points. We chose 100 μm for the x–y slice separation in the longitudinal direction, after making sure that the results in our structures were similar when decreasing the longitudinal mesh size. The core of the numerical procedure is a Newton–Raphson algorithm, which solves the coupled Equations 22.2 through 22.4 in 1D along the vertical direction. The algorithm considers at each mesh point the influence of the next neighbors in the lateral and longitudinal directions. The 1D solver scans iteratively the lateral direction until convergence, providing a 2D map of the electrical variables at the z-slice i, z_i. The electrical variables are used to calculate the heat sources needed by the thermal solver, while the computed gain (or absorption) and the carrier-induced refractive index changes will be used by the optical solver to propagate the electric field to the slice z_{i+1}. In the resolution of the z_{i+1} slice, the values of the electrical variables previously obtained for z_i and z_{i+2} are considered in order to account for the longitudinal current flow. Finally, when convergence of the whole algorithm is reached, a complete 3D map of the electrical variables (electrostatic potential and electron and hole densities), consistent with the photon and temperature distributions, is produced.

22.3.3 Solver for Optical Equations

As mentioned above, the shape of the field at the amplifier input ($z_i = 1$) is either provided by the user or obtained by solving the Helmholtz equation in the lateral direction making use of the effective index approximation. Then, the WA-FDBPM is applied to propagate the field along a longitudinal optical mesh, which is much denser than the electrothermal mesh. The method uses the modal complex index along the lateral axis provided by the electrothermal solution of the initial slice. The WA-FDBPM uses the (1,1) Padé approximant, allowing the propagation of the field a few degrees around the longitudinal direction (1°–6°). The propagation continues until reaching the next slice of the electrothermal mesh, where the electrical and thermal solvers update the modal complex index. In this way, the tapered section is considered a succession of rectangular slices with increasing width.

To complete the optical solution, perfectly matched layers (PML) (Huang et al., 1996) are used as boundary conditions. The PML method defines a lossy region to which the Helmholtz equation is mapped by an anisotropic complex transformation. The PML attenuates the lateral traveling waves ensuring that the field at the borders of the device in the lateral direction vanishes. Due to the lateral symmetry of the optical field, a zero derivative is applied as the boundary condition at $x = 0$.

22.3.4 Solver for Thermal Equations

The heat flow equation (Equation 22.7) is solved in a larger region and using a different mesh than those of the electrical solver. In addition to the epilayer, the substrate, the metal layers, and the heatsink are also considered in the thermal solution. The boundary conditions for the heat-flow equation in the case of a p-down mounted device are negligible heat flow at the n-metal and lateral chip external interfaces and uniform temperature at the bottom of the heatsink. Etched regions are thermally simulated by introducing the thermal conductivity of the planarization isolator.

As mentioned in Section 22.3.1, during the first iteration of the Quasi-3D algorithm, the 2D heat flow equation is solved at each x–y slice with the local heat sources obtained from the electrical solver. The 2D heat flow equation is discretized with a finite difference method, and the system of equations is solved by using a stabilized biconjugate gradient method. The calculated temperature distribution replaces the initialization values provided by the 1D laser simulator. Once the electro-optical solution process has reached, the output facet and the output power P_{out} has been calculated, the excess power P_{exc} is calculated using Equation 22.8, and the excess power heat source is distributed along the complete device volume. Then, a 3D thermal solver is applied. The 3D thermal solver is based on scanning the 2D solver along the longitudinal direction allowing the longitudinal heat flow between slices until reaching convergence. Successive iterations of the complete algorithm use this 3D thermal solver as indicated in Figure 22.2.

22.4 Simulation Example: 1.5-μm InGaAsP/InP Tapered Amplifier

As an illustrative example, in this section we analyze a typical 1.5-μm InGaAsP/InP tapered amplifier with the HAROLD 4.0 simulation tool. After a summary of the most relevant material, and geometrical and simulation parameters, we present the simulation results of a reference device and based on these results, we illustrate the potential of the simulation tools by analyzing the effect on the device performance of two important design parameters: the taper angle α and the confinement factor Γ. The analysis is presented here only for the sake of illustration, so we will present an overview of some interesting effects without a detailed analysis.

22.4.1 Device Geometry and Simulation Parameters

As a reference device, we have considered an SOA with an InGaAsP/InP multi-QW epitaxial structure similar to that of the MOPA reported in Faugeron et al. (2015). It features an asymmetric cladding structure especially conceived for shifting the broad vertical profile of the optical mode to the n-side of the diode, thus minimizing the overlapping with the highly absorbing p-doped layers. Table 22.1 provides the geometrical parameters of the device as well as a brief summary of the most influential material and device parameters used in the simulation. The taper angle has been selected so as to fit the calculated free diffraction angle assuming an index step $\Delta n_{\text{eff}} = 7 \times 10^{-3}$. We have considered a p-side down mounting configuration as more appropriate for the heat management required in a power device. For the Auger recombination

TABLE 22.1 Geometrical Parameters of the Reference SOA and Summary of the Most Relevant Material and Device Parameters Used in the Simulation

Symbol	Parameter	Value	Units
L_{RW}	Length of the ridge waveguide (RW) section	1	mm
W_{RW}	Width of the RW section	3.5	μm
L_{Tap}	Length of the taper section	2	mm
α_{tap}	Taper angle	7	°
Δn_{eff}	Effective index step of the RW section	7×10^{-3}	–
	Mounting configuration (p-up/p-down)	p-Down	–
T_{HS}	Heatsink temperature	18	°C
P_{in}	Input power	20	mW
Γ	Confinement factor	0.037	
α_{scat}	Scattering losses coefficient	0.5	cm^{-1}
$C_n(C_p)$	Electron (hole) Auger recombination coefficient	$8(8) \times 10^{-29}$	cm^6s^{-1}
k_e (k_h)	Electron (hole) free carrier absorption coefficient	$1(30) \times 10^{-18}$	cm^2
n_I	Differential refractive index coefficient	10^{-1}	$\text{cm}^{3/2}$

coefficients and the electron and hole free carrier absorption coefficients, we have used judicious values in the range of published values (Joindot and Beylat, 1993; Piprek et al., 2000). We have accounted for the dependence of the refractive index on the carrier density by means of a square root function with a coefficient n_1 (Borruel et al., 2004) such that it results in a value of about 5 for the LEF at the operating wavelength and carrier densities. The value used for the internal scattering losses together with the free-carrier absorption losses calculated using Equation 22.9 yield a total internal loss of approximately 8 cm^{-1} for a broad area laser with the same epitaxial structure.

22.4.2 Analysis of the Reference SOA

Figure 22.3 shows the simulated power–current (P–I) characteristics of the reference SOA together with the evolution of the electrical-to-optical conversion efficiency ($\eta_{\text{E-O}}$). For injection currents above a transparency level ($I \sim 1.3$ A) the device actually amplifies with an increasing conversion efficiency. The output power shows an approximately linear increase in the current range $4 < I < 9.5$ A. For higher injection levels, both the output power and the conversion efficiency start a saturation process.

The simulation provides the means to make a more in-depth analysis revealing details of the behavior of relevant variables inside the SOA. As an example, Figure 22.4 shows 3D plots of several magnitudes in the QW region for a relatively low injection current ($I = 4$ A). In the RW section, due to the index guiding, the profile of the photon density (Figure 22.4a) is narrow and stable, and shows an increase of the maximum along the longitudinal direction revealing the optical gain. The maximum drops and the profile broadens when the tapered region starts, and from there up to the output facet, the photon profile evolves smoothly. In this evolution, the maximum photon density increases and so does the profile width, following the evolution of the width of the tapered section. For this relatively low injection level, the electron density profile (Figure 22.4b) is almost flat inside the tapered section. In the RW section, the electron density profile is broader than the width of the section due to the current spreading (not shown). The maximum electron density in this section is lower than the density in the tapered section and decreases slightly along the longitudinal direction. Both effects are a consequence of a higher stimulated recombination ratio in the RW section. The hole density (not shown) is slightly different but shows a similar profile. The mutual dependence between carrier and photon densities is much more pronounced at high-injection levels as explained below. The temperature profile in the lateral direction (Figure 22.4c) reminds the injection profile showing higher values in the injected region and a fast variation at the edge of the region. In the RW section, the lateral profile shows a small dip at the cavity center, whereas in the tapered section, the temperature decreases smoothly from a maximum at the axis to a minimum at the edges of the injected region. In the noninjected regions, the lateral evolution is smooth and shows a slow decrease toward the edges of the chip. In the longitudinal direction, the temperature increases smoothly from the RW section to the output facet. This increase is slow in the RW section and faster in the taper section (see also Figure 22.5b). At

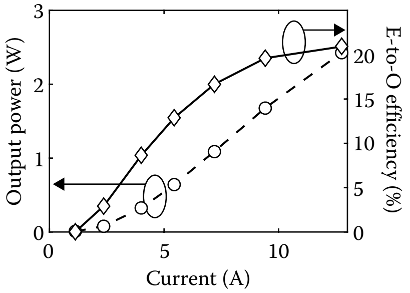


FIGURE 22.3 Power–current characteristics (circles) and electrical-to-optical conversion efficiency (diamonds) of the reference SOA.

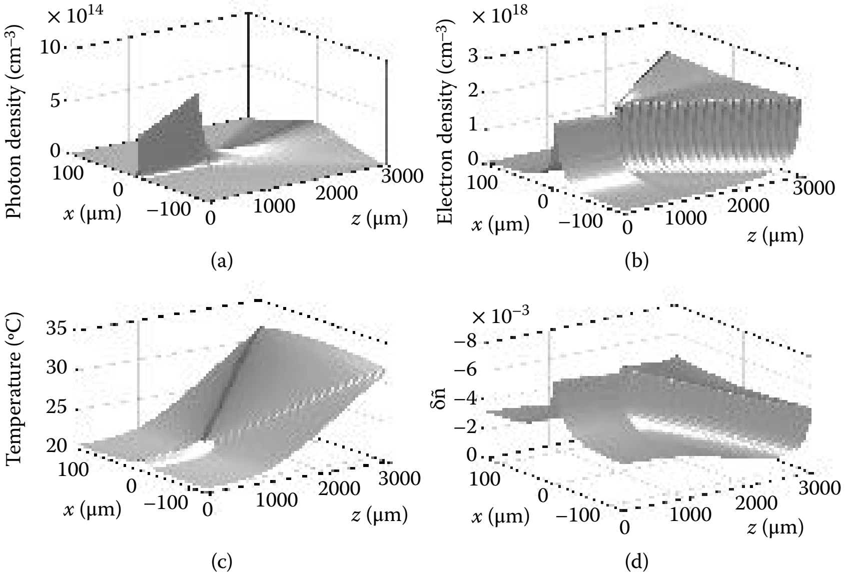


FIGURE 22.4 3D maps of the photon (a) and the averaged electron densities (b), the temperature (c), and the carrier and thermally induced refractive index change (d) in the active region of the reference SOA for an injection $I = 4$ A. The little "ripples" appearing at the edges of the tapered section in some plots are artifacts due to the discretization in the longitudinal direction.

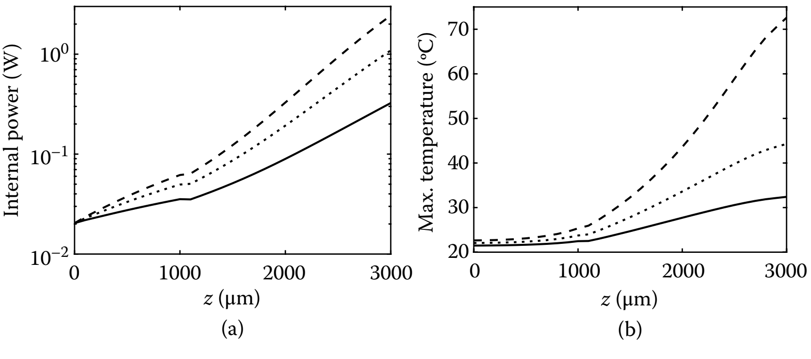


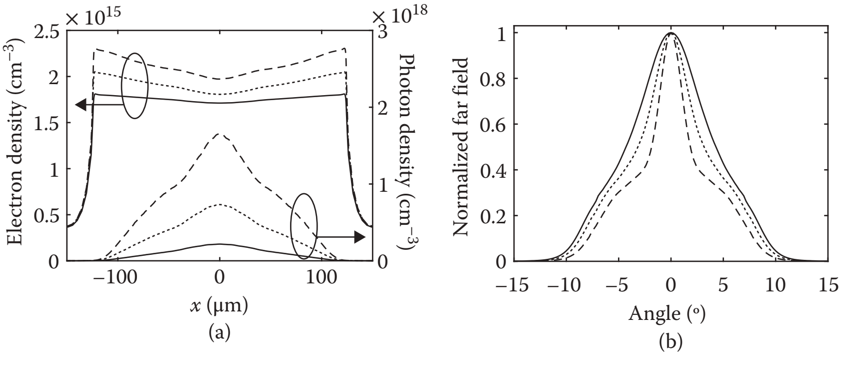
FIGURE 22.5 Evolution along the longitudinal direction of the internal optical power (a) and the maximum QW temperature (b), for different current values. $I = 4$ A (solid), $I = 7.2$ A (dotted), and $I = 12.7$ A (dashed).

this relatively low-injection level, the evolution of the longitudinal temperature profiles are similar in the injected and in the noninjected regions. Finally, Figure 22.4d shows a 3D map of the QW refractive index variation given by Equation 22.9. This variation is a balance between the positive thermal contribution and the negative contribution of the excess carrier density. Therefore the profile is similar to the electron density profile (Figure 22.4b) smoothly modified by the temperature profile (Figure 22.4c). This index variation is the way by which the density of carriers and the local temperature influence the photon density and, hence, the beam profile.

More insight into the evolution of the optical power along the SOA can be provided by plots such as the one in Figure 22.5a, where the internal optical power at each x–y slice is plotted along the longitudinal direction in a logarithmic scale. The evolution of the internal optical power can be described in terms of an effective optical gain g_{eff} by $P(z + \Delta z) = P(z)e^{g_{\text{eff}}\Delta z}$ and therefore g_{eff} is proportional to the slope of the curves in Figure 22.5a. In this description, g_{eff} corresponds to the difference between the modal gain and all the losses occurring in the beam propagation. For the low-injection level considered previously (4 A, solid line), the effective gain is relatively low and slightly decreasing in the RW section and much higher in the tapered section. This can be explained from the maps of photon and carrier densities in Figure 22.4a and b: for approximately the same carrier density in both sections, the photon density is much higher in the RW section and therefore the gain shows an increasing saturation effect which is absent in the tapered section where the photon density is much lower. In a transition region, just at the beginning of the tapered section, there is a dramatic drop of the gain corresponding to an increase of the losses due to the change in the guiding mechanism from index guiding to gain guiding. This drop is due to the poor overlapping of the mode and the injected region at the beginning of the tapered section since the external part of the mode expands to unpumped and, therefore, absorbing regions. In this device, in which the taper angle (and therefore the angle of the gain region) almost matches the free diffraction angle, the gain recovers soon after the transition. For higher-injection levels, Figure 22.5a shows a qualitatively similar behavior of the gain. In both sections, the gain is higher as corresponding to a higher carrier density and evidences some saturation at the end of each section corresponding to a high photon density in these regions.

To complete the overview picture of the longitudinal evolution of internal variables, Figure 22.5b shows the evolution of the maximum QW temperature for the same low-, intermediate-, and high-injection levels of Figure 22.5a. As expected, the QW temperature evolves similarly for the three injection levels, showing a faster increase in the taper section for the highest injection level.

More local details of the behavior of the photon and carrier densities when the injection increases are provided by Figure 22.6. Figure 22.6a shows the electron and photon density profiles along the lateral direction at the output facet of the device. The mutual dependence between carrier and photons pointed above for low injection is much more apparent at high-injection levels and can be explained as follows: The density of carriers injected at these levels is high on average. In comparison with the outer regions, the number of photons generated by stimulated recombination around the axis ($x = 0$) (where the optical mode reaches the maximum) is higher, and therefore the carrier density locally decreases. This is the well-known spatial hole burning (SHB) effect. Due to its negative dependence on the carrier density, the refractive index locally increases in the places more affected by the SHB effect, giving rise to a more abrupt index profile. In turn, the new index profile contributes to a higher confinement of the optical mode in the axial region in a feedback loop in which the local temperature also plays a role through the thermal dependence of the refractive index. As a consequence, the smooth profile of the optical beam at low injection undergoes a slight degradation when the injection increases as illustrated in Figure 22.6b, where the FF profiles at the three injection levels have been compared.



(a)

(b)

FIGURE 22.6 (a) Photon and electron density lateral profiles at the output facet for three injection levels. (b) The corresponding far field profiles. $I = 4$ A (solid), $I = 7.2$ A (dotted), and $I = 12.7$ A (dashed).

22.4.3 Effect of the Confinement Factor and Taper Angle

The confinement factor and the taper angle are two critical design parameters affecting the performance of a tapered SOA. In order to illustrate the capabilities of the simulation tools for the analysis of the influence of these parameters, we have compared the reference SOA ($\Gamma = 3.7\%$ and $\alpha = 7°$) with two devices: (1) an SOA which epitaxial design has been carefully modified so as to result in a confinement factor $\Gamma = 5.6\%$ without increasing the mode overlapping with the p-doped cladding (high Γ-SOA) and (2) a device with $\alpha = 3°$ and the same RW and tapered section lengths as the reference SOA (3°-SOA). Figure 22.7 shows a comparison of the *P–I* characteristics, and the conversion efficiency of the three devices. In comparison with the reference SOA, the high Γ-SOA has a lower transparency current, a higher slope of the *P–I* characteristics, and a higher conversion efficiency. Therefore, at the highest current its output power and conversion efficiency are significantly higher. This can be qualitatively understood just taking into account the higher mode overlapping with the active region. The 3°-SOA shows also a lower transparency current, a higher slope of the *P–I* characteristics, and a higher conversion efficiency. In this case, the explanation is the higher current density corresponding to the same current as a consequence of the lower area of the device.

Figure 22.8 shows the effective optical gain in the three devices at the highest injection level of around 13 A. Note that the output power of each device is different (see Figure 22.7a). In the RW section, the effective gain of the high Γ-SOA is only slightly higher than the effective gain of the reference SOA in spite of the higher value of Γ. The reason for this is the material gain saturation due to the high photon density in this section that makes the gain almost independent of Γ. The effective gain of the 3°-SOA in this RW section reaches higher values corresponding to the higher carrier injection. In the tapered section, the gain of the high Γ-SOA is clearly superior to that of the reference SOA as can be expected from its higher value of Γ. The difference is more apparent at the beginning of the section, where the photon densities are low, than at the end, where the photon densities (and hence the saturation) become high. The most revealing difference between the gain of the 3°-SOA and the other devices is in the transition region at the beginning of the taper section where the mismatch between the free diffraction angle and the taper angle in the 3°-SOA results in high losses evidenced as a low effective gain in this region.

The improved performance of the high Γ-SOA and the 3°-SOA in terms of output power and conversion efficiency is, however, at the expense of a much faster beam degradation as revealed by the evolution of the FF width and the beam propagation factor (M^2) plotted in Figure 22.9 (see [Esquivias et al., 2017] Chapter 28, Volume 2 of this book). In the case of the high Γ-SOA, the beam quality as revealed by M^2 progressively worsens, reaching high values at relatively low currents. In the case of the 3°-SOA, the degradation takes place at intermediate values of the current whereas the reference SOA keeps a relatively low value of M^2 up to a high-injection level.

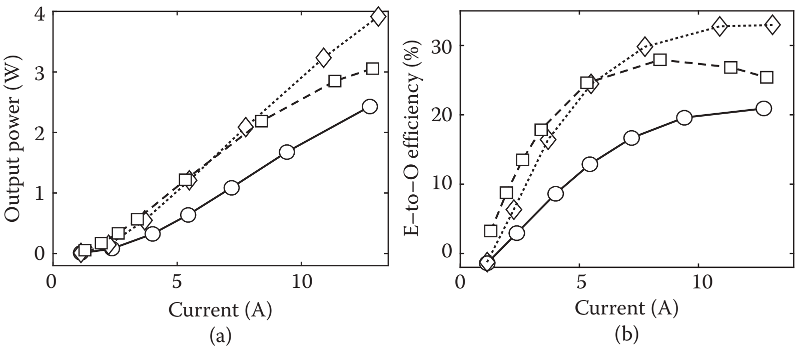


FIGURE 22.7 (a) *P–I* characteristics of the reference SOA (circles), the high Γ-SOA (diamonds), and the 3°-SOA (squares). (b) Their corresponding electrical-to-optical conversion efficiencies.

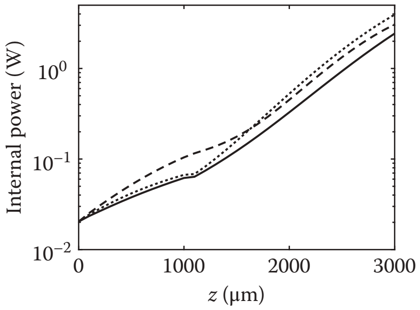
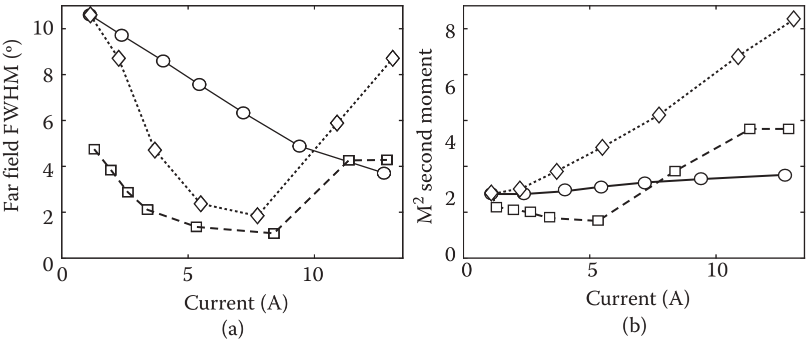


FIGURE 22.8 Evolution along the longitudinal direction of the internal optical power of the three devices at the highest injection ($I = 13$ A), showing the different effective gain (slope of the curves). Reference SOA (solid), high Γ-SOA (dotted), and 3°-SOA (dashed).



(a)

(b)

FIGURE 22.9 (a) Far field full width at half-maximum (FWHM) and (b) M^2 dependence on current for the three devices considered. Reference SOA (circles), high Γ-SOA (diamonds), and 3°-SOA (squares).

As mentioned above, more detailed information of the mechanism underlying beam degradation is brought by the plots in Figure 22.10a where the electron and photon density profiles at the output facet of the three devices are plotted for the highest injection level. The mutual dependence between carriers and photons described in Section 22.4.2 for the reference SOA is much more apparent for the high Γ- and the 3°-SOAs and results in deep holes in the axial region of the electron density profiles together with the corresponding narrow peaks in the photon density. As a consequence, the beam quality severely degrades, as revealed, for example, by the FF profiles plotted in Figure 22.10c and d in comparison with the profile of the reference SOA (Figure 22.10b).

The self-focusing is another manifestation of the beam degradation due to the interplay between carriers and photons. Figure 22.11 shows a comparison between the evolution of the beam profiles in the reference (Figure 22.11a and b) and the 3°-SOA (Figure 22.11c and d) for low- (Figure 22.11a and c) and high injection conditions (Figure 22.11b and d). In these plots, the photon density in each slice perpendicular to the longitudinal axis has been normalized to its maximum value in order to visualize the focusing. The white dashed lines show the borders of the injected region. In contrast with the virtually absent focusing of the beam in the reference SOA, Figure 22.11d shows a strong self-focusing in the 3°-SOA at high current. As a consequence, the slightly increasing astigmatism in the reference SOA when the injection increases (not shown) becomes abrupt and then drops down to negative values in the 3°-SOA.

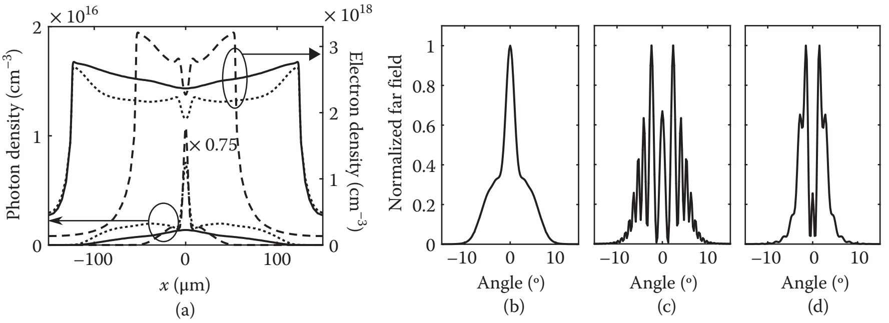


FIGURE 22.10 (a) Photon and electron density lateral profiles at the output facet at the highest injection level ($I = 13$A) for the three devices considered. Reference SOA (solid), high Γ-SOA (dotted), and 3°-SOA (dashed). For clarity, the photon density values of the 3°-SOA device (dashed) have been multiplied by 0.75. (b) through (d) Normalized far field profiles for the reference SOA, the high Γ-SOA, and the 3°-SOA, respectively.

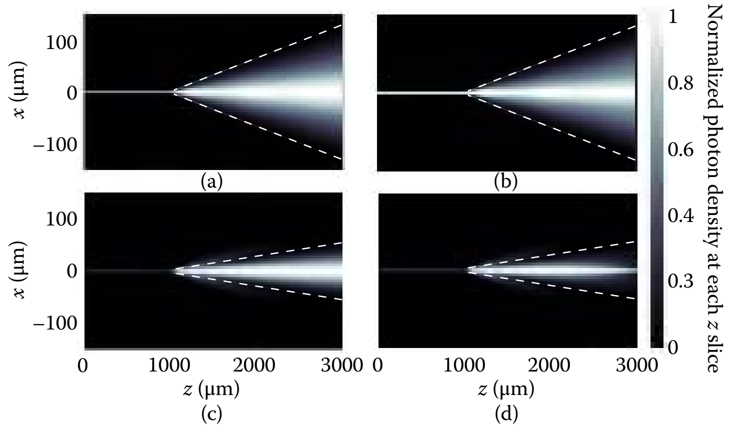


FIGURE 22.11 Maps of the normalized photon density of the reference SOA (top) and the 3°-SOA (bottom) under low-injection (left) and high-injection conditions (right). The white dashed lines show the borders of the injected region. For each value of the longitudinal coordinate z, the photon density was normalized to its maximum along the corresponding lateral x-axis.

22.5 Summary

We have presented an overview of the current state of the art in the modeling of high-power tapered SOAs. We have described the most relevant models used for the simulation of carrier transport, optical field, material properties, and temperature distribution in this type of devices. The particular geometry of tapered amplifiers makes difficult the reduction of the dimensionality of the problem, requiring relatively complex numerical models. The implementation of a steady-state quasi-3D model has been explained in detail, with emphasis on the interaction between the electrical, thermal, and optical solvers. For illustration, the model has been applied to a 1.5-μm tapered SOA, showing the relation between internal carrier and photon distributions and the external measurable characteristics. The influence of the optical confinement factor and the taper angle in the output power and beam quality has also been analyzed.

The main limitation of the steady-state single-frequency models is that they cannot be used to analyze the noise properties of the SOAs, as well as other complex nonlinear phenomena as cross-talk in wavelength division multiplexing applications. Two- or quasi-3D TWMs, coupled to bipolar carrier and thermal equations are the natural evolution of steady-state models to account for dynamical and spectral issues.

Acknowledgments

This work was supported by the European Commission through the FP7-Space project BRITESPACE under grant agreement no. 313200, by Ministerio de Economía y Competitividad of Spain under projects RANGER (TEC2012-38864-C03-02) and COMBINA (TEC2015-65212-C3-2-P), and by the Comunidad de Madrid under program SINFOTON-CM (S2013/MIT-2790). A. Pérez-Serrano acknowledges support from Ayudas a la Formación Posdoctoral 2013 program (FPDI-2013-15740). The authors acknowledge the contribution of L. Borruel, H. Odriozola, S. Sujecki, and E. C. Larkins to the development of the initial model and code for the simulation of tapered lasers.

References

Balsamo, S., Sartori, F., and Montrosset, I. 1996. Dynamic beam propagation method for flared semiconductor power amplifiers. *IEEE J. Quantum Electron.* 2(2):378–384.

Bandelow, U., Gajewski, H., and Hünlich, R. 2005. Fabry–Perot lasers: Thermodynamics-based modeling. In *Optoelectronic Devices. Advanced Simulation and Analysis*. Ed. J. Piprek. 63–85. New York, NY: Springer.

Borruel, L., Arias, J., Romero, B. et al. 2003. Incorporation of carrier capture and escape processes into a self-consistent cw model for Quantum Well lasers. *Microelectron. J.* 34:675–677.

Borruel, L., Odriozola, H., Tijero, J. M. G. et al. 2008. Design strategies to increase the brightness of gain guided tapered lasers. *Opt. Quantum Electron.* 40(2–4):175–189.

Borruel, L., Sujecki, S., Esquivias, I. et al. 2002. Self-consistent electrical, thermal, and optical model of high-brightness tapered lasers. *Proc. SPIE* 4646:355–366.

Borruel, L., Sujecki, S., Moreno, P. et al. 2004. Quasi-3-D simulation of high-brightness tapered lasers. *IEEE J. Quantum Electron.* 40:463–472.

Buus, J. 1982. The effective index method and its application to semiconductor lasers. *IEEE J. Quantum Electron.* 18(7):1083–1089.

Coldren, L. A. and Corzine, S. W. 1995. *Diode Lasers and Photonic Integrated Circuits*, New York, NY: John Wiley & Sons.

Connelly, M. J. 2001. Wideband semiconductor optical amplifier steady-state numerical model. *IEEE J. Quantum Electron.* 37(3):439–447.

Esquivias, I., Pérez-Serrano, A., and Tijero, J. M. G. 2017. High brightness tapered lasers. In: *Handbook of Optoelectronic Device Modeling and Simulation*, vol 2, (Piprek, J., ed) pp.59–80. Boca Raton, FL: Taylor & Francis.

Faugeron, M., Vilera, M., Krakowski, M. et al. 2015. High power three-section integrated master oscillator power amplifier at 1.5 µm. *IEEE Photon. Technol. Lett.* 27:1449–1452.

Fischer, I., Hess, O., Elsäßer, W. et al. 1996. Complex spatio-temporal dynamics in the near-field of a broad-area semiconductor laser. *Europhys. Lett.* 35(8):579–584.

Gehrig, E., and Hess, O. 2001. Spatio-temporal dynamics of light amplification and amplified spontaneous emission in high-power tapered semiconductor laser amplifiers. *IEEE J. Quantum Electron.* 37(10):1345–1355.

Grote, B., Heller, E. K., Scarmozzino, R. et al. 2005. Fabry–Perot lasers: Temperature and many-body effects. In *Optoelectronic Devices. Advanced Simulation and Analysis*. Ed. J. Piprek. 63–85. New York, NY: Springer.

Hadley, G. R. 1992. Wide-angle beam propagation using Padé approximant operators. *Opt. Lett.* 17(20):1426–1428.

Harold™ 2001. *3.0 Reference Manual.* Oxford: Photon Design.

Huang, W. P., Xu, C. L., Lui, W. et al. 1996. The perfectly matched layer (PML) boundary condition for the beam propagation method. *IEEE Photon. Technol. Lett.* 8(5):649–651.

Javaloyes, J. and Balle, S. 2010. Quasiequilibrium time-domain susceptibility of semiconductor quantum wells. *Phys. Rev. A* 81:062505.

Joindot, I. and Beylat, J. L. 1993. Intervalence band absorption coefficient measurements in bulk layer, strained and unstrained multiquantum well 1.55 µm semiconductor lasers. *Electron. Lett.* 29: 604–606.

Koch, S. W., Hader, J., Thränhardt, A. et al. 2005. Gain and absorption: Many-body effects. In *Optoelectronic Devices. Advanced Simulation and Analysis*. Ed. J. Piprek. 63–85. New York, NY: Springer.

Lai, J.-W. and Lin, C.-F. 1998. Carrier diffusion effect in tapered semiconductor-laser amplifier. *IEEE J. Quantum Electron.* 34(7):1247–1256.

Lang, R. J., Hardy, A., Parke, R. et al. 1993. Numerical analysis of flared semiconductor laser amplifiers. *IEEE J. Quantum Electron.* 29:2044–2051.

Lugiato, L. A. and Narducci, L. M. 1985. Single-mode and multimode instabilities in lasers and related optical systems. *Phys. Rev. A* 32(3):1576–1587.

Mariojouls, S., Margott, S., Schmitt, A. et al. 2000. Modeling of the performance of high-brightness tapered lasers. *Proc. SPIE* 3944:395–406.

Moloney, J. V., Indik, R. A., and Ning, C. Z. 1997. Full space-time simulation of high-brightness semiconductor lasers. *IEEE Photon. Technol. Lett.* 9:731–733.

Nagarajan, R., Ishikawa, M., Fukushima, T. et al. 1992. High speed quantum-well lasers and carrier transport effects. *IEEE J. Quantum Electron.* 28:1990–2008.

O'Brien, S., Welch, D. F., Parke, R. A. et al. 1993. Operating characteristics of a high-power monolithically integrated flared amplifier master-oscillator power-amplifier. *IEEE J. Quantum Electron.* 29:2052–2057.

Pérez-Serrano, A., Javaloyes, J., and Balle, S. 2013. Spectral delay algebraic equation approach to broad area laser diodes. *IEEE J. Sel. Top. Quant. Electron.* 19(5):1502808.

Piprek, J., Abraham, P., and Bowers, J. E. 2000. Self-consistent analysis of high-temperature effects on strained-layer multiquantum-well InGaAsP–InP lasers. *IEEE J. Quantum Electron.* 36:366–374.

Razaghi, M., Ahmadi, A., and Connelly, M. J. 2009. Comprehensive finite-difference time-dependent beam propagation model of counter propagating picosecond pulses in a semiconductor optical amplifier. *IEEE /OSA J. Lightwave Technol.* 27(15):3162-3174.

Schwertfeger, S., Klehr, A., Hoffmann, T. et al. 2011. Picosecond pulses with 50 W peak power and reduced ASE background from an all-semiconductor MOPA system. *Appl. Phys. B* 103:603–607.

Selberherr, S. 1984. *Analysis and Simulation of Semiconductor Devices*, Vienna: Springer-Verlag.

Spreemann, M., Lichtner, M., Radziunas, M. et al. 2009. Measurement and simulation of distributed-feedback tapered master-oscillator power-amplifiers. *IEEE J. Quant. Electron.* 45:609–616.

Sujecki, S., Borruel, L., Wykes, J. et al. 2003. Nonlinear properties of tapered laser cavities. *IEEE J. Sel. Top. Quant. Electron.* 9:823–834.

Sumpf, B., Hasler, K.-H., Adamiec, P. et al. 2009. High-brightness quantum well tapered lasers. *IEEE J. Sel. Top. Quant. Electron.* 15(3):1009–1020.

Tijero, J. M. G., Borruel, L., Vilera, M. et al. 2015. Analysis of the performance of tapered semiconductor optical amplifiers: Role of the taper angle. *Opt. Quant. Electron.* 47(6):1437–1442.

Van Roey, J., van der Donk, J., and Lagasse, P. E. 1981. Beam-propagation method: Analysis and assessment. *J. Opt. Soc. Am.* 71:803–810.

Vilera, M., Pérez-Serrano, A., Tijero, J. M. G. et al. 2015. Emission characteristics of a 1.5 µm all semiconductor tapered master oscillator power amplifier. *IEEE Photon. J.* 7(2):1500709.

Wachutka, G. K. 1990. Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling. *IEEE Trans. Comput. Aided Des. Integr. Circuits Syst.* 9:1141–1149.

Walpole, J. N. 1996. Semiconductor amplifiers and lasers with tapered gain regions. *Opt. Quant. Electron.* 28:623–645.

Wenzel, H., Sumpf, B., and Erbert, G. 2003. High-brightness diode lasers. *C. R. Phys.* 4:649–661.

Williams, K. A., Penty, R. V., White, I. H. et al. 1999. Design of high-brightness tapered laser arrays. *IEEE J. Sel. Top. Quant. Electron.* 5:822–831.

23

Quantum-Dot Semiconductor Optical Amplifiers

Benjamin Lingnau
and
Kathy Lüdge

23.1 Introduction.. 715
23.2 QD Semiconductor Optical Amplifier Model........................... 716
Electric Field Propagation: Delay-Differential-Equation Model • Quantum-Dot-in-a-Well Material Equations • QD Charge-Carrier Scattering • Fit Functions for Scattering Rates • Modeling of Spontaneous Emission
23.3 Application Examples... 731
Static Characterization of the QDSOA • Gain Recovery Dynamics in QDSOAs • Rabi Oscillations in QD Semiconductor Amplifiers
23.4 Conclusion... 741

23.1 Introduction

The rapidly increasing need for telecommunications and data-streaming applications within our society still demands a deeper understanding of the physical processes behind this emerging technology. In the last decades, there have been many attempts to optimize the performance of optoelectronic devices by using innovative nanostructured semiconductor gain materials [1,2]. A lot of these innovative devices already have found their way into photonic applications [3,4]. In particular, the demand for faster and more energy-efficient data transfer set the trend to replace the well-established cable-based data transmission with energy efficient optical technologies [5]. As one example, optical amplifiers are needed in optical networks to raise the signal power level and compensate for the inevitable optical losses in glass–fiber connections. Semiconductor-based optical amplifiers (SOAs) are cheap, relatively easy to fabricate, and consume only little energy within an electric circuit. They are structurally similar to semiconductor laser devices, with the difference lying in the absence of an optical cavity (see Chapter 20 SOA Fundamentals).

In this chapter, we describe the modeling of optical amplifiers that contain semiconductor quantum dots (QDs) as active media. Those nearly zero-dimensional QD structures gained a lot of attention as the material of choice for highly energy-efficient and small-footprint optoelectronic devices [6]. The scattering mechanisms and the unique electronic structure of semiconductor QDs have been found to make such devices prime candidates for the implementation of next-generation optoelectronic applications and novel high-speed data transmission schemes. QDs are the final step in miniaturization of the optically active gain material and thus in the confinement of charge carriers below the de Broglie wavelength of electrons (a few nanometers). In addition to the modeling of the quantum-dot semiconductor optical amplifiers (QDSOAs) as active media described within this chapter, other QD-based optical devices such as QD laser and QD light-emitting diodes are treated separately in a different section of this book. Nevertheless, our microscopically based material model for the QDs can equally be used as the building block in laser devices [7,8].

Our main focus will be the characterization of both the performance and the special nonlinear features of QDSOAs that are induced by the low dimensional and thus highly confined QD gain medium. While linear amplification with a low noise figure is required, to ensure low distortion of the input signal for data transmission [9–11], nonlinear optical applications, such as four-wave-mixing [12–15] and cross-gain modulation [16–18] for wavelength conversion, as well as regenerative amplification, require a nonlinear response of the optical amplifier. QDSOAs generally have a high gain bandwidth due to the growth-induced inhomogeneity in the dot size and material composition. These fluctuations lead to an inhomogeneous broadening of the localized QDs states and consequently allow for a broadband amplification. Additionally, due to the coupling to a charge-carrier reservoir with charge-carrier scattering times in the picosecond range, ultrafast gain recovery [19–24] and nonlinear signal processing are possible [15,16,25]. The comparably slow dephasing time of the microscopic interband polarization in the localized QD states [26–28] allows the possibility to directly observe quantum-mechanical effects, such as Rabi oscillations [29–33] or self-induced transparency [34,35]. This could potentially open up new applications in the signal processing of ultra-short, ultra-strong optical pulses.

The optical amplifier devices we consider in this work are single-pass devices, where the optical signal coupled into one side of the waveguide structure is ideally passing exactly once across the device. Thus, the edge-emitting QDSOAs are longer than typical laser devices, in order to provide a long enough interaction time between optical signal and the active medium. The model used for the numeric simulations is based on a traveling-wave approach for the pulse propagation, using Maxwell's wave equation in the slowly varying amplitude and rotating wave approximation. The polarization of the active medium is modeled using microscopic considerations for the charge-carrier dynamics, i.e., the light–matter interaction is modeled on the basis of Maxwell–Bloch equations, taking into account the microscopic polarization of the QD medium.

The structure of the chapter is as follows. First, a delay-differential equation model for an efficient and accurate description of the electric field propagation through the QDSOA is presented (Section 23.2). Subsequently, the charge-carrier dynamics is described in the framework of microscopically calculated nonlinear charge-carrier scattering rates before our approach for a quantitative modeling of the amplified spontaneous emission (ASE) inside the QDSOA is discussed. In Section 23.3, we present exemplary results on the performance of the QDSOA, focusing on the unique properties of semiconductor QDs as active medium. After a static characterization of a chosen sample device, we discuss the ultrafast gain dynamics of QDSOAs, identifying the main timescales that determine its gain recovery. We then present results on the coherent propagation of strong optical pulses. The long dephasing time of QDs leads to strong signatures of the coherent interaction, leading to modifications in the pulse shape and the gain dynamics due to the appearance of Rabi oscillations. A brief conclusion is provided.

23.2 QD Semiconductor Optical Amplifier Model

QDSOAs differ from conventional devices in the choice of active medium. The charge-carrier dynamics in QDs can be very complex due to the localization of electrons within the QDs embedded in the surrounding quantum-well or bulk material. In this chapter, we, therefore, aim to describe the charge-carrier dynamics and derive a set of coupled differential equations for the electronic states involved. The strong nonlinearity of the resulting equations leads to important differences in the performance of QDSOAs when compared to conventional devices. It also opens up the possibility of novel applications, based on their ultrafast gain recovery for linear applications, or their broad gain spectrum for nonlinear or wavelength conversion applications.

In this section, we derive a QDSOA model, in which we focus on the unique dynamics of the QD gain medium. The wave equation used for the propagation of the light will be similar to conventional SOA models (see Chapter 20 on SOA fundamentals in this book). In other works, models without spatial resolution have been employed for the description of semiconductor amplifiers [17,36–38]. As soon as strong spatial inhomogeneities arise, such models are, however, bound to fail. On the other hand, numerically solving the partial differential equation for the electric field propagation along the waveguide axis is computationally very expensive[39]. We, therefore, develop a delay-differential equation model that combines spatial resolution and reasonable computation time. The light–matter interaction will be described within a Maxwell–Bloch framework, which couples the light propagation dynamics to the QD active medium.

In the following, we derive the electric field propagation equations and the material equations describing the active medium. Furthermore, we derive Boltzmann-like equations that govern the charge-carrier scattering dynamics within a quantum-dot-in-a-well system. Spontaneous emission will be included in a phenomenological way to yield stochastic differential equations.

23.2.1 Electric Field Propagation: Delay-Differential-Equation Model

In order to calculate the dynamics of the QDSOA device, we must solve the wave equation of the optical field inside the waveguide. Here we concentrate on narrow-area ridge waveguide structures where the propagation along the longitudinal axis (labeled z) dominates, and the transverse mode profile is assumed to be constant. The electric field is thus governed by the one-dimensional wave equation

$$\frac{\partial^2}{\partial t^2}\boldsymbol{\mathcal{E}}(\boldsymbol{r},t) - {c_0}^2\frac{\partial^2}{\partial z^2}\boldsymbol{\mathcal{E}}(\boldsymbol{r},t) = -\frac{1}{\varepsilon_0}\frac{\partial^2}{\partial t^2}\boldsymbol{\mathcal{P}}(\boldsymbol{r},t)\,, \tag{23.1}$$

where $\boldsymbol{\mathcal{E}}$ and $\boldsymbol{\mathcal{P}}$ denote the real electric field and polarization, respectively, and c_0 and ε_0 are the vacuum speed of light and dielectric constant, respectively. The field quantities are expanded in terms of plane waves,

$$\mathcal{E}(z,t) = \frac{1}{2}\left[E^+(z,t)e^{ikz} + E^-(z,t)e^{-ikz}\right]e^{-i\omega t} + \text{c.c.}\,, \tag{23.2}$$

where we have introduced the slowly varying field amplitudes $E^\pm$, describing the forward (+) and backward (−) propagating electric field. The wave number is given by k, and ω is the optical frequency of the reference frame. From Equation 23.1, we can derive the propagation equations for the electric field amplitudes within the slowly varying envelope approximation (SVEA), neglecting all but the lowest order of derivatives. The SVEA limits the description to field envelopes that change on time scales and lengths much larger than the respective optical period and wavelength, respectively. The propagation equation for the slowly varying field envelope is then written as

$$\left(\frac{\partial}{\partial t} \pm v_g\frac{\partial}{\partial z}\right)E_\pm(z,t) = \frac{i\omega\Gamma}{2\varepsilon_{\text{bg}}\varepsilon_0}P_\pm(z,t) =: S_\pm(z,t)\,, \tag{23.3}$$

with the group velocity $v_g \equiv \frac{c_0}{\sqrt{\varepsilon_{\text{bg}}}}$, with the background permittivity ε_{bg}. $P_\pm(z,t)$ is the macroscopic slowly varying polarization amplitude. The transverse optical confinement factor Γ is introduced phenomenologically to the above equation. It describes the overlap of the electric field mode with the active medium, integrated over the transverse (x and y) coordinates. We summarize the right-hand side in a general source term $S_\pm(z,t)$.

The numerical solution of the above partial differential equation using a finite-difference method requires a spatial discretization into very fine sections of length $v_g\text{d}t$, where $\text{d}t$ is the numerical time-step, in order to ensure numerical stability [41]. For common device lengths on the order of 1 mm, a high number of spatial discretization points is needed. The numerical integration, therefore, becomes very expensive in terms of computation time and memory requirements [42,43]. A more elegant approach is the formulation of the problem as a delay-differential equation system [40,44], which we do in the following.

The partial differential equation Equation 23.3 can be formally solved by integrating along its characteristic lines, given by

$$z = \pm v_g t + \mathrm{const}. \tag{23.4}$$

Expanding the total derivative $\frac{\mathrm{d}}{\mathrm{d}t} = \frac{\partial z}{\partial t}\frac{\partial}{\partial z} + \frac{\partial}{\partial t}$, we can thus write

$$\frac{\mathrm{d}}{\mathrm{d}t}E_\pm(z,t) = \left[\pm v_g \frac{\partial}{\partial z} + \frac{\partial}{\partial t}\right] E_\pm(z,t) = S_\pm(z,t)\,. \tag{23.5}$$

Now we describe the optical amplifier of length ℓ by a number of Z sections along the propagation axis, such that the distance between two discretization points is given by $\Delta z := \ell/Z$. Integrating Equation 23.5 over the time interval $\Delta t = \frac{\Delta z}{v_g}$ thus yields

$$\begin{aligned} E_\pm(z,t) &= E_\pm(z \mp \Delta z, t-\Delta t) + \int_{-\Delta t}^{0} \left[\frac{\mathrm{d}}{\mathrm{d}t}E_\pm\left(z \pm v_g\tau, t+\tau\right)\right]\mathrm{d}\tau' \\ &= E_\pm(z \mp \Delta z, t-\Delta t) + \int_{-\Delta t}^{0} S_\pm\left(z \pm v_g\tau, t+\tau\right)\mathrm{d}\tau \\ &\approx E_\pm(z \mp \Delta z, t-\Delta t) + \frac{\Delta t}{2}\Big[S_\pm(z,t) + S_\pm(z \mp \Delta z, t-\Delta t)\Big]. \end{aligned} \tag{23.6}$$

The integral over the source term was approximated by its values at the end points of the integration interval. This approximation is valid for negligible change of $S_\pm$ along the integration path, i.e., for a sufficiently small space discretization step. The electric field at time t now depends on the values of $E_\pm$, $S_\pm$ at time $t-\Delta t$, which introduces a time delay into the equations.

The electric field in each of the spatial sections along the amplifier device thus couples to the time-delayed electric field in the neighboring sections, with the time Δt describing the time needed for the electric field propagation along the length of one section [40,44]. The resulting discretization scheme is illustrated in Figure 23.1. The advantage of this approach is the decoupling of the time and space discretization steps, ensuring numerical stability even for $\Delta z \gg v_g \mathrm{d}t$. The temporal dynamics at each point within the amplifier device is usually much faster than the characteristic timescale over which a propagating pulse changes its shape. Thus, a reduced number of spatial discretization points can be chosen compared to the finite-difference method, which significantly improves the simulation efficiency. The implementation of the delay-differential equations, however, means that the history of the electric fields as well as the source terms must be saved over a time interval Δt.

23.2.2 Quantum-Dot-in-a-Well Material Equations

The amplifier active medium is composed of an ensemble of of semiconductor quantum dots. For InAs quantum dots grown in the Stranski–Krastanov (SK) mode by molecular beam epitaxy (MBE), their areal density is in the order of $2-5\times10^{10}$ cm^{-2} [45]. The considered QDs are assumed to have two optically active localized states. These are denoted by GS and ES, for the ground and first excited state, respectively, which are located within the bands of the surrounding quantum well, as sketched in Figure 23.2. The resulting energy structure is sketched in Figure 23.3.

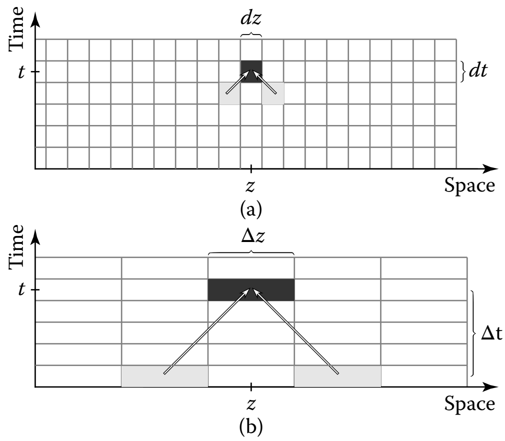

Time
dz
t
}dt
z
Space
(a)
Time
Δz
t
Δt
z
Space
(b)

FIGURE 23.1 Electric field propagation schemes: (a) Traditional time-domain traveling-wave integration scheme. The device is discretized into spatial sections with width dz, usually related to the numerical time step dt via the electric field group velocity d$z = v_g$dt. The field $E(z, t)$ is then determined from the field in neighboring discretization points at the previous time step $E(z \pm \mathrm{d}z, t - \mathrm{d}t)$. (b) The discretization into fewer spatial sections separated by Δz leads to a coupling to neighboring points with a time delay Δt. The time evolution in each section is still calculated with a time step dt, separating space and time discretization steps. (After J. Javaloyes and S. Balle, *Opt. Express*, 20, 8496–8502, 2012.)

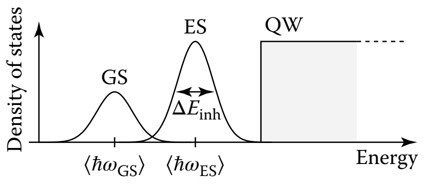

Density of states
ES
QW
GS
ΔE_{inh}
$\langle \hbar\omega_{\mathrm{GS}} \rangle$
$\langle \hbar\omega_{\mathrm{ES}} \rangle$
Energy

FIGURE 23.2 Sketch of the conduction band density of states of a quantum-dot-in-a-well material system. The QD transitions are assumed to consist of a ground-state (GS) and excited-state (ES) transition, which are inhomogeneously broadened with a full-width-at-half maximum (FWHM) of ΔE_{inh}. The two-dimensional quantum-well (QW) states act as a charge-carrier reservoir for the QD states.

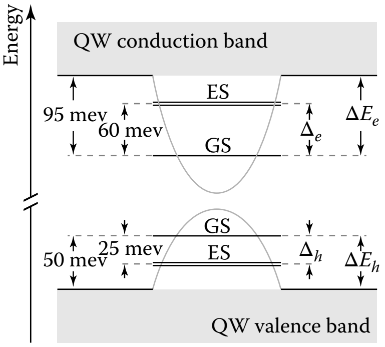

Energy
QW conduction band
ES
95 mev
60 mev
GS
Δ_e
ΔE_e
GS
25 mev
ES
50 mev
Δ_h
ΔE_h
QW valence band

FIGURE 23.3 Energy scheme of the considered dot-in-a-well (DWELL) structure. The QD ground states (GS) lie ΔE_b below the quantum-well (QW) band edge, with an energy spacing of Δ_b between GS and first excited state (ES).

Due to fluctuations in QD size and composition, their emission spectrum is inhomogeneously broadened around the central emission energy of $\approx$ 950 meV for InAs/InGaAs QDs. Typical inhomogeneous linewidths lie around 30 meV—higher than their homogeneous linewidth [27,46]. An accurate description of the QD medium, therefore, requires a separate treatment of quantum dots in dependence of their respective transition energy. Both the interaction with the optical field as well as the charge-carrier scattering dynamics are influenced by the transition and localization energies. A natural way to deal with the inhomogeneous broadening is, therefore, to distribute the whole QD ensemble into a number of subgroups, characterized by their optical transition energy. This procedure is illustrated in Figure 23.4. We label each subgroup by an index j and its respective mean transition energy $\hbar\omega^j$. We introduce the probability mass function $f(j)$ denoting the fraction of QDs within the jth subgroup. Following a Gaussian distribution with an FWHM of ΔE_{inh}, the distribution function $f(j)$ is then given by

$$f(j) = \frac{1}{\mathcal{N}} \exp\left(-4 \ln 2 \left[\frac{\hbar\omega^j - \langle \hbar\omega \rangle}{\Delta E_{\text{inh}}} \right]^2 \right), \tag{23.7}$$

with the normalization constant $\mathcal{N}$ chosen such that $\sum_j f(j) \overset{!}{=} 1$, calculated numerically. The inhomogeneous broadening of the optical spectrum is given by the sum of the individual single-particle state broadenings of electrons and holes (subscripts e and h, respectively),

$$\Delta E_{\text{inh}} = \Delta\varepsilon_e + \Delta\varepsilon_h\,, \tag{23.8}$$

where $\Delta\varepsilon_b$, $b \in \{e, h\}$ is the corresponding electron and hole state broadening. Only the total broadening ΔE_{inh} is experimentally readily accessible, e.g., by measurements of the QD luminescence spectra [47]. For the individual state broadening, we assume widths proportional to the localization energy of the given state:

$$\Delta\varepsilon_b = \Delta E_{\text{inh}} \frac{\Delta E_b}{\Delta E_e + \Delta E_h}\,. \tag{23.9}$$

We now proceed by deriving the dynamic equation for each of the QD subgroups within each amplifier section. In addition to the QD subgroup index j, we introduce the state index $m \in \{\text{GS}, \text{ES}\}$, distinguishing the charge carriers in the GS and first ES. The material dynamics are described within the Maxwell–Bloch approach [48], characterizing the QD active medium by their occupation probabilities ρ_m^j and the microscopic polarization amplitudes p_m^j. The microscopic polarization is induced by the electric field and describes the transition probability under emission of a photon, and thus couples the

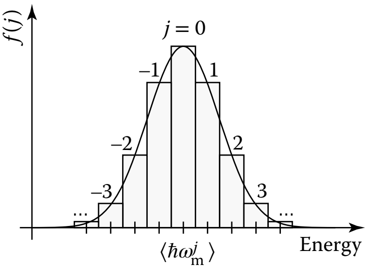


FIGURE 23.4 Illustration of the QD subgroups to model inhomogeneous broadening. The QDs are distributed into $j_{\max}$ subgroups, assumed to follow a Gaussian distribution around the mean transition energy $\langle \hbar\omega_m \rangle$. The probability mass function $f(j)$ gives the probability for a QD to be found in the jth subgroup.

charge-carrier equations to the light field. Similar to the electric field, we expand p^j_m in terms of forward and backward propagating amplitudes,

$$p^j_m(z,t) = \left[p^j_{m,+}(z,t)e^{ikz} + p^j_{m,-}(z,t)e^{-ikz}\right] e^{-i\omega t}. \tag{23.10}$$

We assume a coupling of the individual polarization amplitudes $p^j_{m,\pm}$ only to the corresponding co-propagating part of the electric field. This is the rotating wave approximation, neglecting terms oscillating with $e^{\pm 2ikz}$ and $e^{\pm 2i\omega t}$, which would appear in the coupling of counter-propagating polarization and field terms.

The Maxwell–Bloch equations of the QD active medium polarization amplitudes and occupation probabilities then read [49,50]

$$\frac{\mathrm{d}}{\mathrm{d}t}p^j_{m,\pm}(z,t) = -\left[i(\omega^j_m - \omega) + \frac{1}{\tau_2}\right] p^j_{m,\pm}(z,t) - i\frac{\mu_m}{2\hbar}\left(\rho^j_{e,m}(z,t) + \rho^j_{h,m}(z,t) - 1\right) E_\pm(z,t), \tag{23.11}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\rho^j_{b,m}(z,t) &= \frac{1}{\hbar}\mathrm{Im}\left[p^j_{m,+}(z,t)\mu^*_m E^*_+(z,t)\right] + \frac{1}{\hbar}\mathrm{Im}\left[p^j_{m,-}(z,t)\mu^*_m E^*_-(z,t)\right] \\ &\quad - W_m\rho^j_{e,m}(z,t)\rho^j_{h,m}(z,t) + \frac{\partial}{\partial t}\rho^j_{b,m}(z,t)\Big|_{\mathrm{sc}}.\end{aligned} \tag{23.12}$$

Here, we again denote the localized QD states by $m \in \{\mathrm{GS}, \mathrm{ES}\}$ and their subgroup index by j. The above equations describe the micoscopic polarization amplitudes in the reference frame of ω, whereas ω^j_m denotes the transition frequency of the respective optical transition. The polarization decays with the time constant τ_2. This dephasing time describes the loss of the quantum-mechanical coherence between different QDs due to scattering processes with other electrons or phonons. In semiconductors, these processes are usually in the order of a few tens of femtoseconds, due to the high density of charge carriers as potential scattering partners. The polarization is driven by the electric field $E_\pm(z,t)$ and the carrier inversion, with the interaction strength given by the transition dipole moment μ_m.

The QD charge-carrier dynamics is given by stimulated recombination induced by the microscopic polarization, and bimolecular spontaneous recombination with a rate W_m. The term $\frac{\partial}{\partial t}\rho^j_{b,m}(z,t)\Big|_{\mathrm{sc}}$ denotes charge-carrier scattering contributions, which we will address in the following section.

The above equations couple to the macroscopic light field via the macroscopic slowly varying polarization amplitudes $P_\pm(z,t)$. These are given by the macroscopic sum over the individual microscopic contributions:

$$P_\pm(z,t) = \frac{2N^{\mathrm{QD}}}{h^{\mathrm{QW}}} 2\sum_{j,m} \nu_m f(j)\mu^*_m p^j_{m,\pm}(z,t), \tag{23.13}$$

where N^{QD} is the areal QD density per quantum-well layer and h^{QW} the layer height. The sum $\sum_{j,m}\nu_m f(j)$ gives the sum over all QD subgroups and confined states, with ν_m the degeneracy (excluding spin) of the mth state and $f(j)$ the distribution function as defined in Equation 23.7. The macroscopic polarization couples back to the electric field equation as a source term, as shown in Equation 23.3:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}E_\pm(z,t) = S_\pm(z,t) &= \frac{i\omega\Gamma}{2\varepsilon_{\mathrm{bg}}\varepsilon_0}P_\pm(z,t) \\ &= \frac{2N^{\mathrm{QD}}}{h^{\mathrm{QW}}}\frac{i\omega\Gamma}{\varepsilon_{\mathrm{bg}}\varepsilon_0}\sum_{j,m}\nu_m f(j)\mu^*_m p^j_{m,\pm}(z,t).\end{aligned} \tag{23.14}$$

It is tempting to identify a small-signal gain $g(z, t, \omega)$, as used in traditional rate equation models, from an earlier equation:

$$g(z, t, \omega) = \frac{i\omega\Gamma}{2\varepsilon_{\mathrm{bg}}\varepsilon_0} \frac{P_\pm(z, t)}{E_\pm(z, t)}. \tag{23.15}$$

However, the independent dynamics of the polarization would lead to a strongly varying (and even diverging) value of $g(z, t, \omega)$ in time. Nevertheless, in a static limit, by setting $\frac{\partial}{\partial t} p^j_m = 0$, the adiabatic small-signal gain can be calculated from Equation 23.11, which yields

$$g(z, t, \omega) = \frac{\hbar\omega\Gamma}{\varepsilon_0\varepsilon_{\mathrm{bg}} h^{\mathrm{QW}}} 2N^{\mathrm{QD}} \sum_{j,m} \nu_m f(j) \frac{\tau_2 |\mu_m|^2}{2\hbar^2} \frac{\left[\rho^j_{e,m}(z, t) + \rho^j_{h,m}(z, t) - 1\right]}{1 + [\tau_2(\omega - \omega^j_m)]^2}. \tag{23.16}$$

This expression, however, neglects the dynamics of the polarization, assuming it to follow the incident electric field instantaneously. For coherent interactions, such as Rabi oscillations discussed later in this chapter, the above equation, therefore, cannot be applied. Furthermore, Equation 23.16 is strictly valid only for a monochromatic electric field at a given frequency ω. When simulating broadband amplification, such as optical pulses, the full polarization dynamics must, therefore, be taken into account. For slowly varying and near-monochromatic electric field envelopes, such as encountered in lasers, the adiabatic elimination of the polarization equations is a valid and widely used approximation, and Equation 23.16 can be used to calculate the optical gain.

The Maxwell–Bloch equations only describe the optically active QD states. In order to describe the carrier dynamics in the surrounding charge-carrier reservoir, we model the quantum-well charge-carrier density w_b by a rate equation,

$$\frac{\mathrm{d}}{\mathrm{d}t} w_b(z, t) = \eta \frac{J}{e_0} - A^S \sqrt{w_e(z, t) w_h(z, t)} - B^S w_e(z, t) w_h(z, t) + \frac{\partial}{\partial t} w_b(z, t)\Big|_{\mathrm{sc}}, \tag{23.17}$$

with the pump current density J, the electron charge $-e_0$, and a pump current efficiency η accounting for losses in the surrounding semiconductor layers, as well as imperfect carrier injection into the active region, which is not modeled explicitly. Linear and bimolecular recombination rates in the reservoir are given by A^S and B^S, respectively, which are important to describe carrier losses over a wide range of currents. We neglect Auger recombination, as with typical values ($C = 10^{-28}\ \mathrm{cm}^6 \cdot \mathrm{s}^{-1}$) the losses are found to be dominated by A^S and B^S in the current ranges considered in this chapter. The term $\frac{\partial}{\partial t} w_b(z, t)\Big|_{\mathrm{sc}}$ accounts for scattering of charge carriers into the QD states, which is derived in the next section. The above equations are defined for each space discretization point along the amplifier device, allowing for a spatially inhomogeneous distribution of the charge-carrier distribution, as encountered in long amplifier devices [51,52].

23.2.3 QD Charge-Carrier Scattering

QDs differ from conventional quantum-well or bulk gain media in the presence of localized states and thus a strongly modified density of states. Electrically injected charge carriers reach the active region in high-energy states near the bulk material band edges. The optically active QD states must, therefore, be populated by means of scattering processes. While the charge-carrier distributions within the quasi-continuous bulk and quantum-well bands quickly thermalize due to electron–phonon interaction [53,54], an effective scattering of carriers into the QD states by scattering with phonons is limited due to the existence of the “phonon bottleneck”[55]. Instead, at elevated charge-carrier densities, Coulomb-mediated carrier-carrier scattering provides an ultrafast scattering mechanism between the QD states and the surrounding semiconductor material [56]. In the following, we derive expressions for the Coulomb scattering mechanism.

The starting point for calculating the carrier-carrier scattering is the many-body Hamiltonian in second quantization [49,57],

$$H_{\text{sys}} = H_{\text{kin}} + H_{\text{C}} = \sum_{\substack{a\\ s}} \varepsilon_a a^{\dagger}_{as} a_{as} + \frac{1}{2} \sum_{\substack{abcd\\ ss'}} W_{abcd}\, a^{\dagger}_{as} a^{\dagger}_{bs'} a_{cs'} a_{ds}\,, \tag{23.18}$$

where a_x, $a^{\dagger}_x$ are the electron annihilation and creation operators in the state x with the energy ε_x, respectively. The Hamiltonian consists of the kinetic (free-carrier) contribution H_{kin}, and the Coulomb-interaction Hamiltonian H_{C}, which includes the many-body interaction between the charge carriers in the semiconductor. In the sums, the labels a, b, c, d denote all possible electronic states, with s, s' denoting their spins. The occupation probability for a given state (ν, s) is given by

$$\rho_{\nu s} = \langle a^{\dagger}_{\nu s} a_{\nu s} \rangle\,. \tag{23.19}$$

The screened Coulomb interaction matrix element is given by

$$W_{abcd} = \iint \mathrm{d}^3 r\, \mathrm{d}^3 r'\ \phi^*_a(\boldsymbol{r})\phi^*_b(\boldsymbol{r}') \frac{e_0{}^2}{4\pi\varepsilon_0\varepsilon_{\text{bg}}} \frac{e^{-\kappa|\boldsymbol{r}-\boldsymbol{r}'|}}{|\boldsymbol{r}-\boldsymbol{r}'|} \phi_c(\boldsymbol{r}')\phi_d(\boldsymbol{r})\,, \tag{23.20}$$

with the single-particle wave functions $\phi_x(\boldsymbol{r})$, approximated as harmonic oscillator wave functions, which was shown to yield surprisingly accurate results [58]. The vacuum and background permittivity are given by ε_0 and ε_{bg}, respectively, and $-e_0$ is the electron charge. The screening wave number κ describes the screening of the Coulomb interaction potential by the surrounding charge-carrier plasma, which can be calculated in a self-consistent way [57,59,60]. Here, we implement the screening wave number in the static quasi-equilibrium limit as

$$\kappa = \frac{e_0{}^2}{2\varepsilon_0\varepsilon_{\text{bg}}} \sum b \frac{\partial w_b}{\partial E^{\text{eq}}_{\text{F},b}} = \frac{e_0{}^2}{2\varepsilon_0\varepsilon_{\text{bg}}} \sum_b D_b f(E^{\text{QW}}_{b,0}, E^{\text{eq}}_{\text{F},b}, k_B T^{\text{eq}})\,, \tag{23.21}$$

with the quasi-Fermi level $E^{\text{eq}}_{\text{F},b}$, quasi-equilibrium temperature T^{eq}, and 2D density of states D_b of the corresponding electron and hole plasma. Within the quasi-equilibrium approximation, the screening can be expressed by the occupation probability at the band edge, $E^{\text{QW}}_{b,0}$, increasing the screening of the Coulomb interaction with increasing charge-carrier density. The screening becomes very important at elevated charge-carrier densities where the unscreened Coulomb potential would greatly overestimate the interaction between the charge carriers.

An exact numeric solution to Heisenberg's equation of motion within the given problem is generally not possible. The Coulomb interaction Hamiltonian couples the dynamic evolution of n-operator expectation values to $n+2$-operator expectation values, leading to an infinite number of coupled differential equations. It is, therefore, necessary to apply further approximations. We truncate this infinite chain of dynamic equations by factorizing six-operator expectation values into factors of two-operator expectation values, i.e., occupation probabilities. The Coulomb matrix elements thus enter the dynamic equations in up to second order. First-order contributions lead to renormalization effects of the single-particle energies and polarization amplitudes due to the Coulomb interaction with surrounding charge carriers [59,61]. These effects do not lead to a net change in the charge-carrier distribution and are therefore neglected.

By applying the Markov approximation to the resulting equations of motion, the explicit time dependence of four-operator expectation values is discarded. Instead, a quasi-static dependence of these expectation values on the occupation probabilities is assumed, by setting their time derivative to zero in the adiabatic limit. The scattering processes are thus assumed to follow any changes in the charge-carrier distribution instantaneously. The scattering contribution to the charge-carrier states then yields a Boltzmann-type equation:

$$\frac{\partial}{\partial t}\rho_{\nu\sigma}\Big|_{\mathrm{sc}} = \frac{2\pi}{\hbar}\sum_{\substack{bcd\\ s'}}\mathrm{Re}\left[W_{\nu bcd}\left(W^{*}_{\nu bcd} - W^{*}_{\nu bdc}\right)\right]\delta(\varepsilon_{\nu} + \varepsilon_{b} - \varepsilon_{c} - \varepsilon_{d})$$
$$\times\left[(1-\rho_{\nu\sigma})(1-\rho_{bs'})\rho_{cs'}\rho_{d\sigma} - \rho_{\nu\sigma}\rho_{bs'}(1-\rho_{cs'})(1-\rho_{d\sigma})\right], \tag{23.22}$$

which describes the Coulomb scattering in the second-order Born–Markov approximation [23,62–64]. The summation terms in Equation 23.22 describe the simultaneous scattering between states $d \leftrightarrow \nu$ and $c \leftrightarrow b$. The delta function ensures energy conservation, such that the total energy of the final states equals that of the initial states. The Coulomb scattering is thus revealed to be of Auger-type, requiring the simultaneous scattering of two electrons, with no net change in the total charge-carrier energy. The individual scattering processes are proportional to the occupation probabilities in the initial states c, d and proportional to the probability to find vacant final states ν, b, accounting for Pauli blocking. The second term in the sum in Equation 23.22 describes the reverse scattering processes, with electrons scattering out of the states ν, b.

Equation 23.22 can be written in the form of a Boltzmann equation for the occupation probability $\rho(t)$ of any given state in the system,

$$\frac{\partial}{\partial t}\rho(t)\Big|_{\mathrm{sc}} = S^{\mathrm{in}}[1-\rho(t)] - S^{\mathrm{out}}\rho(t), \tag{23.23}$$

combining the summation terms into an in-scattering rate S^{in} and a corresponding out-scattering rate S^{out}. These rates are calculated from the sums in Equation 23.22, which include all possible individual scattering processes. An exact treatment of the scattering dynamics within the second-order Born–Markov approximation would still require the dynamic tracking of the complete charge-carrier distribution, making numerical treatment difficult. Fortunately, the given QD-quantum-well system allows the distinction between qualitatively different scattering processes in order to break up the sums in Equation 23.22 into different parts which can be handled more easily.

For the dot-in-a-well (DWELL) structures considered in this chapter, two general classes of charge-carrier scattering processes can be distinguished: the capture of a quantum-well electron into a confined QD state, and the intradot electron relaxation, each with their respective inverse escape processes. This is illustrated in Figure 23.5a and b, respectively. The accompanying Auger-electron can involve either quantum-well states only, or transitions between quantum-well and other QD states. Note that depending on the involved energy differences, not all of these scattering channels are possible. For example, in the depicted case of the intra-dot relaxation in Figure 23.5b, the Auger transition in the valence band from the quantum well to the GS is not possible, as it would violate energy conservation. The possible scattering processes contributing to the total scattering rate thus strongly depend on the exact energy scheme of the QD-quantum-well system. Note that throughout this work impact ionization and Auger-assisted recombination, i.e., the direct scattering between conduction and valence bands, is not considered.

Following the earlier discussion, the scattering dynamics of the localized QD states are rewritten in the electron–hole picture as

$$\frac{\partial\rho_{b,\mathrm{GS}}}{\partial t}\Big|_{sc} = S^{\mathrm{cap,in}}_{b,\mathrm{GS}}(\{\rho_{\mathrm{QW}}\})(1-\rho_{b,\mathrm{GS}}) - S^{\mathrm{cap,out}}_{b,\mathrm{GS}}(\{\rho_{\mathrm{QW}}\})\rho_{b,\mathrm{GS}}$$
$$+ S^{\mathrm{rel,in}}_{b,\mathrm{GS}}(\{\rho_{\mathrm{QW}}\})\rho_{b,\mathrm{ES}}(1-\rho_{b,\mathrm{GS}}) - S^{\mathrm{rel,out}}_{b,\mathrm{GS}}(\{\rho_{\mathrm{QW}}\})(1-\rho_{b,\mathrm{ES}})\rho_{b,\mathrm{GS}}, \tag{23.24}$$

$$\frac{\partial\rho_{b,\mathrm{ES}}}{\partial t}\Big|_{sc} = S^{\mathrm{cap,in}}_{b,\mathrm{ES}}(\{\rho_{\mathrm{QW}}\})(1-\rho_{b,\mathrm{ES}}) - S^{\mathrm{cap,out}}_{b,\mathrm{ES}}(\{\rho_{\mathrm{QW}}\})\rho_{b,\mathrm{ES}}$$
$$+ S^{\mathrm{rel,in}}_{b,\mathrm{ES}}(\{\rho_{\mathrm{QW}}\})\rho_{b,\mathrm{GS}}(1-\rho_{b,\mathrm{ES}}) - S^{\mathrm{rel,out}}_{b,\mathrm{ES}}(\{\rho_{\mathrm{QW}}\})(1-\rho_{b,\mathrm{GS}})\rho_{b,\mathrm{ES}}. \tag{23.25}$$

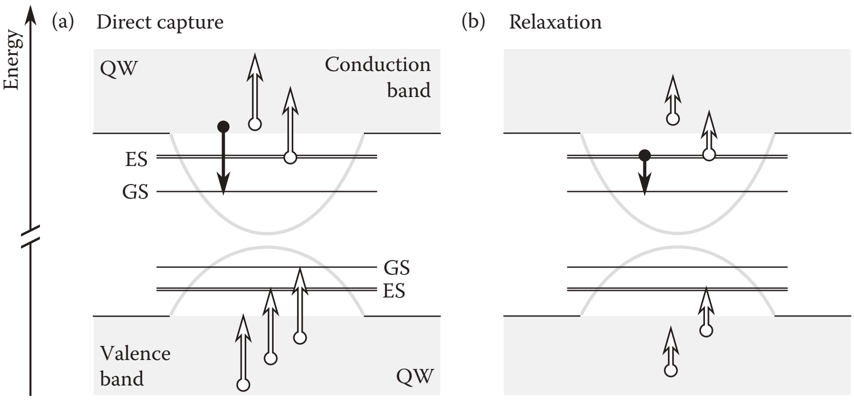


FIGURE 23.5 Possible scattering channels in quantum-dot (QD) quantum-well (QW) systems. (a) Direct capture into the QD ground state (GS). (b) Intradot relaxation from excited state (ES) to GS. The electron scattering process considered is shown by the black arrow, with the white arrows denoting the possible Auger processes. For all processes shown, the corresponding reverse scattering is also possible. Not shown is the direct capture into the QD ES, analogous to (a).

Here, ρ_b denotes either electron or hole occupation probabilities for $b \in \{e, h\}$. The scattering rates $S^{\text{cap,in}}$ denote the direct capture of quantum-well electrons into the QD states, $S^{\text{rel,in}}$ the intradot relaxation between the QD states, with S^{out} the scattering rate of the respective reverse processes. We can relate the relaxation processes of the ES to the corresponding GS terms,

$$S^{\text{rel,in}}_{b,\text{ES}}(\{\rho_{\text{QW}}\}) = -\frac{1}{2} S^{\text{rel,out}}_{b,\text{GS}}(\{\rho_{\text{QW}}\}), \tag{23.26}$$

$$S^{\text{rel,out}}_{b,\text{ES}}(\{\rho_{\text{QW}}\}) = -\frac{1}{2} S^{\text{rel,in}}_{b,\text{GS}}(\{\rho_{\text{QW}}\}), \tag{23.27}$$

with a factor $\frac{1}{2}$ compared to the GS contribution, due to the twofold degeneracy of the ES. All scattering rates in the above equations depend on the whole quantum-well distribution in both bands, denoted by $\{\rho_{\text{QW}}\}$.

Thus far the derived scattering expressions only describe the dynamics of QD states and their interaction with the quantum-well charge carriers. The dynamics of quantum-well carriers can in principle be expressed by Equation 23.22 as well. However, this would require resolving all quantum-well states and tracking their population distribution in time, which greatly increases the dimensionality of the system state. This problem can be resolved by assuming a specific distribution of the carrier population within the quantum well.

The intraband scattering between quantum-well states is typically in the order of $\approx$ 100 fs [65–69]. As long as this scattering is faster than the charge-carrier exchange between the quantum well and QDs, the quantum well can be assumed to be in quasi-equilibrium with good accuracy:

$$\rho_{b,\text{QW}}(\varepsilon^{\text{2D}}_{b,\boldsymbol{k}}) \approx f(\varepsilon^{\text{2D}}_{b,\boldsymbol{k}}, E^{\text{eq}}_{\text{F},b}, T^{\text{eq}}) \equiv \left[1 + \exp\left(\frac{\varepsilon^{\text{2D}}_{b,\boldsymbol{k}} - E^{\text{eq}}_{\text{F},b}}{k_{\text{B}} T^{\text{eq}}}\right)\right]^{-1}, \tag{23.28}$$

with the corresponding single-particle energies $\varepsilon^{\text{2D}}_{b,\boldsymbol{k}}$ and the quasi-Fermi level $E^{\text{eq}}_{\text{F},b}$. From this quasi-Fermi distribution the 2D-charge-carrier density w_b in the quantum wells can be calculated by taking the density of states in the quantum well as

$$\mathcal{D}_b(E) = D_b \Theta(E - E^{\text{QW}}_{b,0}) = \frac{m^*_b}{\pi\hbar^2} \Theta(E - E^{\text{QW}}_{b,0}), \tag{23.29}$$

under the assumption that the quantum-well sub-band spacing is large enough that only the lowest sub-band needs to be taken into account. The energy $E^{\mathrm{QW}}_{b,0}$ is the corresponding quantum-well band edge and Θ is the Heaviside function. The quantum-well charge-carrier density can then be written as

$$\begin{aligned} w_b &= \frac{2}{A_{\mathrm{act}}} \sum_{\boldsymbol{k}^{2\mathrm{D}}} \left[1 + \exp\left(\frac{\varepsilon^{2\mathrm{D}}_{b,\boldsymbol{k}} - E^{\mathrm{eq}}_{\mathrm{F},b}}{k_{\mathrm{B}} T^{\mathrm{eq}}} \right) \right]^{-1} \\ &= \int\limits_{-\infty}^{\infty} \mathrm{d}\varepsilon^{2\mathrm{D}}_{b,\boldsymbol{k}}\, \mathcal{D}_b(\varepsilon^{2\mathrm{D}}_{b,\boldsymbol{k}}) \left[1 + \exp\left(\frac{\varepsilon^{2\mathrm{D}}_{b,\boldsymbol{k}} - E^{\mathrm{eq}}_{\mathrm{F},b}}{k_{\mathrm{B}} T^{\mathrm{eq}}} \right) \right]^{-1} \\ &= D_b k_{\mathrm{B}} T^{\mathrm{eq}} \log\left[1 + \exp\left(\frac{E^{\mathrm{eq}}_{\mathrm{F},b} - E^{\mathrm{QW}}_{b,0}}{k_{\mathrm{B}} T^{\mathrm{eq}}} \right) \right], \end{aligned} \tag{23.30}$$

where the sum over all quantum-well $\boldsymbol{k}$-states was expressed as the integral over the charge-carrier energy. A_{act} is the active region in-plane area, with the factor 2 accounting for spin degeneracy. By inverting the above expression, the quasi-Fermi level $E^{\mathrm{eq}}_{\mathrm{F},b}$ can be expressed in terms of the charge-carrier density in the quantum well,

$$E^{\mathrm{eq}}_{\mathrm{F},b} = E^{\mathrm{QW}}_{b,0} + k_{\mathrm{B}} T^{\mathrm{eq}} \log\left[\exp\left(\frac{w_b}{D_b k_{\mathrm{B}} T^{\mathrm{eq}}} \right) - 1 \right]. \tag{23.31}$$

Thus, the quantum-well charge-carrier population can be expressed as a function of the carrier density and the quasi-equilibrium temperature:

$$\rho_{b,\mathrm{QW}}(\varepsilon^{2\mathrm{D}}_{b,\boldsymbol{k}}) \equiv \rho_{b,\mathrm{QW}}(\varepsilon^{2\mathrm{D}}_{b,\boldsymbol{k}}, w_b, T^{\mathrm{eq}}) = \left[1 + \exp\left(\frac{\varepsilon^{2\mathrm{D}}_{b,\boldsymbol{k}} - E^{\mathrm{eq}}_{\mathrm{F},b}(w_b, T^{\mathrm{eq}})}{k_{\mathrm{B}} T^{\mathrm{eq}}} \right) \right]^{-1}. \tag{23.32}$$

By entering this relation into the expressions for the scattering rates Equation 23.22, also the individual scattering rates can be expressed as functions of only the 2D charge-carrier densities w_b and their quasi-equilibrium temperature T^{eq}, eliminating the need to keep track of the microscopic carrier population distribution.

Furthermore, it is now possible to relate the in- and out-scattering rates of a given scattering process to each other [70,71]. The out-scattering contribution in Equation 23.22 is equivalent to the in-scattering contribution under the replacement $\rho \to 1 - \rho$, which for the quantum well in quasi-equilibrium can be expressed as

$$1 - \rho_{\mathrm{QW}}(\varepsilon^{2\mathrm{D}}_{b,\boldsymbol{k}}) = \rho_{\mathrm{QW}}(\varepsilon^{2\mathrm{D}}_{b,\boldsymbol{k}}) \exp\left(\frac{\varepsilon^{2\mathrm{D}}_{b,\boldsymbol{k}} - E^{\mathrm{eq}}_{\mathrm{F},b}(w_b, T^{\mathrm{eq}})}{k_{\mathrm{B}} T^{\mathrm{eq}}} \right). \tag{23.33}$$

For the QD scattering processes the out-scattering rates can thus be written as [72]:

$$S^{\mathrm{cap,out}}_{b,m}(w_e, w_h, T^{\mathrm{eq}}) = S^{\mathrm{cap,in}}_{b,m}(w_e, w_h, T^{\mathrm{eq}}) \exp\left(\frac{\varepsilon^{\mathrm{QD}}_{b,m} - E^{\mathrm{eq}}_{\mathrm{F},b}}{k_{\mathrm{B}} T^{\mathrm{eq}}} \right) \tag{23.34}$$

$$S_b^{\text{rel,out}}(w_e, w_h, T^{\text{eq}}) = S_b^{\text{rel,in}}(w_e, w_h, T^{\text{eq}}) \exp\left(\frac{\varepsilon_{b,\text{GS}}^{\text{QD}} - \varepsilon_{b,\text{ES}}^{\text{QD}}}{k_{\text{B}} T^{\text{eq}}}\right), \tag{23.35}$$

where $\varepsilon_{b,m}^{\text{QD}}$ denotes the energy of the localized QD state, with $m \in \{\text{GS}, \text{ES}\}$ distinguishing between GS and ES. The out-scattering of charge carriers thus becomes more probable at elevated charge-carrier temperatures [73,74]. Note that in the derivation of above expressions, only a quasi-equilibrium within the quantum well must be assumed without making assumptions about the QD occupations. Equations 23.34 and 23.35 are therefore also valid in nonequilibrium situations between QD and quantum well.

The resulting scattering contribution to the dynamics of the QD occupation probability is, therefore, written as

$$\begin{aligned} \left.\frac{\partial \rho_{b,\text{GS}}^{j}(z,t)}{\partial t}\right|_{\text{sc}} &= S_{b,\text{GS}}^{\text{cap,in}}\left[1 - \rho_{b,\text{GS}}^{j}\right] - S_{b,\text{GS}}^{\text{cap,out}} \rho_{b,\text{GS}}^{j} \\ &\quad + S_b^{\text{rel,in}} \rho_{b,\text{ES}}^{j}\left[1 - \rho_{b,\text{GS}}^{j}\right] - S_b^{\text{rel,out}} \rho_{b,\text{GS}}^{j}\left[1 - \rho_{b,\text{ES}}^{j}\right] \end{aligned} \tag{23.36}$$

$$\begin{aligned} \left.\frac{\partial \rho_{b,\text{ES}}^{j}(z,t)}{\partial t}\right|_{\text{sc}} &= S_{b,\text{ES}}^{\text{cap,in}}\left[1 - \rho_{b,\text{ES}}^{j}\right] - S_{b,\text{ES}}^{\text{cap,out}} \rho_{b,\text{ES}}^{j} \\ &\quad - \frac{1}{2}\left\{S_b^{\text{rel,in}} \rho_{b,\text{ES}}^{j}\left[1 - \rho_{b,\text{GS}}^{j}\right] - S_b^{\text{rel,out}} \rho_{b,\text{GS}}^{j}\left[1 - \rho_{b,\text{ES}}^{j}\right]\right\}. \end{aligned} \tag{23.37}$$

The scattering contribution to the quantum-well equations can be calculated from charge-carrier number conservation. It is simply given by the total charge-carrier density captured in the QD states:

$$\left.\frac{\partial}{\partial t} w_b(z,t)\right|_{\text{sc}} = -2N^{\text{QD}} \sum_{j,m} \nu_m f(j) \left(S_{b,m}^{\text{cap,in}}(z,t)\left[1 - \rho_{b,m}^{j}(z,t)\right] - S_{b,m}^{\text{cap,out}}(z,t) \rho_{b,m}^{j}(z,t)\right). \tag{23.38}$$

23.2.4 Fit Functions for Scattering Rates

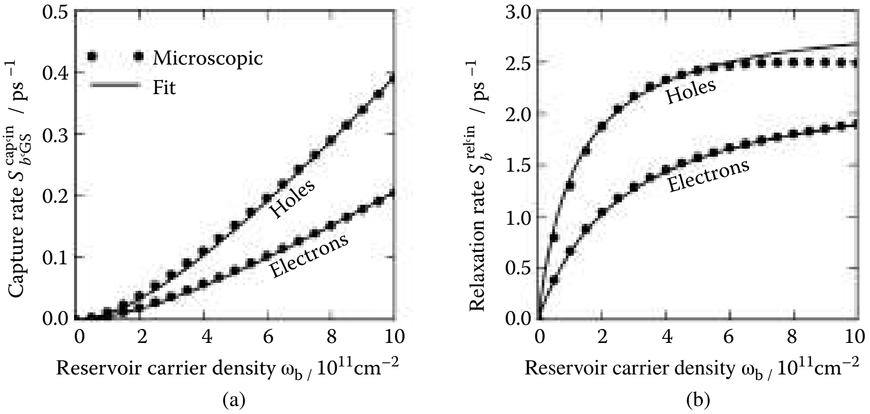
The microscopic calculation of the QD charge-carrier scattering rates requires a summation over all possible electronic states in the DWELL system, and such a calculation is, therefore, quite expensive in terms of computation time. While the scattering rates can be implemented by means of lookup tables, having a simpler analytic expression would be advantageous.

A first look at the scattering rates in dependence of the reservoir densities w_b reveals for the capture rates a quadratic increase at low densities, and a transition to nearly linear increase at higher values of w_b. The relaxation rates, on the other hand, show a linear increase at first and subsequent saturation. This is depicted in Figure 23.6 (circles). Taking these characteristics into account, we fit the scattering rates, using the following functions

$$S_{b,m}^{\text{cap,in}}(w_b) = \frac{A {w_b}^2}{B + w_b} \tag{23.39}$$

$$S_{b,m}^{\text{rel,in}}(w_b) = \frac{C w_b}{D + w_b}. \tag{23.40}$$

The corresponding out-scattering rates are calculated via Equations 23.34 and 23.35. Table 23.1 gives the fitting parameters extracted from the microscopically calculated rates. The comparison shown in Figure 23.6 shows good agreement between the microscopically calculated rates and the fit functions. For high reservoir carrier densities the fits show a slight deviation, especially pronounced in the hole



(a)

(b)

FIGURE 23.6 Fits of the quantum-dot scattering rates. Shown are (a) the ground-state direct-capture and (b) the intradot relaxation rate for electrons and holes, in dependence of the electron reservoir density w_b at $T = 300$ K. The microscopically calculated rates (circles) are shown together with the simple fit functions (lines). The fit parameters are given in Table 23.1.

TABLE 23.1 Fitting Parameters for the Scattering Rates

	Electrons		Holes	
	GS	ES	GS	ES
A (10^{-11}cm^2·ns^{-1})	32	54	60	79
B (10^{11}cm^{-2})	5.6	2.9	5.3	2.1
C (ns^{-1})	2400		3000	
D (10^{11}cm^{-2})	2.7		1.2	

Confinement energies: $\Delta E_e = 95$ meV, $\Delta_e = 60$ meV,
$\Delta E_h = 50$ meV, $\Delta_h = 25$ meV.

relaxation rate. These values, however, correspond to very strong electrical pumping. For currents typically used in these devices, the fit functions presented are in very good agreement with the microscopic calculations.

The relatively simple expressions extracted from the fits to the microscopically calculated scattering rates show that the common approach of capture rates which are directly proportional to the reservoir charge-carrier density [17,75] should even yield quantitatively acceptable results. The saturation of the intradot relaxation rates at elevated carrier densities suggests the use of constant relaxation rates when simple QD models are desired. Nevertheless, when the accurate description of dynamics over a large range of operating conditions is required, the microscopic description of the QD scattering should be preferred.

23.2.5 Modeling of Spontaneous Emission

The spontaneous emission created in optical amplifier devices will be subject to stimulated amplification when emitted along the propagation axis. This ASE can reach significant optical power levels and is important for the device characteristics and performance. Apart from adding an optical noise background to the device output, it can often become strong enough to influence the charge-carrier dynamics [76–78]. These effects will deteriorate the device performance and the optical signal quality.

A consistent description of the ASE is important. In general there exist two appropriate modeling approaches: the deterministic description of the ASE power spectral density in the frequency space [17,79] and the stochastic description in timedomain [43,78]. Here, we will employ the stochastic description, which simplifies the inclusion of time-varying input signals.

We, therefore, phenomenologically add an additional source term on the right-hand side of Equation 23.5, modeling the stochastic spontaneous emission added to the propagating electric field:

$$\frac{\mathrm{d}}{\mathrm{d}t}E_{\pm}(z,t') = S_{\pm}(z,t') + S_{\pm}^{\mathrm{sp}}(z,t')\,. \tag{23.41}$$

The electric field propagation along the one space-discretization section is again determined by integration of Equation 23.41 over the interval Δz:

$$E_{\pm}(z,t) \approx E_{\pm}(z \mp \Delta z, t-\Delta t) + \frac{\Delta t}{2}\left[S_{\pm}(z,t) + S_{\pm}(z \mp \Delta z, t-\Delta t)\right] + \int_{-\Delta t}^{0} S_{\pm}^{\mathrm{sp}}\left(z \pm v_g\tau, t+\tau\right)\mathrm{d}\tau\,. \tag{23.42}$$

The spontaneous emission source term must account for all optical transitions in the inhomogeneously broadened QD ensemble. Let $\eta_m^j(z,t)$ describe the spontaneous emission contribution of the jth subgroup of the mth localized state. We write for the electric field spontaneously added to the propagating field along Δz:

$$\int_{-\Delta t}^{0} S_{\pm}^{\mathrm{sp}}\left(z \pm v_g\tau, t+\tau\right)\mathrm{d}\tau \equiv \sum_{m,j} \eta_m^j(z,t)\,. \tag{23.43}$$

The spontaneous emission of an optical transition has a finite linewidth given by its homogeneous broadening. The homogeneous linewidth of a given transition is directly linked to the dephasing time τ_2 of the microscopic polarization, with the linewidth simply given by $2\tau_2^{-1}$. In order to correctly implement the spectral properties of the ASE, the spontaneously emitted field $\eta_m^j(z,t)$ must, therefore, have the correct linewidth and center frequency. It is thus not possible to describe the spontaneous emission by white noise, which would produce a flat noise spectrum; it must instead be modeled using colored noise. We implement this colored noise by two-dimensional Ornstein–Uhlenbeck processes [80], which describe quantities which are driven by white noise but relax with a given rate γ toward zero. This relaxation rate leads to a finite “memory” of the process, which translates into a finite spectral width around a center frequency. The time evolution of each of the respective noise signals is modeled by the following stochastic differential equation:

$$\frac{\mathrm{d}}{\mathrm{d}t}\eta_m^j(z,t) = -(i\omega_m^j + \gamma)\eta_m^j(z,t) + \sqrt{D_{\mathrm{sp},m}^j(z,t)}\,\tilde{\xi}_m^j(z,t)\,, \tag{23.44}$$

where $\tilde{\xi}(z,t)$ is a complex Gaussian white noise process, which is δ-correlated both in z and t. The relaxation rate of η_m^j is given by γ, and its center frequency by ω_m^j. The noise signal then fulfills the following properties [80]:

$$\langle \mathrm{Re}\,\eta_m^j(z,t)\rangle = \langle \mathrm{Im}\,\eta_m^j(z,t)\rangle = 0 \tag{23.45}$$

$$\langle |\eta_m^j(z,t)|^2\rangle = \frac{D_{\mathrm{sp},m}^j(z,t)}{\gamma} \tag{23.46}$$

$$\langle \eta_m^j(z,t)\eta_m^{j\,*}(z',t+\tau)\rangle \approx \frac{D_{\mathrm{sp},m}^j(z,t)}{\gamma}e^{-\gamma|\tau|}e^{i\omega_m^j\tau}\delta_{z,z'}\,. \tag{23.47}$$

Equation 23.47 is valid only under the assumption of a slowly varying noise amplitude $\partial_t D_{\mathrm{sp}}(z,t) \ll \gamma$, such that within one correlation time γ^{-1} the spontaneous emission amplitude can be assumed as constant. We can use the Wiener–Khinchin-theorem [81] this relation can be used to calculate the power spectrum $\mathcal{S}_{\eta_m^j}(z,\omega)$ of $\eta_m^j(z,t)$:

$$\mathcal{S}_{\eta_m^j}(z,\omega) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \langle \eta_m^j(z,t)\eta_m^{j\,*}(z,t+\tau)\rangle e^{-i\omega\tau}\mathrm{d}\tau \tag{23.48}$$

$$= \frac{D_{\mathrm{sp},m}^j(z)}{\pi}\frac{1}{(\omega_m^j-\omega)^2+\gamma^2}, \tag{23.49}$$

which yields a Lorentzian line shape with a FWHM of 2γ. We thus identify $\gamma = (\tau_2)^{-1}$, such that the noise linewidth equals the homogeneous linewidth of the QD transitions.

Using the noise correlation properties, the average power that is added to the electric field by the noise can be calculated. Combining Equations 23.42 and 23.43 and summarizing the deterministic source terms in a combined variable, $\widetilde{S}_{\pm}^{\mathrm{stim}}$, yields for the electric field:

$$E_{\pm}(z,t) = E_{\pm}(z\mp\Delta z, t-\Delta t) + \widetilde{S}_{\pm}^{\mathrm{stim}}(z,t) + \sum_{m,j}\eta_m^j(z,t) \tag{23.50}$$

$$\left\langle\left|E_{\pm}(z,t)\right|^2\right\rangle = \left|E_{\pm}(z\mp\Delta z,t-\Delta t)+\widetilde{S}_{\pm}^{\mathrm{stim}}(z,t)\right|^2 + \sum_{m,j}\left\langle\left|\eta_m^j(z,t)\right|^2\right\rangle$$

$$= \left|E_{\pm}(z\mp\Delta z,t-\Delta t)+\widetilde{S}_{\pm}^{\mathrm{stim}}(z,t)\right|^2 + \sum_{m,j}\tau_2 D_{\mathrm{sp},m}^j(z,t), \tag{23.51}$$

where we have used the zero mean property, $\langle\eta_m^j\rangle = 0$. On average, the spontaneous emission thus increases the squared modulus of the electric field along one space discretization step during the propagation time Δt by $\sum_{m,j}\tau_2 D_{\mathrm{sp},m}^j(z,t)$. Or, written in terms of a time derivative,

$$\frac{\partial}{\partial t}|E_{\pm}(z,t)|^2\Big|_{\mathrm{sp}} = \frac{\tau_2}{\Delta t}\sum_{m,j} D_{\mathrm{sp},m}^j(z,t). \tag{23.52}$$

In the photon picture, the average change of the electric field energy density due to the spontaneous emission can be calculated:

$$\frac{\partial}{\partial t}u(z,t)\Big|_{\mathrm{sp}} = \frac{\varepsilon_{\mathrm{bg}}\varepsilon_0}{2}\frac{\partial}{\partial t}|E_{\pm}(z,t)|^2\Big|_{\mathrm{sp}}$$

$$= \beta\frac{2N^{\mathrm{QD}}\Gamma}{h^{\mathrm{QW}}}\sum_{m,j}\nu_m f(j)\hbar\omega_m^j W_m \rho_{\mathrm{e},m}^j(z,t)\rho_{\mathrm{h},m}^j(z,t). \tag{23.53}$$

Here we model spontaneous emission by bimolecular recombination processes, given by the sum over all QD states. The spontaneous emission ratio β gives the fraction of photons spontaneously emitted into the waveguide mode. Comparing Equations 23.52 and 23.53 yields for the individual noise strengths

$$D_{\mathrm{sp},m}^j(z,t) = \frac{\Delta t}{\tau_2}\frac{2\beta\Gamma\hbar\omega_m^j 2N^{\mathrm{QD}}}{\varepsilon_{\mathrm{bg}}\varepsilon_0 h^{\mathrm{QW}}}\nu_m f(j)\frac{1}{2}\left[R_{\mathrm{sp},m}^j(z,t)+R_{\mathrm{sp},m}^j(z\mp\Delta z,t-\Delta t)\right], \tag{23.54}$$

where the average of the spontaneous emission rate at the endpoints of the integration interval $[z, z \mp \Delta z]$ was taken, defined by

$$R^{j}_{\mathrm{sp},m}(z,t) := W_m \varrho^{j}_{\mathrm{e},m}(z,t)\varrho^{j}_{\mathrm{h},m}(z,t)\,. \tag{23.55}$$

The spontaneous emission noise thus depends on the optical frequency and on the occupation of the individual QD subgroups. The resulting propagation equation for the electric field is thus given by

$$E_{\pm}(z,t) \approx E_{\pm}(z \mp \Delta z, t - \Delta t) + \frac{\Delta t}{2}\Big[S_{\pm}(z,t) + S_{\pm}(z \mp \Delta z, t - \Delta t)\Big] + \sum_{m,j} \eta^{j}_{m}(z,t)\,. \tag{23.56}$$

23.3 Application Examples

The QDSOA model derived in the previous section will now be applied to an exemplary device. We implement Equations 23.11 through 23.17, which describe the active medium dynamics, along with Equation 23.56 to describe the electric field propagation including spontaneous emission noise. The charge-carrier scattering is described within the full microscopic framework as described in Equations 23.36 through 23.38, with the individual scattering rates depending on the reservoir carrier densities as well as the device temperature.

In this section we will at first investigate the static characteristics of the amplifier, focusing on the unique properties of the QD active medium. We thus calculate the pump-current dependent-gain spectra for the sample device, along with ASE spectra. Our modeling approach allows for a characterization over a large range of pump currents, accounting for the changing carrier dynamics by the nonlinear scattering rates. We analyze the performance of the QDSOA for optical signals centered on either the GS or ES energies, exploiting the broad gain spectrum due to the different localized states. Subsequently, we characterize the dynamic gain recovery after perturbation of the gain medium by a strong pulse. The gain recovery dynamics provide an important insight into the ultrafast amplification capabilities of an amplifier. Furthermore, we show how the internal charge-carrier dynamics imprint their signature onto the gain dynamics. The last part of this section addresses the phenomenon of coherent pulse-shaping by Rabi oscillations in QDSOAs, enabled by the comparably long dephasing time of the interband polarization. The interaction of ultrashort pulses with the QD gain medium, therefore, leads to strong modifications of the pulse shape. These modifications differ from classical descriptions and can be very complex due to the individual dynamics of QDs within the inhomogeneous ensemble.

23.3.1 Static Characterization of the QDSOA

In linear amplification applications, the device performance of SOAs is generally limited by two competing effects. On the one hand, the maximum achievable optical output power is limited by the charge carriers available for stimulated emission. The gain of the amplifier will, therefore, decrease when the optical power becomes too large. This effect is known as gain saturation. On the other hand, a too small optical signal will significantly reduce the signal-to-noise ratio, as the spontaneous emission background will dominate the output. Noise effects thus play an important role in the amplification of optical data signals [82,83]. A strong noise background will negatively impact the signal quality by distorting the corresponding optical output, and potentially corrupting the transmitted data stream. It is, therefore, important to investigate the ASE and gain of a given amplifier device to assess its suitability in a given application.

We implement the previously derived QD amplifier model to simulate the static characteristics of a specific device. We solve the delay-differential equation system using the simple forward-Euler method. The Gaussian white noise for implementing the spontaneous emission is generated by the Box–Muller algorithm [84]. For each spatial section of the amplifier device, we save a history array containing $\Delta t/\mathrm{d}t$

entries (up to Δt before the current time t) of the electric field $E_{\pm}(z, t)$, the field source term $S_{\pm}(z, t)$, and the spontaneous emission rate $R^{j}_{\mathrm{sp},m}(z, t)$.

The modeled amplifier is a 3-mm-long DWELL structure, consisting of ten 5-nm-thick InGaAs quantum wells, each embedding a density of $3 \times 10^{10}\,\mathrm{cm}^{-2}$ InAs QDs with a shallow-etched, 4-μ-wide ridge waveguide. Refer to Figure 23.3 for an illustration of the modeled energy structure. The device parameters that will be used here and in the following sections are given in Table 23.2, unless otherwise noted.

Additional phenomenological dependencies of the dephasing time and the device temperature on the applied pump current are included to account for heating effects as well as carrier-induced dephasing. We assume a temperature that increases linearly with the applied current,

$$T_{\ell}(j) = 295\,\mathrm{K} + \frac{j}{14}\frac{\mathrm{K}}{\mathrm{mA}}. \tag{23.57}$$

The increase in temperature significantly alters the detailed balance relationship, Equations 23.34 and 23.35, which determines the out-scattering rates out of the localized QD states. This, in turn, determines the quasi-equilibrium distribution that is reached in the steady state. Furthermore, the dephasing time of the QD interband polarization is known to decrease under strong excitation and with temperature [27,28]. We thus introduce a current-dependent dephasing time,

$$\tau_2(j) = \frac{\tau_2^0}{1 + \frac{j}{300\,\mathrm{mA}}}, \tag{23.58}$$

TABLE 23.2 Model Parameters Used in the Simulations, unless Otherwise Stated.

Symbol	Value	Meaning
N^{QD}	$3 \times 10^{10}\,\mathrm{cm}^{-2}$	QD density per layer
a_L	10	Number of layers
h^{QW}	5 nm	QW layer height
w_{wg}	4μ	Waveguide width
n_{bg}	3.77	Background index
ΔE_{inh}	30 meV	QD inhomogeneous broadening FWHM
η	0.4	Current injection efficiency
A^S	$0.7\,\mathrm{ns}^{-1}$	QW linear recombination rate
B^S	$50\,\mathrm{nm}^2\,\mathrm{ns}^{-1}$	QW bimolecular recombination rate
W_m	$1\,\mathrm{ns}^{-1}$	QD spontaneous recombination rate
β	3.5×10^{-4}	Spontaneous emission ratio
$\langle\hbar\omega_{\mathrm{GS}}\rangle$	963 meV (1288 nm)	QD GS center emission
$\langle\hbar\omega_{\mathrm{ES}}\rangle$	1048 meV (1183 nm)	QD ES center emission
μ_m	$0.6\,\mathrm{nm}\,e_0$	QD transition dipole moment
τ_2	$200\,\mathrm{fs} \times \left(1 + \frac{j}{300\,\mathrm{mA}}\right)^{-1}$	QD polarization dephasing time
T_{ℓ}	$295\,\mathrm{K} + \frac{j}{14}\frac{\mathrm{K}}{\mathrm{mA}}$	Lattice temperature
Γ	0.045	Geometric confinement factor
$\Delta E_e (\Delta E_h)$	95 meV (50 meV)	Electron (hole) QD GS localization energy
$\Delta_e (\Delta_h)$	60 meV (25 meV)	Electron (hole) QD GS–ES energy spacing
Z	31	Number of space discretization steps
dt	1 fs	Numeric time step

QW, quantum well; QD, quantum dot; GS, ground state; ES, excited state; FWHM, full-width half maximum.

where τ_2^0 denotes the corresponding dephasing time at $j = 0$. The above effects lead to broadening and subsequent decrease of the optical gain at high pump currents, as frequently observed in experiments [85]. Note that the dephasing time can in principle be calculated from microscopic scattering processes similar to the formalism presented in Section 23.2.3. A rigorous treatment of these effects, however, requires microscopic approaches that go beyond the employed Born–Markov approximation [86–90], which are beyond the scope of this chapter.

With the above additions, we calculate the key characteristics of the amplifier device, starting with the small-signal optical gain spectrum. The small-signal gain $G(\omega)$ is calculated from the stationary QD distribution (i.e., without a disturbing optical signal) via

$$G(\omega) = \exp\left[\frac{2}{v_g}\int_0^{\ell} g(z,t,\omega)\,\mathrm{d}z\right], \tag{23.59}$$

with the instantaneous amplitude gain

$$g(z,t,\omega) = \frac{\hbar\omega\Gamma}{\varepsilon_0\varepsilon_{\mathrm{bg}}h^{\mathrm{QW}}}2N^{\mathrm{QD}}\sum_{j,m}\nu_m f(j)\frac{\tau_2|\mu_m|^2}{2\hbar^2}\frac{\left[\rho^j_{e,m}(z,t)+\rho^j_{h,m}(z,t)-1\right]}{1+[\tau_2(\omega-\omega^j_m)]^2} \tag{23.60}$$

determined from the steady-state solution to Equation 23.11. The resulting spectra are shown in Figure 23.7a, clearly showing the GS gain peak around 1300 nm, which saturates and decreases for higher pump currents, at which the ES gain becomes dominant. We plot the optical gain at the GS and ES transitions, at $\lambda = 1288$ nm and $\lambda = 1183$ nm, respectively, under the dependence of the pump current, shown in Figure 23.7b.

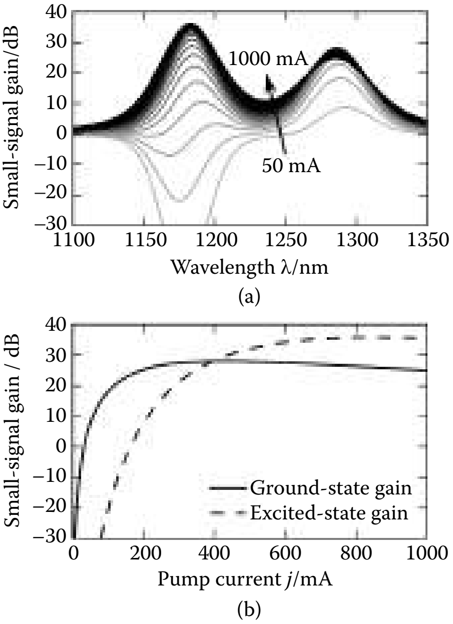


FIGURE 23.7 (a) Static gain spectra of the quantum dot semiconductor optical amplifier device for increasing pump currents up to $j = 1000$ mA in steps of 50 mA. (b) Small-signal gain at the ground (solid) and excited state (dashed) wavelengths ($\lambda_{\mathrm{GS}} = 1288$ nm, $\lambda_{\mathrm{ES}} = 1183$ nm) under the dependence of the pump current.

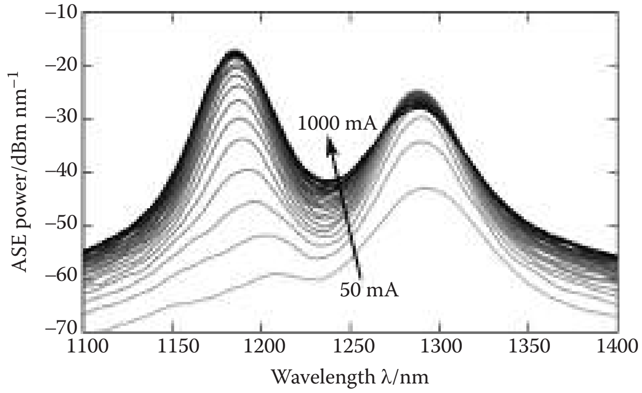


FIGURE 23.8 Simulated amplified spontaneous emission (ASE) spectra for increasing pump currents up to $j = 1000$ mA in steps of 50 mA.

Additionally, we can calculate the ASE spectra, shown in Figure 23.8. These are calculated by direct integration of the dynamic equations and subsequent Fourier transform of the electric field time series at the output facets. The resulting spectra can be used for a direct quantitative comparison with experimental data. The output power at the corresponding end facets is related to the electric field via

$$P_{\pm}^{\text{out}} = A_{\text{beam}} \frac{\varepsilon_0 n_{\text{bg}} c_0}{2} |E_{\pm}^{\text{out}}|^2 , \tag{23.61}$$

where $A_{\text{beam}} = \frac{a_L h^{\text{QW}} w_{\text{wg}}}{\Gamma}$ is the beam area in the waveguide and $E_{\pm}^{\text{out}} = \left\{E_+(\ell), E_-(0)\right\}$ denotes the forward- and back-propagating electric fields at their respective end facets. The ASE spectra mimic the qualitative behavior of the gain spectra, with clearly distinguishable GS and ES peaks and a saturation for strong pump currents (see Figure 23.8, where increasing gray levels indicate higher pump current). In addition, an initial blueshift of the emission peaks is visible especially for the ES emission. This is due to state filling, i.e., the shift of the quasi-Fermi levels toward higher energies with increasing carrier densities. At low currents, only the lowest energy states are occupied by charge carriers, and only with increasing current can the higher lying energy states be filled. In the model, this is accounted for by the detailed balance relation between the in- and out-scattering rates (Equations 23.34 and 23.35), which ensures the relaxation toward a quasi-Fermi distribution in the steady state. The quasi-Fermi level shifts toward higher energies with increasing reservoir carrier density, thus leading to a higher occupation in higher energy states.

When an optical signal of significant power is amplified by the gain medium, its charge-carrier distribution can be substantially perturbed. This leads to a power-dependent response of the device and thus induces nonlinearities. In order to quantify this response, we simulate the amplifier under the injection of a constant optical input signal with varying power. As we have seen previously, the existence of localized QD states makes it possible for the amplifier to work on GS and ES wavelengths. We will, therefore, investigate the device performance under amplification of signals at either of the two corresponding wavelengths. We put in an electric field at the input facet as

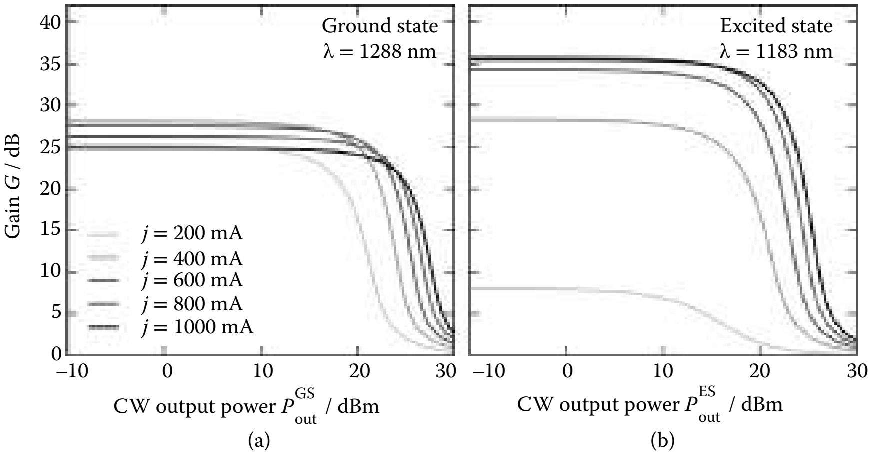
$$E_+(0, t) = E_{\text{in}}^{m} e^{-i(\langle\omega^m\rangle - \omega)t} , \tag{23.62}$$

with the frequency detuning $(\langle\omega^m\rangle - \omega)$ $(m \in \{\text{GS}, \text{ES}\})$ of the input signal relative to the carrier frequency chosen to yield signals centered on the energy of either QD state. After the propagation of the signal through the amplifier device and after a steady state is reached, we can simply evaluate the power-dependent device gain by the power ratio

$$G = \frac{|E_+(\ell, t)|^2}{|E_+(0, t)|^2} . \tag{23.63}$$

Figure 23.9a shows the GS gain in dependence on the achievable optical output power for different pump currents. The small-signal gain, i.e., at low input powers, shows the behavior that we have seen before in Figure 23.7, with a saturation and subsequent decrease of the gain for increasing current. In the linear amplification regime, for lower optical power, the gain is flat. Under these small-signal conditions, the optical power is not strong enough to perturb the gain medium appreciably. Once the optical power becomes large enough, the nonlinear or gain saturation regime is reached [10,11,79]. In Figure 23.9, the onset of the GS saturation regime increases from 18 dBm output power at $j = 200$ mA to 25 dBm at $j = 1000$ mA. Here, the number of charge carriers injected into the optically active QD states is not sufficient to replenish the QD states that are depleted by the signal, and the optical gain thus gradually decreases.

A general trend toward higher achievable output power with increasing current and corresponding shift of the nonlinear regime toward higher optical power can be observed both in the GS and ES (see Figure 23.9a and b). This is a direct consequence of the increased in-scattering rates and the higher reservoir charge-carrier density [10]. The faster and more efficient refilling of the QD states after depletion by the optical signal shifts the saturation regime toward higher power. The ES shows a transition to the nonlinear gain regime at lower optical power than the GS, as shown in Figure 23.9b. This can be understood by the weaker confinement of the QD ES relative to the reservoir, which increases the sensitivity of the ES occupation to changes in the reservoir charge-carrier density. Nevertheless, the ES allows a high gain and a sufficiently high saturation power to allow for its application in optical communication networks. Furthermore, the ultra-broad gain bandwidth of QDSOAs could be exploited for a simultaneous amplification of data signals with wide wavelength spacing. The low cross-talk between the GS and ES transitions makes an error-free amplification of two suitable data signals possible, as was recently shown [91].



(a) (b)

FIGURE 23.9 Optical gain G of the quantum-dot amplifier in dependence of the optical output power P_{out}^{m} for different pump currents j. Shown is the response to optical signals centered on (a) the ground state, and (b) the excited-state wavelength.

23.3.2 Gain Recovery Dynamics in QDSOAs

We now wish to characterize the dynamic performance of the modeled QD semiconductor amplifier. When optical amplifiers are used in optical data signal transmission, e.g., for propagation loss compensation, their dynamics must be fast enough to follow the signal symbol rate in order to minimize patterning effects and reduce transmission errors, as detailed in Chapter 25 on patterning effects. In experimental setups the gain recovery dynamics can be characterized by observing the time evolution of the optical gain after a perturbation of the active medium. This is commonly done in a pump-probe setup [19,92–94]. The principle is as follows: A strong and temporally narrow optical pump pulse is injected into the waveguide and depletes charge carriers during its amplification. A weaker probe pulse injected with a time delay $\Delta t_{\mathrm{p-p}}$ with respect to the pump pulse then experiences a different gain, depending on the perturbation of the carrier distribution after this delay time. By repeating the measurement for a range of $\Delta t_{\mathrm{p-p}}$ and measuring the probe pulse intensity, the time evolution of the gain after propagation of the pump pulse can be extracted experimentally.

The gain-recovery dynamics is obtained from the QDSOA model by simulating the amplifier in a pump-probe experiment. The perturbation of the gain medium is induced by a strong Gaussian pump pulse injected into the input facet at $z = 0$:

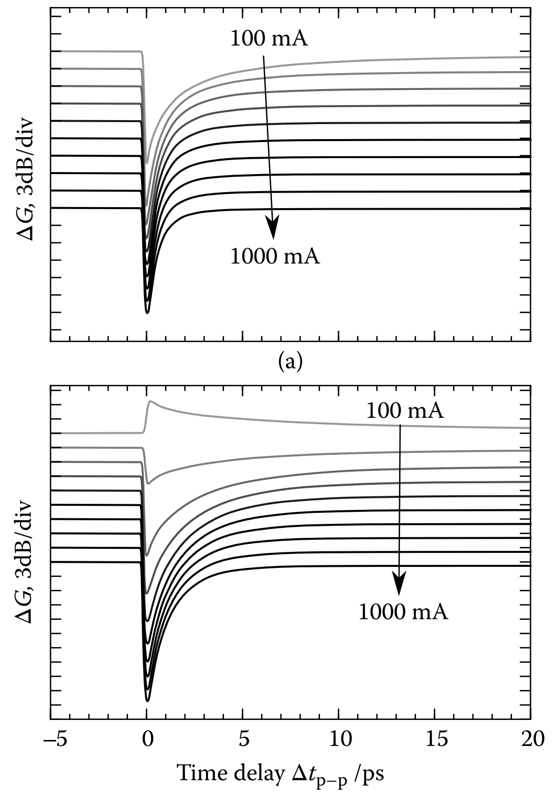
$$E_{+}(0,t) = E_{\mathrm{in}} \exp\left[-4\ln 2\left(\frac{t-t_0}{\Delta_{\mathrm{FWHM}}}\right)^2\right] e^{-i\delta\omega t}, \tag{23.64}$$

with an input amplitude E_{in}, arrival time t_0, and pulse width Δ_{FWHM}. We additionally allow for a detuning $\delta\omega$ of the pump pulse with respect to the chosen optical reference frequency which we can use to tune the pump wavelength. In order to probe the gain after the pump pulse has passed the active medium, the probe pulse does not have to be modeled explicitly. Instead, we can directly calculate the gain that the probe would experience along the device:

$$G(\Delta t_{\mathrm{p-p}},\omega) = \exp\left[\frac{2}{v_g}\int_0^{\ell} g\left(z, t_0 + \Delta t_{\mathrm{p-p}} + \frac{z}{v_g}, \omega\right) \mathrm{d}z\right]. \tag{23.65}$$

Here, we used the instantaneous amplitude gain, as defined in Equation 23.60, integrated along the device in a copropagating frame with a delay time $\Delta t_{\mathrm{p-p}}$ relative to the pump pulse. Note that in pump-probe experiments the temporal resolution is limited by the temporal width of the probe pulse. Measuring the integrated probe pulse power averages the extracted gain over the pulse profile.

As an example, we simulate the amplifier under the influence of a pump pulse with a width of $\Delta_{\mathrm{FWHM}} = 300$ fs and a pulse energy of 1 pJ, centered on the GS gain peak. In Figure 23.10a the resulting gain recovery curves are shown in dependence of the pump current. Here, we show the gain difference ΔG with respect to the unperturbed case. A strong initial reduction in gain can be observed due to the depletion of charge carriers by the pump pulse, with a subsequent recovery toward the unperturbed value on a picosecond timescale. This fast recovery is the signature of the charge-carrier refilling by carrier scattering, which as a consequence is strongly current-dependent and mimics the current dependence of the scattering rates. Pump-probe gain recovery measurements can, therefore, be used to extract effective charge-carrier scattering timescales of the active QD medium [95]. In addition to the GS performance, we take a look at the dynamics of the ES gain in Figure 23.10b. Here, for $j = 100$ mA, the pump-pulse induces an initial increase of the gain, as seen in the initially increasing light gray curve. At this current, the ES is below transparency, and the pump pulse is absorbed, which increases the inversion and thus the gain. For higher pump currents (increasing gray levels in Figure 23.10) the dynamics follows the same trends as the GS, albeit with a slower recovery. Here again the less efficient carrier refilling of the ES becomes apparent. The intricate scattering



(b)

FIGURE 23.10 Gain recovery after perturbation with a strong pump pulse at $\Delta t_{\text{p}-\text{p}} = 0$. (a) Ground-state gain after excitation at the ground-state wavelength centered at $\lambda = 1288$ nm, and (b) excited-state gain after excitation at the excited-state wavelength centered at $\lambda = 1183$ nm. The current was increased from $j = 100$ mA to $j = 1000$ mA in steps of 100 mA. Shown is the gain difference with respect to the unperturbed case, with individual curves shifted vertically by 3 dB for improved readability.

dynamics between the different QD and reservoir states in the active medium, therefore, directly affects the performance of QD optical amplifiers, emphasizing the importance of its accurate modeling, as described in Section 23.2.3.

With the QDSOA model at hand, we can take a closer look at the dynamics of the QD gain medium. The gain recovery can be seen to consist of three main components, i.e., it can be fitted very accurately by a triexponential recovery, as shown in Figure 23.11a. Here, we do not show the fit for $t \lesssim 0.1$ ps as for small times the pulse is still interacting with the medium, which the fit cannot describe. We can attribute the three time constants to the intradot relaxation, being the fastest component, the charge-carrier capture from the reservoir, and a slow recovery of the whole system, governed mainly by the reservoir carrier lifetime. An experimental determination of the individual timescales, however, is usually difficult due to their narrow separation. This is already evident from the three effective timescales that determine the gain recovery. In principle, we would expect at least seven different timescales: GS and ES capture rates and intradot relaxation for both electrons and holes, along with the recovery of the reservoir. The three extracted timescales must, therefore, be interpreted as effective values.

Our model also allows us to analyze the charge-carrier dynamics within the gain medium directly, which is not possible within the experimental setup. We thus evaluate the variation of the carrier occupation in

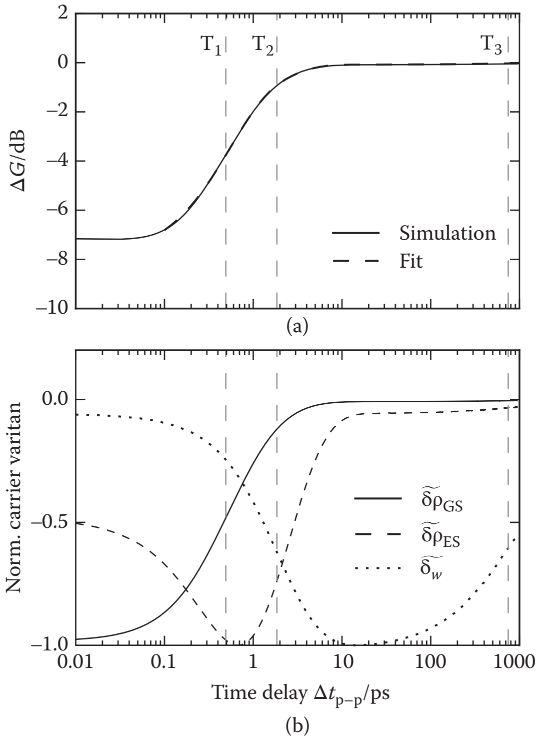


FIGURE 23.11 (a) Fit to the gain recovery after a perturbation of the ground state at $j = 500$ mA. The three characteristic timescales, $\tau_1 = 0.49$ ps, $\tau_2 = 1.8$ ps, $\tau_3 = 750$ ps, correspond to the intradot relaxation, charge-carrier capture, and reservoir refilling, respectively. (b) Time evolution of the normalized carrier variations with respect to their respective steady-state values in the GS, ES, and reservoir states.

the QD GS and ES as well as in the reservoir states at the output facet, by defining a normalized carrier variation,

$$\widetilde{\delta\rho}_m(t) = \frac{\sum_b \left[\rho_{b,m}(t) - \rho^0_{b,m}\right]}{\max\left|\sum_b \left[\rho_{b,m}(t) - \rho^0_{b,m}\right]\right|}, \tag{23.66}$$

where ρ^0_m denotes the corresponding QD occupation in the unperturbed steady state. A similar expression for the reservoir density variation, $\widetilde{\delta w}$, can be written down. During the perturbation the normalized variations reach their minimum value of −1 and during recovery slowly grow back to a value of zero. The variations of the GS, ES, and reservoir carriers are plotted in Figure 23.11b, as solid, dashed, and dotted lines, respectively. We can clearly observe the three qualitative steps in the recovery of the gain medium we discussed earlier. The GS occupation is minimal right after the pulse has passed the device. On a timescale close to τ_1, the GS starts a fast recovery under a reduction of the ES occupation, illustrating the intradot relaxation from the ES states to the GS. These relaxation processes dominate the ultrafast gain recovery of QDSOAs. The ES then starts to recover with a timescale $\approx \tau_2$ due to charge-carrier capture from the reservoir, which is subsequently depleted. The timescale τ_3 then constitutes the recovery time of the reservoir, which equilibrates the whole system again.

For the modeled device, the two fast timescales of the GS gain recovery decrease from $\tau_1 = 1.4$ ps, $\tau_2 = 7.5$ ps at a current of 100 mA to $\tau_1 = 0.43$ ps, $\tau_2 = 1.2$ ps at $j = 1000$ mA (not shown here). Together with the increasing saturation power at higher currents, a high operating current could be seen as optimal for high-power ultrafast applications. However, the saturation and decrease of the optical gain at elevated pump currents limits the range of possible operating conditions.

23.3.3 Rabi Oscillations in QD Semiconductor Amplifiers

When the quantum-mechanic phase is maintained over macroscopically accessible timescales, coherent effects can induce phenomena that significantly differ from classical predictions. The precondition for this is a sufficiently long dephasing time of the quantum-mechanic coherence. The comparably long lifetime of the interband polarization in semiconductor QDs makes them promising candidates for applications in quantum-optics [96]. Recently, coherent pulse propagation in macroscopic semiconductor devices has been observed in quantum-cascade lasers [97] and quantum-dash [30] and QD semiconductor amplifiers [32], relying on ultrashort, strong optical pulses. It is important to note that in the experiments using optical amplifiers, the measurements were performed at room temperature, opening up possible future quantum-coherent applications using uncooled devices.

In this section, we present a theoretical description of Rabi oscillations induced by ultrashort pulses in a QDSOA at room temperature [32]. Rabi oscillations denote a periodic exchange of energy between the optical field and the active medium, which leads to characteristic modifications of the optical pulse traveling through the amplifier device. The principle of Rabi oscillations can be derived from the dynamic equation for a single QD subgroup:

$$\frac{\partial}{\partial t}p_m^j(z,t) = -\left[i(\omega_m^j - \omega) + \frac{1}{\tau_2}\right]p_m^j - i\frac{\mu_m}{2\hbar}\left(\rho_{e,m}^j + \rho_{h,m}^j - 1\right)E(z,t), \tag{23.67}$$

$$\frac{\partial}{\partial t}\rho_{b,m}^j(z,t) = \frac{1}{\hbar}\mathrm{Im}\left[p_m^j\mu_m^* E^*(z,t)\right] - W_m\rho_{e,m}^j\rho_{h,m}^j + \left.\frac{\partial}{\partial t}\rho_{b,m}^j\right|_{\mathrm{sc}}. \tag{23.68}$$

We reduce the above equations by introducing the inversion $d = (\rho_{e,m}^j + \rho_{h,m}^j - 1)$ and neglecting losses. We also limit ourselves to a resonant excitation, i.e., $\omega = \omega_m^j$:

$$\frac{\partial}{\partial t}p_m^j(z,t) = -i\frac{\mu_m}{2\hbar}d(z,t)E(z,t), \tag{23.69}$$

$$\frac{\partial}{\partial t}d(z,t) = \frac{2}{\hbar}\mathrm{Im}\left[\mu_m^* E^*(z,t)p_m^j(z,t)\right]. \tag{23.70}$$

For an initial inversion of d_0 and zero polarization, the above equations have the solution

$$p_m^j(z,t) = -\frac{i}{2}\sin\Theta(z,t), \tag{23.71}$$

$$d(z,t) = d_0\cos\Theta(z,t), \tag{23.72}$$

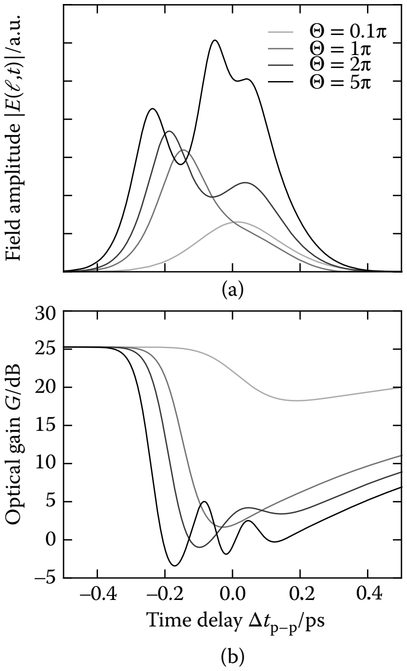
with the pulse area defined as

$$\Theta(z,t) = \int_{-\infty}^{t}\frac{\mu_{\mathrm{GS}}}{2\hbar}\left|E(z,t')\right|\mathrm{d}t'. \tag{23.73}$$

Equation 23.72 shows that, with a proper choice of the pulse area, it is possible, e.g., to invert the charge-carrier distribution by choosing $\Theta = \pi$. When choosing $\Theta = n2\pi, n \in \mathbb{N}$, i.e., integer multiples of 2π, the system returns to its initial state after the exciting pulse has passed through. This important property shows that the coherent interaction between the optical field and the active material is a reversible process, as long as the polarization dephasing time is much larger than the pulse width. In most cases, however, this condition is not strictly fulfilled, leading to an additional damping of the Rabi oscillations[48].

We now apply our QD amplifier model to the amplification of strong optical pulses. We use the same pulse width of 300 fs as in the previous section, but with higher input amplitudes and we evaluate the output pulse shape as well as the integrated optical gain, as defined in Equation 23.65. For different input pulse areas Θ, the field amplitude at the output facet is shown in Figure 23.12a. At a small pulse area $\Theta = 0.1\pi$ (23 fJ pulse energy, light gray line) the pulse retains its Gaussian shape during propagation. For increasing pulse energies (darker lines in Figure 23.12), a pronounced shoulder at the leading edge of the pulse appears and the pulse shape noticeable diverts from its original Gaussian envelope. For $\Theta = 2\pi$ a pronounced dip in amplitude is visible, and a second dip appears for $\Theta = 5\pi$ (black line), where a complex pulse shape is obtained. A look into the dynamics of the optical gain, shown in Figure 23.12b, reveals a strong decrease in gain after injection of the optical pulse, as already seen in the previous section. For strong pulses, the gain exhibits oscillations and intermittently becomes negative, i.e., absorbing (see black line in Figure 23.12). This is a clear indication of the occurrence of Rabi oscillations, i.e., the coherent interaction between the microscopic polarization and the inversion. For an incoherent interaction, i.e., neglecting the coherent dynamics of the polarization, the optical gain cannot be reduced below transparency in optically pumped media. The coherent dynamics thus imprints a clear signature onto the QD amplifier behavior, which can differ strongly from conventional expectations.

We take a closer look at the dynamics of the amplifier under the amplification of strong pulses in Figure 23.13. The dynamics of the microscopic polarization and the carrier occupation across the inhomogeneously broadened QD GS at the output facet are visualized as gray scale values in Figure 23.13b and c, respectively. The optical pulse can be seen to induce a complex response of the gain medium. The QD



(a)

(b)

FIGURE 23.12 Rabi oscillations in the QD semiconductor optical amplifier. (a) Output electric field amplitudes for different input pulse areas $\Theta \in \{0.1\pi, 1\pi, 2\pi, 5\pi\}$, corresponding to pulse energies of 0.023, 2.3, 9.4, 58 pJ, respectively. (b) Integrated optical gain in dependence of the time delay $\Delta t_{\text{p-p}}$. The pump current was set to $j = 200$ mA.

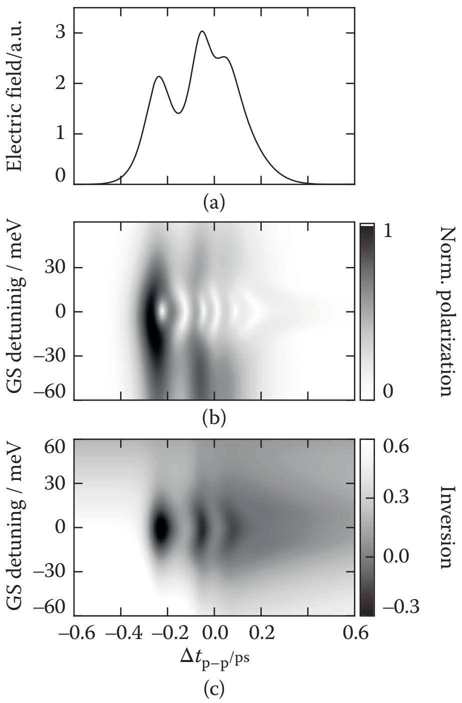


FIGURE 23.13 Quantum-dot (QD) ground-state dynamics for an input pulse area $\Theta = 5\pi$. (a) Output electric field amplitude. (b) Time evolution of the microscopic polarization amplitude p_m^j in dependence of the QD GS transition energy. The color code denotes the normalized polarization amplitude. (c) QD ground-state carrier dynamics in dependence of the QD GS transition energy. The color code denotes the QD subgroup inversion. The pump current was set to $j = 200$ mA.

subgroups within the ensemble which are resonant to the pulse (detuning of 0 meV), exhibit pronounced Rabi oscillations visible in both the polarization and inversion. The outer subgroups show this behavior to a lesser extent but with a different temporal profile. The resulting macroscopic polarization that drives the pulse is a superposition of the individual microscopic polarization amplitudes across the whole QD ensemble (see Equation 23.13). The response of the gain medium to such strong optical pulses can, therefore, be very complex and go beyond what a simple two-level system would predict. This underlines the importance of taking both homogeneous as well as inhomogeneous effects into account when modeling light–matter interaction in QD devices in the high-power and ultrafast regime.

23.4 Conclusion

In this chapter, we have presented a model for QDSOAs that takes into account the intricate scattering dynamics between the localized QD states and the surrounding carrier reservoir states. The Coulomb scattering, which dominates the carrier dynamics, can be microscopically calculated to yield Boltzmann-like terms describing the charge-carrier scattering, which can be implemented numerically in terms of lookup tables, or approximated by simplified fit functions. The description of the electric field propagation along the device is modeled using a delay-differential equation approach, which allows a separation of numeric time step and spatial discretization length for an efficient integration algorithm.

The complex charge-carrier dynamics determine both the static and dynamic behavior of QDSOAs and is responsible for their ultrafast gain recovery. An accurate modeling of these dynamics is, therefore, necessary for a quantitatively accurate description of these optoelectronic devices. The simulations of an exemplary QDSOA device show that the signature of the internal charge-carrier dynamics can be directly seen in dynamic gain recovery measurements. The strong pump-current dependence of the scattering timescales lead to the existence of optimal operating conditions when either high gain or ultrafast gain recovery is desired. The long dephasing time of QDs compared to other semiconductor materials leads to a macroscopically accessible measurement of quantum-coherence manifesting itself in the break-up of ultrashort strong optical pulses, even at room temperature. The induced Rabi oscillations lead to a complex response of the QD gain medium which requires a detailed modeling of the polarization dynamics across the inhomogeneously broadened QD ensemble. The thoroughly nonclassical response of the QD medium to ultrashort pulses due to long-lived coherence might pave the way to novel quantum-optical devices using semiconductor QDs as the medium of choice.

References

1. P. Harrison, *Quantum Wells, Wires and Dots: Theoretical and Computational Physics of Semiconductor Nanostructures*. Chichester, UK: John Wiley & Sons (2005).
2. K. Lüdge, *Nonlinear Laser Dynamic—From Quantum Dots to Cryptography*. Weinheim: Wiley-VCH (2012).
3. D. Bimberg, Quantum dot based nanophotonics and nanoelectronics, *Electron. Lett.*, 44 (2008), p. 168.
4. J. J. Coleman, The development of the semiconductor laser diode after the first demonstration in 1962, *Semicond. Sci. Technol.*, 27 (2012), p. 090207.
5. D. Bimberg, D. Arsenijevic, G. Larisch, H. Li, J. A. Lott, P. Moser, H. Schmeckebier, and P. Wolf, Green nanophotonics for future datacom and Ethernet networks, *Proc. SPIE.*, 9134 (2014), pp. 913402–913913.
6. T. D. Steiner, *Semiconductor Nanostructures for Optoelectronic Applications, Semiconductor Materials and Devices Library*. Norwood, MA: Artech House (2004).
7. B. Lingnau, Nonlinear and Nonequilibrium Dynamics of Quantum-Dot Optoelectronic Devices, PhD thesis, Technical University of Berlin (2015).
8. B. Lingnau, W. W. Chow, E. Schöll, and K. Lüdge, Feedback and injection locking instabilities in quantum-dot lasers: A microscopically based bifurcation analysis, *New J. Phys.*, 15 (2013), p. 093031.
9. T. Akiyama, N. Hatori, Y. Nakata, H. Ebe, and M. Sugawara, Pattern-effect-free amplification and cross-gain modulation achieved by using ultrafast gain nonlinearity in quantum-dot semiconductor optical amplifiers, *Phys. Stat. Sol. B*, 238 (2003), pp. 301–304.
10. T. W. Berg, J. Mørk, and J. M. Hvam, Gain dynamics and saturation in semiconductor quantum dot amplifiers, *New J. Phys.*, 6 (2004), p. 178.
11. A. V. Uskov, E. P. O'Reilly, M. Lämmlin, N. N. Ledentsov, and D. Bimberg, On gain saturation in quantum dot semiconductor optical amplifiers, *Opt. Commun.*, 248 (2005), p. 211.
12. T. Akiyama, H. Kuwatsuka, N. Hatori, Y. Nakata, H. Ebe, and M. Sugawara, Symmetric highly efficient (~0dB) wavelength conversion based on four-wave mixing in quantum dot optical amplifiers, *IEEE Photon. Technol. Lett.*, 14 (2002), pp. 1139–1141.
13. N. Majer, K. Lüdge, and E. Schöll, Maxwell-Bloch approach to four-wave mixing in quantum dot semiconductor optical amplifiers, In *IEEE Proceedings of the 11th International Conference on Numerical Simulation of Optical Devices (NUSOD)*, edited by J. Piprek, pp. 153–154. Rome (2011).
14. O. Qasaimeh, Theory of four-wave mixing wavelength conversion in quantum dot semiconductor optical amplifiers, *IEEE Photon. Technol. Lett.*, 16 (2004), pp. 993–995.
15. H. Schmeckebier, C. Meuer, D. Arsenijevic, G. Fiol, C. Schmidt-Langhorst, C. Schubert, G. Eisenstein, and D. Bimberg, Wide-range wavelength conversion of 40-Gb/s NRZ-DPSK signals using a 1.3-μm quantum-dot semiconductor optical amplifier, *IEEE Photon. Technol. Lett.*, 24 (2012), pp. 1163–1165.

16. G. Contestabile, A. Maruta, S. Sekiguchi, K. Morito, M. Sugawara, and K. Kitayama, Cross-gain modulation in quantum-dot SOA at 1550 nm, *IEEE J. Quant. Electron.*, 46 (2010), pp. 1696–1703.
17. J. Kim, M. Laemmlin, C. Meuer, D. Bimberg, and G. Eisenstein, Theoretical and experimental study of high-speed small-signal cross-gain modulation of quantum-dot semiconductor optical amplifiers, *IEEE J. Quant. Electron.*, 45 (2009), pp. 240–248.
18. C. Meuer, H. Schmeckebier, G. Fiol, D. Arsenijevic, J. Kim, G. Eisenstein, and D. Bimberg, Cross-gain modulation and four-wave mixing for wavelength conversion in undoped and p-doped 1.3-µm quantum dot semiconductor optical amplifiers, *IEEE Photon. J.*, 2 (2010), pp. 141–151.
19. P. Borri, W. Langbein, J. M. Hvam, F. Heinrichsdorff, M. H. Mao, and D. Bimberg, Ultrafast gain dynamics in InAs–InGaAs quantum-dot amplifiers, *IEEE Photon. Technol. Lett.*, 12 (2000), 594–596.
20. S. Dommers, V. V. Temnov, U. Woggon, J. Gomis, J. Martinez-Pastor, M. Lämmlin, and D. Bimberg, Complete ground state gain recovery after ultrashort double pulses in quantum dot based semiconductor optical amplifier, *Appl. Phys. Lett.*, 90 (2007), p. 033508.
21. M. Lämmlin, GaAs-Based Semiconductor Optical Amplifiers with Quantum Dots as an Active Medium, PhD thesis, Technical University of Berlin (2006).
22. N. Majer, S. Dommers-Völkel, J. Gomis-Bresco, U. Woggon, K. Lüdge, and E. Schöll, Impact of carrier-carrier scattering and carrier heating on pulse train dynamics of quantum dot semiconductor optical amplifiers, *Appl. Phys. Lett.*, 99 (2011), p. 131102.
23. N. Majer, K. Lüdge, and E. Schöll, Cascading enables ultrafast gain recovery dynamics of quantum dot semiconductor optical amplifiers, *Phys. Rev. B.*, 82 (2010), p. 235301.
24. M. van der Poel, E. Gehrig, O. Hess, D. Birkedal, and J. M. Hvam, Ultrafast gain dynamics in quantum-dot amplifiers: Theoretical analysis and experimental investigations, *IEEE J. Quant. Electron.*, 41 (2005), pp. 1115–1123.
25. C. Meuer, C. Schmidt-Langhorst, R. Bonk, H. Schmeckebier, D. Arsenijevic, G. Fiol, A. Galperin, J. Leuthold, C. Schubert, and D. Bimberg, 80 Gb/s wavelength conversion using a quantum-dot semiconductor optical amplifier and optical filtering, *Opt. Express*, 19 (2011), pp. 5134–5142.
26. P. Borri, W. Langbein, S. Schneider, U. Woggon, R. L. Sellin, D. Ouyang, and D. Bimberg, Ultralong dephasing time in InGaAs quantum dots, *Phys. Rev. Lett.*, 87 (2001), p. 157401.
27. P. Borri, W. Langbein, S. Schneider, U. Woggon, R. L. Sellin, D. Ouyang, and D. Bimberg, Exciton relaxation and dephasing in quantum-dot amplifiers from room to cryogenic temperature, *IEEE J. Sel. Top. Quantum Electron.*, 8 (2002), pp. 984–991.
28. T. Koprucki, A. Wilms, A. Knorr, and U. Bandelow, Modeling of quantum dot lasers with microscopic treatment of Coulomb effects, *Opt. Quant. Electron.*, 42 (2011), pp. 777–783.
29. P. Borri, W. Langbein, S. Schneider, U. Woggon, R. L. Sellin, D. Ouyang, and D. Bimberg, Rabi oscillations in the excitonic ground-state transition of InGaAs quantum dots, *Phys. Rev. B*, 66 (2002), p. 081306(R).
30. A. Capua, O. Karni, G. Eisenstein, and J. P. Reithmaier, Rabi oscillations in a room-temperature quantum dash semiconductor optical amplifier, *Phys. Rev. B*, 90 (2014), p. 045305.
31. H. Kamada, H. Gotoh, J. Temmyo, H. Ando, and T. Takagahara, Exciton Rabi oscillation in single isolated quantum dots, *Phys. Stat. Sol. A*, 190 (2002), pp. 485–490.
32. M. Kolarczik, N. Owschimikow, J. Korn, B. Lingnau, Y. Kaptan, D. Bimberg, E. Schöll, K. Lüdge, and U. Woggon, Quantum coherence induces pulse shape modification in a semiconductor optical amplifier at room temperature, *Nature Commun.*, 4 (2013), p. 2953.
33. T. H. Stievater, X. Li, D. G. Steel, D. Gammon, D. S. Katzer, D. Park, C. Piermarocchi, and L. J. Sham, Rabi oscillations of excitons in single quantum dots, *Phys. Rev. Lett.*, 87 (2001), p. 133603.
34. A. Icsevgi and W. E. Lamb, Propagation of light pulses in a laser amplifier, *Phys. Rev.*, 185 (1969), pp. 517–545.
35. S. Schneider, P. Borri, W. Langbein, U. Woggon, J. Förstner, A. Knorr, R. L. Sellin, D. Ouyang, and D. Bimberg, Self-induced transparency in InGaAs quantum-dot waveguides, *Appl. Phys. Lett.*, 83 (2003), pp. 3668–3670.

36. T. W. Berg and J. Mørk, Quantum dot amplifiers with high output power and low noise, *Appl. Phys. Lett.*, 82 (2003), pp. 3083–3085.
37. T. Erneux, E. A. Viktorov, P. Mandel, T. Piwonski, G. Huyet, and J. Houlihan, The fast recovery dynamics of a quantum dot semiconductor optical amplifier, *Appl. Phys. Lett.*, 94 (2009), p. 113501.
38. S. B. Kuntze, A. J. Zilkie, L. Pavel, and J. S. Aitchison, Nonlinear state-space model of semiconductor optical amplifiers with gain compression for system design and analysis, *J. Lightwave Technol.*, 26 (2008), pp. 2274–2281.
39. M. Radziunas and H. J. Wünsche, Dynamics of multi-section DFB semiconductor laser: Traveling wave and mode approximation models, *Proc. SPIE*, 4646 (2002), p. 27.
40. J. Javaloyes and S. Balle, Multimode dynamics in bidirectional laser cavities by folding space into time delay, *Opt. Express*, 20 (2012), pp. 8496–8502.
41. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vettering, *Numerical Recipes*, 3rd ed. Cambridge: Cambridge University Press (2007).
42. M. Radziunas, Numerical bifurcation analysis of the traveling wave model of multisection semiconductor lasers, *Physica D*, 213 (2006), pp. 98–112.
43. M. Rossetti, P. Bardella, and I. Montrosset, Time-domain travelling-wave model for quantum dot passively mode-locked lasers, *IEEE J. Quant. Electron.*, 47 (2011), p. 139.
44. M. Rossetti, P. Bardella, and I. Montrosset, Modeling passive mode-locking in quantum dot lasers: A comparison between a finite- difference traveling-wave model and a delayed differential equation approach, *IEEE J. Quant. Electron.*, 47 (2011), p. 569.
45. H. Y. Liu, M. Hopkinson, C. N. Harrison, M. J. Steer, R. Frith, I. R. Sellers, D. J. Mowbray, and M. S. Skolnick, Optimizing the growth of 1.3 μm InAs/InGaAs dots-in-a-well structure, *J. Appl. Phys.*, 93 (2003).
46. Q. T. Vu, H. Haug, and S. W. Koch, Relaxation and dephasing quantum kinetics for a quantum dot in an optically excited quantum well, *Phys. Rev. B*, 73 (2006), p. 205317.
47. W. Lei, M. Offer, A. Lorke, C. Notthoff, C. Meier, O. Wibbelhoff, and A. D. Wieck, Probing the band structure of InAs/GaAs quantum dots by capacitance-voltage and photoluminescence spectroscopy, *Appl. Phys. Lett.*, 92 (2008), p. 193111.
48. M. O. Scully, *Quantum Optics*. Cambridge: Cambridge University Press (1997).
49. W. W. Chow and S. W. Koch, *Semiconductor-Laser Fundamentals*, Berlin: Springer (1999).
50. H. Haug and S. W. Koch, *Quantum Theory of the Optical and Electronic Properties of Semiconductors*, 2nd ed. Singapore: World Scientific (2004).
51. W. C. W. Fang, C. G. Bethea, Y. K. Chen, and S. L. Chuang, Longitudinal spatial in homogeneities in high-power semiconductor lasers, *IEEE J. Sel. Top. Quantum Electron.*, 1 (1995), pp. 117–128.
52. J. N. Fehr, M. A. Dupertuis, T. P. Hessler, L. Kappei, D. Marti, F. Salleras, M. S. Nomura, B. Deveaud, J.-Y. Emery, and B. Dagens, Hot phonons and Auger related carrier heating in semiconductor optical amplifiers, *IEEE J. Quant. Electron.*, 38 (2002), p. 674.
53. P. Sotirelis and K. Hess, Electron capture in GaAs quantum wells, *Phys. Rev. B*, 49 (1994), pp. 7543–7547.
54. M. Vallone, Quantum well electron scattering rates through longitudinal optic-phonon dynamical screened interaction: An analytic approach, *J. Appl. Phys.*, 114 (2013), p. 053704.
55. R. Heitz, H. Born, F. Guffarth, O. Stier, A. Schliwa, A. Hoffmann, and D. Bimberg, Existence of a phonon bottleneck for excitons in quantum dots, *Phys. Rev. B*, 64 (2001), p. 241305(R).
56. K. Schuh, P. Gartner, and F. Jahnke, Combined influence of carrier-phonon and coulomb scattering on the quantum-dot population dynamics, *Phys. Rev. B*, 87 (2013), p. 035301.
57. G. D. Mahan, *Many-Particle Physics*. New York, NY: Plenum (1990).
58. A. Wojs, P. Hawrylak, S. Fafard, and L. Jacak, Electronic structure and magneto-optics of self-assembled quantum dots, *Phys. Rev. B*, 54 (1996), p. 5604.
59. H. Haug and S. W. Koch, Semiconductor laser theory with many-body effects, *Phys. Rev. A*, 39 (1989), p. 1887.

60. H. Haug and S. Schmitt-Rink, Electron theory of the optical properties of laser-excited semiconductors, *Prog. Quant. Electron.*, 9 (1984), p. 3.
61. W. W. Chow and S. W. Koch, Theory of semiconductor quantum-dot laser dynamics, *IEEE J. Quant. Electron.*, 41 (2005), pp. 495–505.
62. H. Haug and A. P. Jauho, *Quantum Kinetics in Transport and Optics of Semiconductors*. Berlin: Springer (1996).
63. E. Malic, K. J. Ahn, M. J. P. Bormann, P. Hövel, E. Schöll, A. Knorr, M. Kuntz, and D. Bimberg, Theory of relaxation oscillations in semiconductor quantum dot lasers, *Appl. Phys. Lett.*, 89 (2006), p. 101107.
64. T. R. Nielsen, P. Gartner, and F. Jahnke, Many-body theory of carrier capture and relaxation in semiconductor quantum-dot lasers, *Phys. Rev. B*, 69 (2004), p. 235314.
65. L. Bányai, Q. T. Vu, B. Mieck, and H. Haug, Ultrafast quantum kinetics of time-dependent RPA-screened Coulomb scattering, *Phys. Rev. Lett.*, 81 (1998), pp. 882–885.
66. R. Binder, D. Scott, A. E. Paul, M. Lindberg, K. Henneberger, and S. W. Koch, Carrier-carrier scattering and optical dephasing in highly excited semiconductors, *Phys. Rev. B*, 45 (1992), p. 1107.
67. F. X. Camescasse, A. Alexandrou, D. Hulin, L. Bányai, D. B. Tran Thoai, and H. Haug, Ultrafast electron redistribution through Coulomb scattering in undoped GaAs: Experiment and theory, *Phys. Rev. Lett.*, 77 (1996), pp. 5429–5432.
68. M. G. Kane, Nonequilibrium carrier-carrier scattering in two-dimensional carrier systems, *Phys. Rev. B*, 54 (1996), pp. 16345–16348.
69. D. B. Tran Thoai and H. Haug, Coulomb quantum kinetics in pulse-excited semiconductors, *Z. Phys. B*, 91 (1992), pp. 199–207.
70. E. Schöll, *Nonequilibrium Phase Transitions in Semiconductors*. Berlin: Springer (1987).
71. K. Lüdge and E. Schöll, Quantum-dot lasers—Desynchronized nonlinear dynamics of electrons and holes, *IEEE J. Quant. Electron.*, 45 (2009), pp. 1396–1403.
72. K. Lüdge, Modeling quantum dot based laser devices. In *Nonlinear Laser Dynamics–From Quantum Dots to Cryptography*, Chapter 1, pp. 3–34. Weinheim: Wiley-VCH (2012).
73. M. Rossetti, A. Fiore, G. Sek, C. Zinoni, and L. Li, Modeling the temperature characteristics of InAs/GaAs quantum dot lasers, *J. Appl. Phys.*, 106 (2009), p. 023105.
74. J. Urayama, T. B. Norris, H. Jiang, J. Singh, and P. Bhattacharya, Temperature-dependent carrier dynamics in self-assembled InGaAs quantum dots, *Appl. Phys. Lett.*, 80 (2002), pp. 2162–2164.
75. T. Erneux, E. A. Viktorov, and P. Mandel, Time scales and relaxation dynamics in quantum-dot lasers, *Phys. Rev. A*, 76 (2007), p. 023819.
76. P. Baveja, D. Maywar, A. Kaplan, and G. P. Agrawal, Self-phase modulation in semiconductor optical amplifiers: Impact of amplified spontaneous emission, *IEEE J. Quant. Electron.*, 46 (2010), pp. 1396–1403.
77. T. W. Berg and J. Mørk, Saturation and noise properties of quantum-dot optical amplifiers, *IEEE J. Quant. Electron.*, 40 (2004), pp. 1527–1539.
78. A. M. de Melo and K. Petermann, On the amplified spontaneous emission noise modeling of semiconductor optical amplifiers, *Opt. Commun.*, 281 (2008), pp. 4598–4605.
79. C. Meuer, J. Kim, M. Lämmlin, S. Liebich, A. Capua, G. Eisenstein, A. R. Kovsh, S. S. Mikhrin, I. L. Krestnikov, and D. Bimberg, Static gain saturation in quantum dot semiconductor optical amplifiers, *Opt. Express*, 16 (2008), p. 8269.
80. C. W. Gardiner, *Handbook of Stochastic Methods*. New York, NY: Springer (1985).
81. N. Wiener, Generalized harmonic analysis, *Acta Math.*, 55 (1930), pp. 117–258.
82. R. Bonk, T. Vallaitis, J. Guetlein, C. Meuer, H. Schmeckebier, D. Bimberg, C. Koos, and J. Leuthold, The input power dynamic range of a semiconductor optical amplifier and its relevance for access network applications, *IEEE Photon. J.*, 3 (2011), pp. 1039–1053.

83. S. Wilkinson, B. Lingnau, J. Korn, E. Schöll, and K. Lüdge, Influence of noise on the signal properties of quantum-dot semiconductor optical amplifiers, *IEEE J. Sel. Top. Quantum Electron.*, 19 (2013), p. 1900106.
84. G. E. P. Box and M. E. Muller, A note on the generation of random normal deviates, *Ann. Math. Statist.*, 29 (1958), pp. 610–611.
85. H. Shahid, D. Childs, B. J. Stevens, and R. A. Hogg, Negative differential gain due to many body effects in self-assembled quantum dot lasers, *Appl. Phys. Lett.*, 99 (2011), p. 061104.
86. E. Goldmann, M. Lorke, T. Frauenheim, and F. Jahnke, Negative differential gain in quantum dot systems: Interplay of structural properties and many-body effects, *Appl. Phys. Lett.*, 104 (2014), p. 242108.
87. S. W. Koch, T. Meier, F. Jahnke, and P. Thomas, Microscopic theory of optical dephasing in semiconductors, *Appl. Phys. A*, 71 (2000), pp. 511–517.
88. M. Lorke, T. R. Nielsen, J. Seebeck, P. Gartner, and F. Jahnke, Influence of carrier-carrier and carrier-phonon correlations on optical absorption and gain in quantum-dot systems, *Phys. Rev. B*, 73 (2006), p. 085324.
89. H. H. Nilsson, J. Z. Zhang, and I. Galbraith, Homogeneous broadening in quantum dots due to Auger scattering with wetting layer carriers, *Phys. Rev. B*, 72 (2005), p. 205331.
90. H. Tahara, Y. Ogawa, F. Minami, K. Akahane, and M. Sasaki, Long-time correlation in non-Markovian dephasing of an exciton-phonon system in InAs quantum dots, *Phys. Rev. Lett.*, 112 (2014), p. 147404.
91. H. Schmeckebier, B. Lingnau, S. König, K. Lüdge, C. Meuer, A. Zeghuzi, D. Arsenijevic, M. Stubenrauch, R. Bonk, C. Koos, C. Schubert, T. Pfeiffer, and D. Bimberg, Ultra-broadband bidirectional dual-band quantum-dot semiconductor optical amplifier. In *Optical Fiber Communication Conference and Exposition* (2015), p. Tu3I.7.
92. A. Girndt, A. Knorr, M. Hofmann, and S. W. Koch, Theoretical analysis of ultrafast pump-probe experiments in semiconductor amplifiers, *Appl. Phys. Lett.*, 66 (1995), p. 550.
93. E. P. Ippen and C. V. Shank, Techniques for measurement, edited by S. L. Shapiro, In *Ultrashort Light Pulses: Picosecond Techniques and Applications*, pp. 83–122. Berlin: Springer (1977).
94. J. M. A. Mecozzi and J. Mørk, Theory of heterodyne pump-probe experiments with femtosecond pulses, *J. Opt. Soc. Am. B*, 13 (1996), pp. 2437–2452.
95. P. Borri, S. Schneider, W. Langbein, and D. Bimberg, Ultrafast carrier dynamics in InGaAs quantum dot materials and devices, *J. Opt. A: Pure Appl. Opt.*, 8 (2006), p. S33.
96. T. Brandes, Coherent and collective quantum optical effects in mesoscopic systems, *Phys. Rep.*, 408 (2005), pp. 315–474.
97. H. Choi, V. M. Gkortsas, L. Diehl, D. Bour, S. Corzine, J. Zhu, G. Höfler, F. Capasso, F. X. Kärtner, and T. B. Norris, Ultrafast Rabi flopping and coherent pulse propagation in a quantum cascade laser, *Nature Photon.*, 4 (2010), pp. 706–710.

24

Wave Mixing Effects in Semiconductor Optical Amplifiers

Simeon N. Kaunga-Nyirenda
Michal Dlubek
Jun Jun Lim
Steve Bull
Andrew Phillips
Slawomir Sujecki
and
Eric Larkins

24.1 Introduction 747
24.2 Review of Nonlinear Phenomena in SOAs and Applications ... 748
Carrier Density Modulation • Carrier Heating • Spectral Hole Burning • Other Nonlinear Effects • Summary of Important Nonlinear Applications
24.3 Challenges in Modeling Nonlinear Effects 751
24.4 Improved Modeling of Nonlinear Phenomena in SOAs 752
Optical Model • Electrical Model • Light–Matter Interactions • Inclusion of Material Dispersion in the Time-Domain Model • Numerical Implementation
24.5 Case Study: FWM 757
CDM Characteristics • Experimental Procedure for Measuring CDM • Dependence on Operating Conditions
24.6 Summary and Conclusions 766

24.1 Introduction

The growing demand for telecommunications services continues to push optical transmission rates beyond 10 G-symbols/s and to use more efficient modulation formats (e.g., quadrature phase shift keying [QPSK], differential QPSK [DQPSK]) to cope with the ever-increasing demand for bandwidth. The growing demand for bandwidth is also driving a move to perform high bandwidth signal processing functions directly in the optical domain. All-optical signal processing generally requires nonlinear optical functionality. Semiconductor optical amplifiers (SOAs) are attractive for such functions because of their strong nonlinear optical response. Optical signal processing applications that employ SOAs include wavelength conversion, phase conjugation, optical switching, Boolean logic functions, all-optical recovery, and all-optical demultiplexing. SOAs offer large bandwidth, ease of integration, compactness, low power consumption, electrical pumping/control, and potentially low cost. Interest in SOAs continues to grow as photonic integrated circuit (PIC) integration densities continue to increase, with applications in signal leveling and signal regeneration or integration with electronics as optical interconnects for high-speed cross-chip communications. SOAs are also useful in electrically reconfigurable optical switching/routing networks (Williams et al., 2008). For these applications, however, SOAs may introduce nonlinear signal impairments that adversely affect system performance, and these impairments must be managed.

This chapter focuses on the numerical modeling and simulation of carrier plasma–induced nonlinear optical phenomena in SOAs. The chapter starts with a brief description of nonlinear processes in SOAs and their applications. We then discuss SOA modeling, paying particular attention to challenges encountered in the modeling and simulation of nonlinear effects. An efficient optical model based on a one-dimensional (1D) bidirectional traveling wave algorithm is then presented. The optical model is coupled to the semiconductor material through light–matter interactions. The electrical properties are simulated using a simplified 1D (unipolar) electrical model, which only takes into account the dynamics of the total carrier density in the active medium. The unipolar model assumes flat quasi-Fermi levels and charge neutrality in order to remove the additional computational complexity associated with the use of a self-consistent bipolar electrical model with band bending (due to charge imbalance in the well). Nevertheless, while a bipolar model will certainly improve the overall accuracy of the model, the focus of this work was the development and demonstration of an efficient model for the nonlinear optical response. For this purpose, the simplicity of the unipolar model is an advantage.

The material polarization is represented by an effective susceptibility that self-consistently includes the carrier-induced perturbations of the net optical gain and refractive index spectra, both of which are functions of the optical field. No particular functional dependence on the optical field is assumed. This allows a general representation of the light–matter interaction under quasi-equilibrium conditions, which is applicable for all optical intensity levels. The most widely used power series expansion of the polarization is usually implemented with a single nonlinear term, which is only applicable when the optical intensity is sufficiently weak. The gain and spontaneous emission spectra are calculated using the band structure details of the semiconductor. These spectra were appropriately broadened to account for the spectral broadening of the carrier energies in the bands due to carrier lifetime dependent dephasing and various scattering events. The refractive index spectrum is obtained from the broadened gain spectrum using a nonlinear Kramers–Kronig transformation.

The need to perform simulations in the time domain poses a particular challenge for the inclusion of the frequency-dependent complex material polarization (i.e., gain and refractive index dispersion) and spontaneous emission spectra. We use recursive digital filter functions to accurately represent the spectrally dependent material quantities in the time-domain simulation. This method captures the asymmetric nature of the spectral material response and can also be used to incorporate frequency responses obtained from experimental measurements as well as complicated responses that cannot easily be represented analytically. The use of spatiotemporal filter coefficients ensures that physical effects such as spatial-hole burning and spectral changes in the gain caused by band-filling and bandgap renormalization effects are correctly included. A further advantage of the recursive filter method is that the filter coefficients can be obtained offline from the main simulation engine and tabulated, improving the computational efficiency and minimizing the resources required by the simulations. The refractive index spectra are also represented more accurately than approaches using a constant linewidth enhancement factor (at best a small-signal representation of the Kramers–Kronig relation, which ignores the spectral dependence of the index changes).

The chapter concludes with a case study exploring carrier density modulation (CDM) caused by optical wave beating in SOAs. We use the time-domain model developed to obtain an understanding of wave mixing processes in SOAs and the dependence of CDM on different operating parameters. The simulation results are compared with experimental results from a commercial buried heterostructure (BH) 1550 nm InGaAs multiple quantum well (MQW)-SOA made by the Centre for Integrated Photonics (SOA-NL-OEC-1550).

24.2 Review of Nonlinear Phenomena in SOAs and Applications

Nonlinear optical processes occur when the complex dielectric response of the material depends on the strength of the optical field passing through the medium (Boyd, 2007). When an optical wave interacts with the semiconductor gain medium, there is a shift in the carriers of the gain medium, resulting in a field-induced dipole moment that becomes the source of an electromagnetic field at the microscopic level. This interaction between the optical field and the material is through the induced macroscopic material polarization (the polarization term in the wave equation; see Section 24.4.1). This macroscopic polarization is the sum of the distinct microscopic polarizations

$$P(\vec{r}, t) = \frac{2}{V_k} \sum_k d_{\text{cv}}(k) p(k, \vec{r}, t), \tag{24.1}$$

where V_k is the normalized volume, $d_{\text{cv}}(k)$ is the optical dipole matrix element, and p is the interband polarization. k, $\vec{r}$, and t are the electron wave vector, position vector, and time, respectively. The random high-speed temporal fluctuations of the microscopic dipole elements are eliminated by spatial averaging over the incoherent dipole ensemble. Only coherent (i.e., externally driven) fluctuations remain and their response can be described with a macroscopic polarization (Jackson, 1975). This induced macroscopic polarization, $\tilde{P}^{\pm}$, is a frequency-dependent quantity that is a function of the total optical field, $\tilde{E}$, interacting with the medium via the time-varying carrier distributions. In the frequency domain, the polarization is related to the total electric field by

$$\tilde{P}^{\pm}(\omega, N, \tilde{E}) = \varepsilon_0 \chi_{\text{eff}}(\omega, N, \tilde{E}) \tilde{E}^{\pm}, \tag{24.2}$$

where $\tilde{P}(\omega)$ and $\tilde{E}(\omega)$ are the Fourier transforms of $P(t)$ and $E(t)$, respectively, and ω is the angular frequency. N is the total carrier density in the active region. The effective susceptibility, χ_{eff} takes into account all processes, both linear and nonlinear. No particular functional dependence on the optical field is assumed in Equation 24.2. When the optical field is small, the polarization can be expanded into a power series of the optical field and the effective susceptibility is

$$\chi_{\text{eff}} = \sum_n \chi^{(n+1)} \tilde{E}^n \; : \; n = 0, 1, 2 \ldots \tag{24.3}$$

At low optical intensities, the optical properties of the material tend to behave in a linear manner. However, as the optical intensity increases, the strong nonlinear response, which results from the interaction between strong optical signals and the gain medium, can manifest itself as self- and cross-gain modulation (SGM/XGM), self-, and cross-phase modulation (SPM/XPM), and wave mixing. SOAs are optoelectronic devices whose behavior is governed by the material properties of the amplifying medium and the light propagation through it. These are strongly coupled and influenced by the operating conditions, e.g., input power levels, pumping current and form of the input. The nonlinear processes arise because of the dynamics of the excited carrier populations (and their distributions in real and momentum space) as the material interacts with the optical radiation. These nonlinear phenomena are detrimental to applications that rely on device linearity but can be exploited for optical signal processing and other applications requiring all-optical functionality.

CDM, carrier heating (CH), and spectral hole burning (SHB) are the most important processes giving rise to nonlinear optical phenomena in semiconductor devices.

24.2.1 Carrier Density Modulation

CDM is a coherent interband process that results from the dependence of the stimulated carrier recombination rate on the optical intensity. Stimulated emission and absorption processes, and to some extent the (amplified) spontaneous emission process, modify the total carrier densities in the gain medium, leading to changes in the total gain spectrum and the refractive index of the semiconductor material. A strong optical input signal will increase the stimulated emission rate and decrease the carrier density in the active material. This reduces the gain and perturbs the refractive index, which affect the propagation of both the perturbing signal and other co-/counterpropagating signals. This is the basis for effects such as SGM, XGM, SPM and XGM. When at least two waves distinguishable in frequency, polarization, or wave vector propagate through the SOA, wave beating can excite the material by modulating the carrier density. This modulation in turn creates dynamic gain and index gratings, which are responsible for creating new frequency components.

24.2.2 Carrier Heating

Carrier heating (CH) results from modulation of the carrier energy distributions in the valence and/or conduction bands (sub-bands for quantum wells [QWs] and wires). CH is caused by processes that increase the temperatures of the carrier distributions relative to the lattice temperature, including carrier injection, stimulated emission, and free carrier absorption (FCA). Carrier injection adds high energy ("hot") carriers to the carrier plasma, while stimulated emission removes "cold" carriers close to the band edge. FCA transfers energy from photons to carriers, moving them to higher energies within the bands. Such carriers quickly thermalize and transfer energy to other carriers by carrier–carrier scattering, thereby increasing the plasma energy density or temperature (Uskov et al., 1994).

24.2.3 Spectral Hole Burning

SHB is also a form of modulation of the energy distributions of the carriers within the valence and/or conduction bands (or sub-bands for QWs and wires). While CH perturbs the entire carrier plasma, the plasma can usually still be described with a quasi-equilibrium distribution with an increased plasma temperature. SHB, on the other hand, is an energetically localized perturbation of the plasma, which cannot be described with a quasi-equilibrium distribution. SHB occurs when carriers in the states supporting stimulated recombination are consumed faster than they are replenished by intraband scattering. This creates a spectral hole in the intraband carrier distribution relative to the corresponding thermal carrier distribution.

In contrast to CDM, CH, and SHB are intraband processes. SHB is due to changes in the carrier distribution away from Fermi–Dirac distributions and is responsible for gain saturation. (The gain compression coefficient, frequently used to describe gain saturation at high power, is a manifestation of SHB.) CH results in a Fermi–Dirac distribution that has a higher temperature than that of the lattice. CDM, CH, and SBH all create perturbations that result in the modulation of the gain/absorption and refractive index spectra, whose dynamics also create spatiotemporal index and gain gratings that can diffract waves propagating in the medium to generate new frequency components. Although the gain and index gratings caused by CDM are larger, CH and SHB contribute to the nonlinear optical response. CDM dominates up to frequencies limited by the carrier replenishment rate. CH and SHB are important at higher frequencies, but CDM remains important.

24.2.4 Other Nonlinear Effects

Other processes, e.g., optical Kerr effect, two-photon absorption (TPA), and second harmonic generation (SHG), also contribute to the nonlinear optical response, but their contribution is smaller compared to that of CDM and CH, and are neglected in this work. The inclusion of the Kerr effect in simulations is complicated by the lack of causality. TPA is a multiphoton process that is resonantly enhanced for photon energies of the order of half the bandgap of the semiconductor material. Instantaneous coherent processes like the optical Kerr effect, TPA and SHG are not discussed further here, but are important for some applications (e.g., all-optical gates) at high optical power densities. The reader is referred to Boyd (2007) or other texts on nonlinear optics for more information.

24.2.5 Summary of Important Nonlinear Applications

Optical nonlinearities cause signal impairments in linear applications, but have been exploited to great benefit in the field of functional photonics. SOA nonlinearities are used for all-optical signal processing to eliminate the need for optical–electrical–optical (OEO) conversion in communication networks. Processing tasks such as demultiplexing, clock recovery, regeneration, and routing can be undertaken entirely in the optical domain at higher speeds than are possible with electronics (Manning et al., 1997). One of the key applications of nonlinear processes in SOAs is wavelength conversion.

Wavelength converters increase the flexibility and performance of all-optical networks based on wavelength division multiplexing (WDM) (Kovačević and Acampora, 1995). All-optical wavelength converters are essential for coping with increasing data rates. They increase the network performance by reducing the effects of lightpath blocking due to the wavelength continuity constraints in optical packet and optical circuit switching networks. Of the different approaches to all-optical wavelength conversion, SOA-based wavelength converters have attracted the greatest interest because of their compactness, easy integration, and strong nonlinear response. Different approaches have been used for SOA-based wavelength conversion, including XGM, XPM, and four-wave mixing (FWM). FWM is the most flexible approach because it has a large extinction ratio and retains the phase information of the carrier signal, making it bitrate independent and intrinsically transparent to modulation format (a useful attribute in optical networks). FWM-based SOA wavelength converters also have the potential for processing ultra-high speed analog and digital signals using ultrafast processes like SHB and CH. Spectral inversion also occurs during the FWM-conversion process. FWM-converted signals with inverted spectral distribution have been used for dispersion compensation by positioning a nonlinear SOA at the center of the link (Yanhua et al., 2003).

Another important application of nonlinear processes in SOA is optical performance monitoring in optical networks. The increasing complexity of optical networks and the emergence of new types of traffic and data protocols have rendered digital signal monitoring in the electronic domain impractical (Dlubek, 2008). As a result, optical signal monitoring methods that measure the analogue parameters in the optical domain are required for transparent optical networks. Optical sampling using FWM in SOAs has been used to generate signal histograms for bit error rate (BER) estimation (Dlubek, 2008).

24.3 Challenges in Modeling Nonlinear Effects

SOA modeling is a sophisticated subject due to the number and complexity of processes that take place simultaneously in the device. The dynamic nature of optical nonlinearities requires the use of time-domain models. Time-domain simulations are needed to determine the carrier densities (i.e., drift-diffusion, generation/recombination effects, etc.), whereas the calculation of the complex index needs to be done in the frequency domain. Optical propagation can be done in the frequency domain, but spatiotemporal carrier density (and complex index) variations make time-domain models more tractable.

The challenge for time-domain simulations is the inclusion of the frequency-dependent complex material polarization (gain and refractive index dispersion). Several attempts have been made to include dispersion in time-domain simulations: single-pole Lorentzian filter (Pratt and Carroll, 2000), finite impulse filter (FIR) theory (Jones et al., 1995; Toptchiyski et al., 1999), a first-order (Durhuus et al., 1992) and a second-order (Das et al., 2000) Taylor series expansion about the gain peak wavelength with constant coefficients. The second-order Taylor series expansion and the Lorentzian profile are both inadequate for the asymmetric material gain spectrum. Furthermore, the use of constant filter coefficients is inadequate for representing carrier-induced shifts in the gain spectrum. Other approaches include the spectrum slicing technique (SSM) (Park et al., 2005) and effective Bloch equations (EBE). The SSM method can handle the asymmetric gain spectrum, but is not suited to modeling nonlinear interactions (Park et al., 2005). The EBE fits the gain and refractive index spectra using a series of Lorentzians. A single Lorentzian is used in Ning et al. (1997), which is only useful for signals with a very narrow spectrum. Accurate implementation for high bandwidth signals requires the inclusion of more Lorentzian terms, making the model computationally intensive.

24.4 Improved Modeling of Nonlinear Phenomena in SOAs

The theoretical model described here is applied to an MQW SOA fabricated with a BH (Figure 24.1). The operation of the SOA depends on the interaction of the optical waves with the material. Processes such as optical gain, optical absorption, and changes in the refractive index must be appropriately coupled to the optical model in order to develop numerical models that accurately represent the operation of the actual device. This section describes the different components of a coupled optoelectronic simulation tool for nonlinear SOAs.

24.4.1 Optical Model

The optical field propagation in the SOA active region is described by Maxwell's equations

$$\nabla \times \nabla \times \vec{\mathbf{E}} + \frac{n^2}{c^2}\frac{\partial^2 \vec{\mathbf{E}}}{\partial t^2} = -\frac{1}{\varepsilon_0 c^2}\left(\frac{\partial^2 \vec{\mathbf{P}}}{\partial t^2} + \frac{\partial \vec{\mathbf{j}}}{\partial t}\right), \tag{24.4}$$

where n is the refractive index of the unexcited material (background index), c is the speed of light in vacuum, ε_0 is the free-space permittivity, $\vec{\mathbf{P}}(x, y, z, t)$ is the macroscopic polarization of the medium, and $\vec{\mathbf{j}}(x, y, z, t)$ is the current density. It is assumed that the SOA is designed to support only the fundamental mode with a transverse field distribution $U(x, y)$, which is assumed to be wavelength independent and normalized such that $\int\int |U(x, y)|^2 \,\mathrm{d}x\mathrm{d}y = 1$. For simplicity, all effective and modal parameters (e.g., the transverse confinement factor) are assumed to have been obtained at this point, so that only the longitudinal spatial direction needs to be considered. The strong waveguiding of the BHs further confines the optical field in the transverse directions beyond the vertical confinement of the separate confinement heterostructure (SCH) (Agrawal and Dutta, 1986).

The variation of the injection current with respect to time is very small compared to that of the optical fields, so it is justifiable to assume

$$\frac{\partial j}{\partial t} = \frac{\partial(\sigma E)}{\partial t} \approx 0. \tag{24.5}$$

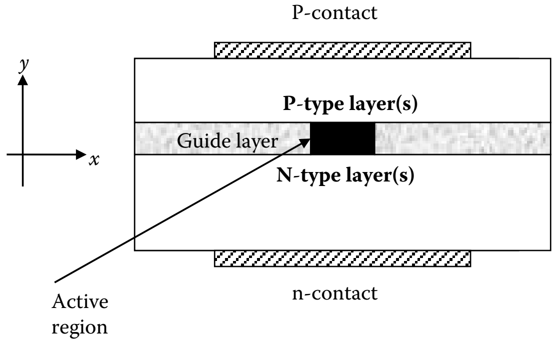


FIGURE 24.1 Schematic of a buried heterostructure (BH) semiconductor optical amplifier (SOA).

An alternative way of looking at this assumption is that optical fields are not usually generated directly by free currents and charges (Liu, 2005). Under these assumptions, the wave equations take the same form as those in a medium free of sources. The assumption in Equation 24.5 is plausible at optical frequencies, but this term may contribute in nonlinear optical beat-frequency devices, e.g., microwave and terahertz (THz) devices. This assumption is also not *a priori* valid for plasmonic devices (i.e., metals). For a straight and symmetric waveguide, the fields can be assumed to remain linearly polarized during propagation through the device. For a single polarization component, this implies that $\nabla \cdot \vec{\mathbf{E}} = \nabla \cdot \vec{\mathbf{P}} = 0$ and the left-hand side of Equation 24.1 is $\nabla \times \nabla \times \vec{\mathbf{E}} = \nabla \left(\nabla \cdot \vec{\mathbf{E}}\right) - \nabla^2 \vec{\mathbf{E}} = -\nabla^2 \vec{\mathbf{E}}$, so the scalar wave equation can be used.

Zero facet reflectivities are the target in traveling-wave SOAs. However, facets with antireflection (AR) coatings still exhibit residual reflectivities, which can result in the formation of an optical cavity with longitudinal resonances. Only real amplitude reflection coefficients are considered in this work. Although any wave reflected at the facet may initially be small, it will be amplified as it back-propagates along the waveguide such that it is not negligible compared to the forward field. Thus, reflected fields contribute to carrier depletion, gain compression, and refractive index change in the active medium. This affects the amplification, pulse shape, and phase of the output field. Reflected fields can also give rise to delayed replicas or echoes of the bits, which act as noise. Fields generated by ASE may also be nonnegligible compared to forward travelling fields. In the general case, the total optical field is comprised of both backward and forward propagating fields. Counterpropagating fields are also used deliberately in some applications relying on XGM and/or XPM.

Dropping the vector nature and using the slowly varying envelope approximation, the total optical field in the active region can be written as

$$E(x,y,z,t) = \frac{1}{2} U(x,y) \left\{ \psi^{+}(z,t) \exp\left[i(\omega t - kz)\right] + \psi^{-}(z,t) \exp\left[i(\omega t + kz)\right] + c.c. \right\}. \tag{24.6}$$

Taking into account that multiple signals may be input into the SOA, this can be rewritten as

$$E(x,y,z,t) = U(x,y) \sum_{q} \left\{ \psi_q^{+}(z,t) \exp\left[i\left(\omega_q t - k_q z\right)\right] + \psi_q^{-}(z,t) \exp\left[i\left(\omega_q t + k_q z\right)\right] + c.c. \right\}, \tag{24.7}$$

where $\lambda_q = \frac{c}{f_q}$ is the center wavelength of the qth signal and $\omega_q = 2\pi f_q$. From Equations 24.6 and 24.7, the total forward and backward slowly varying components are

$$\psi^{+}(z,t) = \sum_{q} \psi_q^{+}(z,t) \exp\left\{ i\left[\left(\omega_q - \omega_{\text{ref}}\right)t - \left(k_q - k_{\text{ref}}\right)z\right]\right\} + c.c. \tag{24.8}$$

$$\psi^{-}(z,t) = \sum_{q} \psi_q^{-}(z,t) \exp\left\{ i\left[\left(\omega_q - \omega_{\text{ref}}\right)t + \left(k_q - k_{\text{ref}}\right)z\right]\right\} + c.c., \tag{24.9}$$

where $k_{\text{ref}} = \frac{n\omega_{\text{ref}}}{c}$ is the optical propagation constant in the longitudinal direction and ω_{ref} is a reference frequency for the baseband transformation of all the signals propagating through the active medium. $\psi^{+}(z,t)$ and $\psi^{-}(z,t)$ are the forward and backward propagating complex envelopes of the optical field. The induced polarization can be written in a similar form as the field

$$\mathrm{P}(x,y,z,t) = \frac{1}{2} U(x,y) \left\{ P^{+}(z,t) \exp\left[i(\omega t - kz)\right] + P^{-}(z,t) \exp\left[i(\omega t + kz)\right] + c.c \right\}, \tag{24.10}$$

where $P^{+}(z,t)$ and $P^{-}(z,t)$ are the slowly varying forward and backward propagating complex polarizations. Substitution of Equations 24.7 and 24.10 into the scalar version of the wave Equation 24.4 and applying the slowly varying envelope approximation (SVEA) and paraxial approximation leads to the following propagation equations

$$\frac{n}{c}\frac{\partial \psi^{\pm}}{\partial t} \pm \frac{\partial \psi^{\pm}}{\partial z} = i\frac{\Gamma_{xy}k_{\mathrm{ref}}}{2\varepsilon_0 n^2}P^{\pm} + \psi^{\pm}_{\mathrm{sp}}(z,t), \tag{24.11}$$

which are subject to the boundary conditions

$$\psi^{+}(0,t) = \sqrt{R_1}\psi^{-}(0,t) + \sqrt{1-R_1}\psi_{\mathrm{in}1}(0,t) \tag{24.12}$$

$$\psi^{-}(L,t) = \sqrt{R_2}\psi^{+}(L,t) + \sqrt{1-R_2}\psi_{\mathrm{in}2}(L,t), \tag{24.13}$$

where $\psi_{\mathrm{in}1}(0,t)$ and $\psi_{\mathrm{in}2}(L,t)$ are optical inputs injected through the end facets. These facets have power reflectivities R_1 and R_{21}, respectively. The last term in Equation 24.11 is the local spontaneous noise contribution to the total propagating field. For simplicity, k_{ref} is referred to as k from here onward.

The transverse confinement factor, Γ_{xy}, in 24.11 is given by

$$\Gamma_{xy} = \frac{\int_{-w/2}^{w/2}\int_{-d/2}^{d/2}\left|U(x,y)\right|^2 \mathrm{d}y\mathrm{d}x}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|U(x,y)\right|^2 \mathrm{d}y\mathrm{d}x}, \tag{24.14}$$

where w and d are the active region width and thickness, respectively.

24.4.2 Electrical Model

The carrier dynamics are calculated using a simplified 1D carrier density model based on the evolution of the total carrier density in the active region (White et al., 1998)

$$\frac{\partial N(z,t)}{\partial t} = \frac{\eta J(z,t)}{ed} - \gamma_{\mathrm{nr}}(N)N - \frac{i\Gamma_{xy}}{4\hbar}\left\{\left(\psi^{+}\right)\left(P^{+}\right)^{*} + \left(\psi^{-}\right)\left(P^{-}\right)^{*} - c.c.\right\}, \tag{24.15}$$

where J is the total injected current density, e is the electron charge, γ_{nr} represents nonradiative recombination (generally carrier dependent), and η is the current injection efficiency. Carrier diffusion effects have been neglected in Equation 24.15 since the longitudinal electrical mesh spacing is larger than the carrier diffusion length. Thus, the total current density is only due to the external source from the electrodes. The current density from the external source is determined as in Dai et al. (1997)

$$J_{ext}(z,t) = \frac{V_{bias} - V_{jcn}(z,t)}{r_{\mathrm{s}}}, \tag{24.16}$$

where r_{s} is the series resistance (in units of $\Omega\mathrm{m}^{-2}$) of the p- and n-layers on either side of the active region and the p- and n-contact resistance. V_{jcn} is the voltage drop across the active layer (p–n junction) and relates to the difference between the two quasi-Fermi levels as

$$V_{jcn}(z,t) = \frac{E_{\mathrm{F}n}(z,t) - E_{\mathrm{F}p}(z,t)}{e}, \tag{24.17}$$

where $E_{\mathrm{F}n}$ and $E_{\mathrm{F}p}$ are the electron and hole quasi-Fermi levels, respectively. The quasi-Fermi levels are calculated within the gain model under the commonly used assumption of charge neutrality. While charge neutrality is a valid assumption in bulk material, it may not be a valid assumption for QWs, especially at high current pumping levels. A significant charge imbalance in the QW may strongly affect the optical properties of the surrounding optical guiding layers (SCH layers) and the entire waveguide due to the carriers injected into the region (Tolstikhin, 2000). Nevertheless, charge neutrality is assumed in order to use the unipolar model to simplify the analysis, which should be acceptable for moderate injection currents. $E_{\mathrm{F}n}$ and $E_{\mathrm{F}p}$ are estimated using the Joyce–Dixon approximation (Joyce and Dixon, 1977)

$$E_{\mathrm{F}n,p} = k_{\mathrm{B}}T\left[\ln\left(\frac{N}{N_{c,v}}\right) + \frac{1}{2}\left(\frac{N}{N_{c,v}}\right) + \frac{1}{24}\left(\frac{N}{N_{c,v}}\right)^2 - 0.0000347\left(\frac{N}{N_{c,v}}\right)^3\right], \tag{24.18}$$

where N_c and N_v are the effective densities of states for electrons and holes, k_{B} is the Boltzmann constant, and T is the temperature. The nonradiative processes considered in Equation 24.15 are Shockley–Read–Hall (SRH) and Auger recombination. The recombination rates are calculated under the charge neutrality condition.

24.4.3 Light–Matter Interactions

The interaction between light and the semiconductor gain medium is represented by the material (interband) polarization through an effective susceptibility (Equation 24.2). A general description of the polarization that is applicable for all levels of optical intensity is necessary. The effective susceptibility consists of the carrier-dependent changes in net gain and refractive index, both of which are functions of the optical field

$$\chi_{\mathrm{eff}}\left(\omega, N, \tilde{E}\right) = -\frac{1}{k}\left\{i\left[g_{net}\left(\omega, N, \tilde{E}\right) - \left(\frac{1-\Gamma_{xy}}{\Gamma_{xy}}\right)\alpha_{conf}\right] + k\Delta n\left(\omega, N, \tilde{E}\right)\right\}, \tag{24.19}$$

where α_{conf} represents the optical losses in the confinement regions. α_{conf} is needed because $1 - \Gamma_{xy}$ of the optical energy is located in the confinement layers outside the gain material. Equation 24.19 allows the induced nonlinear polarization to be described in a general fashion, without any assumptions based on its functional dependence on the optical field. This approach is more straightforward and physically robust than the common truncated series expansion of the polarization based on Equation 24.3. With this representation, the band structure is taken account of through the determination of the gain and refractive index changes.

To complete the analysis, a model is required for the material gain, spontaneous emission, and refractive index spectra. A parabolic model is used for the conduction band, while the valence band structure is calculated using a four-band **k·p**-model (Vurgaftman et al., 2001). The gain and spontaneous emission rates are obtained using Equations 24.20 and 24.21, respectively (Zory, 1993),

$$g(\hbar\omega) = \frac{1}{\hbar\omega}\frac{\bar{n}_{\mathrm{g}}\pi e^2\hbar}{n^2 c\varepsilon_0 m_0^2}\sum_{ic=iv}\left|M_{\mathrm{T}}\right|^2\rho_{\mathrm{red}}^{mD}\left(f_n - f_p\right) \tag{24.20}$$

$$R_{\mathrm{sp}}(\hbar\omega) = \frac{1}{\hbar\omega}\frac{\pi e^2\hbar}{n^2\varepsilon_0 m_0^2}\sum_{ic=iv}\left|M_{\mathrm{avg}}\right|^2\rho_{\mathrm{red}}^{mD}f_n\left(1 - f_p\right) \tag{24.21}$$

$$\left|M_{\mathrm{avg}}\right|^2 = \frac{1}{3}\sum_{\substack{\text{all 3}\\\text{polarizations}}}\left|M_{\mathrm{T}}\right|^2. \tag{24.22}$$

$f_{n,p}(E)$ is a Fermi–Dirac distribution function and gives the occupational probability of the electron of energy state E. ρ_{red}^{mD} is the reduced density of states, with $m = 2$ for QW material, and $m = 3$ for bulk.

$\left|M_T\right|^2$ is the transition matrix element and determines the probability and hence strength of interaction between the states at energy E_e and E_h in the presence of an electromagnetic field.

Carrier lifetime dependent dephasing and intraband scattering processes (carrier–carrier, carrier–phonon, etc.) broaden the range of energy states that may participate in the optical transitions. This energy broadening is taken into account phenomenologically by directly broadening the optical spectra (Zory, 1993). Instead of broadening the gain spectrum directly, the spontaneous emission spectrum was broadened using a hyperbolic secant lineshape function (Lim et al., 2007)

$$L\left(E_{eh} - \hbar\omega\right) = \frac{\tau_{in}}{\hbar\pi} \sec h \left(\frac{E_{eh} - \hbar\omega}{\hbar/\tau_{in}}\right), \tag{24.23}$$

where τ_{in} is the intraband relaxation time mainly due to scattering processes, and E_{eh} is the transition energy. The gain spectrum is obtained from the broadened spontaneous emission spectrum, $R_{sp,broad}(\hbar\omega)$, through

$$g(\hbar\omega) = \frac{1}{(\hbar\omega)^2} \frac{3\pi^2\hbar^3c^2}{2\bar{n}^2} R_{sp,broad}(\hbar\omega) \left[1 - \exp\left(\frac{\hbar\omega - \Delta E_F}{k_B T}\right)\right]. \tag{24.24}$$

The form of Equation 24.24 ensures that the gain passes through zero exactly at the quasi-Fermi level separation ΔE_F.

Gain compression due to SHB is included phenomenologically by introducing a constant gain compression factor, ε_{shb}

$$g(\hbar\omega) = \frac{g(N, T)}{1 + \varepsilon_{shb} S(\hbar\omega)}, \tag{24.25}$$

where $S(\hbar\omega)$ is the density of photons with energy $\hbar\omega$.

The carrier-induced changes in the refractive index spectrum are calculated from the changes in the gain/absorption spectrum through the nonlinear Kramers–Kronig relations (Hutchings et al., 1992) to ensure that the absorption and index changes are self-consistent. This is expressed as

$$\Delta n(\hbar\omega; \xi) = -\frac{c}{2\pi\omega} \int_{-\infty}^{\infty} \frac{\Delta g(\hbar\omega\prime; \xi)}{\omega' - \omega} d\omega', \tag{24.26}$$

where ξ is the external excitation, e.g., an electromagnetic field or temperature. This form is appropriate for different nonlinear processes, e.g., self-action effects (nonlinear refraction) and cascaded processes (Boyd, 2007). In this work, the carrier-induced changes in the gain/absorption spectrum only include band-filling effects and bandgap renormalization, but CH and SHB effects can also be included.

24.4.4 Inclusion of Material Dispersion in the Time-Domain Model

The challenge for time-domain simulations is the inclusion of the frequency-dependent complex material polarization (i.e., gain and refractive index dispersion). In this work, recursive digital filters are used to accurately represent the spectrally dependent material quantities in the time-domain model (Kaunga-Nyirenda et al., 2010). The most general linear digital filter takes a sequence of input points and produces a sequence of output points by the following difference equation (Smith, 2002):

$$y[n] = a_0 x[n] + a_1 x[n-1] + a_2 x[n-2] + \cdots + b_1 y[n-1] + b_2 y[n-2] + \ldots, \tag{24.27}$$

where n is the time index, and $x[]$ and $y[]$ represent the input and output, respectively. The a's and b's that define the filter are called recursive coefficients, with $b_0 \equiv 1$ corresponding to the sample being calculated.

Digital filtering allows the computation of long convolution integrals of two functions in the time domain for discretely sampled signals. The optimization of the filter coefficients involves the fast Fourier transform. Hence, it can be argued that the overhead incurred in obtaining the filter coefficients at every spatial point and time step is much larger than Fourier transforming the field and multiplying by the frequency dependent quantities directly, followed by inverse Fourier transforming back into the time domain. This computational overhead is, however, avoided by obtaining the filter coefficients offline from the main simulation and storing them in tables. For a given set of parameters (e.g., carrier density and temperature), the gain and refractive index changes can be calculated and the filter coefficients obtained for each case and stored in tables. The filter coefficients can then be retrieved during simulation with negligible additional computational burden. This approach is not possible with the direct energy–time transformation approach, which must be done during the simulation.

Figure 24.2 shows the fitted gain and refractive index spectra for an eight-pole filter after applying the filter to an impulse input and taking the Fourier transform of the resulting impulse response. A good fit is observed between the filter response and the desired response. The number of poles is dictated by the index spectrum, which has a sharp turning point and is less smooth than the gain spectrum. The number of poles for the filters for both spectra has to be the same for ease of implementation in the field update equation. Therefore, the number of poles chosen was one that gave a satisfactory fit for the index spectral profile, i.e., 8 (Figure 24.2).

This method accurately captures the asymmetric spectral response of the material polarization (Figure 24.2) and can be used to incorporate frequency responses from experimental measurements as well as complicated responses, not easily represented analytically. The use of spatiotemporal filter coefficients allows correct inclusion of spatial-hole burning and shifts of the gain spectrum due to band-filling and bandgap renormalization.

24.4.5 Numerical Implementation

The optical propagation (Equation 24.11) and carrier density (Equation 24.15) are coupled through the material Equations 24.19 through 24.21. These two equations are solved self-consistently by dividing the active region into small enough (shorter than the diffusion length) sections, such that quantities can be considered constant within each section. The field propagation equations are marched in time using the Lax differencing scheme, which allows use of a larger time step than is allowed by the explicit differencing scheme. The stability condition of the Lax differencing scheme is given by the Courant condition $(c\Delta t/n\Delta z) \leq 1$ (Press et al., 2002). The carrier density equation is integrated using a fourth-order Runga–Kutta method. All carrier-dependent quantities are estimated at the midpoint of each section, while the optical fields are evaluated at the section boundaries (Kaunga-Nyirenda et al., 2010). This ensures that the photons traveling to the right and to the left in a particular section both interact with the same carrier-related quantities (Wong and Carroll, 1987).

24.5 Case Study: FWM

FWM studies in nonlinear media have focused largely on the optical characteristics of mixing processes and experimental observations in the optical domain (Agrawal, 1988; Mukai and Saitoh, 1990; Uskov et al., 1994; Darwish et al., 1996; Gong et al., 2004), with emphasis mainly on the issue of conversion efficiency (Mukai and Saitoh, 1990). Very little attention has been paid to the characteristics of wave mixing in the electrical domain. Electrical measurements have been reported before for semiconductor laser diodes (SLDs) and a simplified model based on frequency modulation was presented in Nietzke et al. (1989). Although the operational characteristics of SLDs and SOAs are different, the material physics is the same, so that CDM and resulting wave mixing have their origins in the same physical processes. The carrier density is usually clamped to its threshold value in SLDs, limiting the modulation depth. On the other hand,

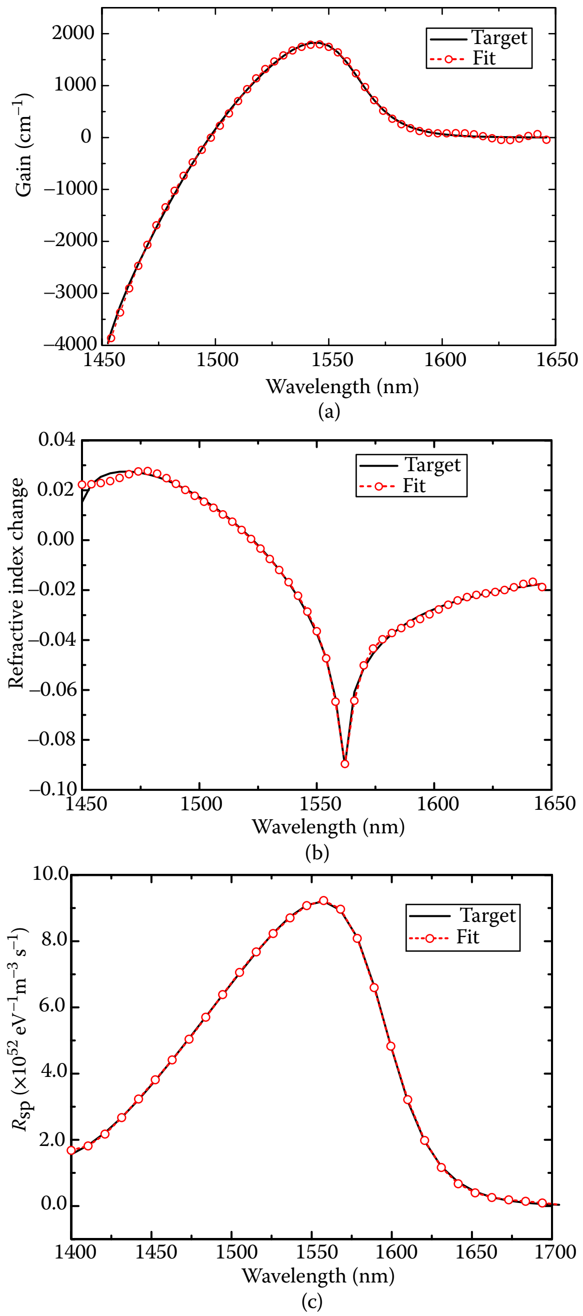


FIGURE 24.2 Fits for (a) gain spectrum, (b) refractive index change, and (c) spontaneous emission rate using recursive filters. The solid lines are the spectral quantity to be fitted, while the dotted lines are fits using the recursive filter. Number of poles for the filter is 8. Fits are obtained from the impulse responses of the filters.

large carrier density excursions are common in SOAs. In SLDs, the presence of strong optical feedback (i.e., large reflections at the facets) complicates the carrier density dynamics by introducing relaxation oscillations (Nietzke et al., 1989). A proper understanding of the characteristics of CDM is, therefore, likely to be stronger and clearer with SOAs (Dlubek et al., 2010). The model developed in Section 24.4 is used to study the characteristics of CDM over different operating conditions. No *a priori* assumptions are made about the existence or form of the CDM—it is obtained directly from the physics.

24.5.1 CDM Characteristics

The CDM characteristics in wave mixing in SOAs are studied using the model developed earlier. The input to the SOA is of the form

$$E_{\text{in}}(0,t) = \left\{ \psi_{\text{p}}(0,t) + \psi_{\text{s}}(0,t)\exp\left(-j\Omega t\right) \right\} \exp\left(-j\omega_{\text{ref}}t\right), \tag{24.28}$$

where $\Omega = 2\pi f_{\text{d}}$ and $f_{\text{d}} = \left|f_{\text{p}} - f_{\text{s}}\right|$ is the detuning frequency between the inputs (referred to as pump and probe). The pump and probe laser frequencies are $f_{\text{p}} = \omega_{\text{p}}/2\pi$ and $f_{\text{s}} = \omega_{\text{s}}/2\pi$, respectively. The time dependence of the pump and probe slowly varying fields is a general form and allows the use of modulated input signals. The input (Equation 24.28) enters through the boundary conditions (Equations 24.12 and 24.13).

The parameters used in the simulations are given in Table 24.1.

Indium gallium arsenide (InGaAs) was used as the active material for the quantum wells. The barrier material was strained InGaAs (−0.67% strain) (Kelly et al., 1996). This matches the material system for the device used in the experiments reported in Dlubek et al. (2010). The indium composition in the InGaAs active material was taken to be 0.62. This reproduced the device characteristics as reported in its data sheet and Kelly et al. (1996). This device has a nonlinear geometry, (Figure 24.3) so an effective length with a

TABLE 24.1 Simulation Parameters Used in the Study of Wave Mixing

Symbol	Description	Value (Units)
L	SOA length	1.1 mm (Lealman, 2009)
W	SOA width	1.3 μm (Lealman, 2009)
D	SOA thickness	0.1 μm (Lealman, 2009)
Δz	Longitudinal grid spacing	10 μm
Γ_{xy}	Optical confinement factor	0.18 (Lealman, 2009)
R_1	Input facet reflectivity	0.05% (Kelly et al., 1996)
R_2	Output facet reflectivity	0.05% (Kelly et al., 1996)
α_{int}	Internal loss	$2.0\ \text{cm}^{-1}$
A	Linear gain coefficient	$2.0 \times 10^{-20}\ \text{m}^2$
N_0	Carrier density at transparency	$1.7521 \times 10^{24}\ \text{m}^{-3}$
τ_{s}	Effective carrier lifetime	0.7 ns
α_{LWEF}	Linewidth enhancement factor	−3
A	Linear radiative recombination coefficient	$1.0 \times 10^{7}\ \text{s}^{-1}$
B	Bimolecular recombination coefficient	$5.6 \times 10^{-16}\ \text{m}^3\text{s}^{-1}$
C	Auger recombination coefficient	$3.0 \times 10^{-41}\ \text{m}^6\text{s}^{-1}$
$\bar{n}_{\text{g}}$	Group index of refraction for the mode	3.6
n	Index of refraction of active region	3.3
H	Current injection efficiency	0.7
ε_{shb}	Gain compression factor	$5.0 \times 10^{-17}\ \text{cm}^3$
τ_{in}	Intraband relaxation time	0.2 ps

Source: Agrawal, G.P., *Journal of the Optical Society of America B*, 5, 147–159, 1988.
Note: Some phenomenological parameters (carrier density at transparency, linear gain coefficient, effective carrier lifetime, linewidth enhancement factor, internal loss) are required as input to Agrawal's truncated series approximation (TSA) model.

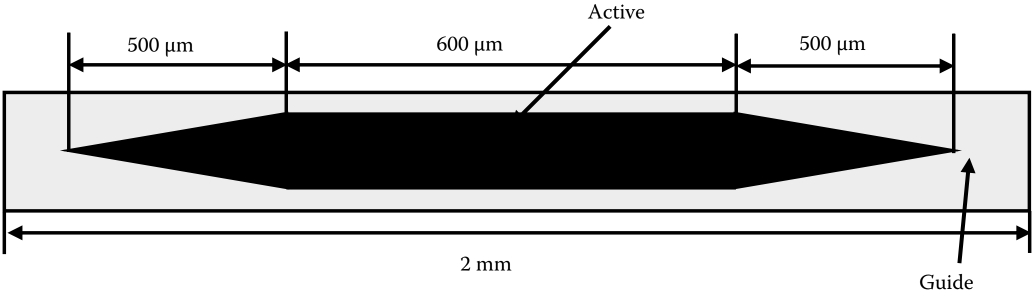


FIGURE 24.3 Sketch of the nonlinear geometry of a large spot semiconductor optical amplifier (SOA).

confinement factor equal to that of the untapered region was used in the simulations. All simulations were performed at 300 K. The carrier temperature was assumed to remain constant and equal to that of the lattice.

The CDM amplitude is obtained from the solution of the carrier density equation. A measurable quantity related to this is the variation in the injection current or voltage across the device, since the temporal fluctuations in the stimulated recombination rate result in temporal fluctuations in the current to resupply the carriers. The time series for the CDM amplitude is obtained at any spatial point along the longitudinal propagation direction as

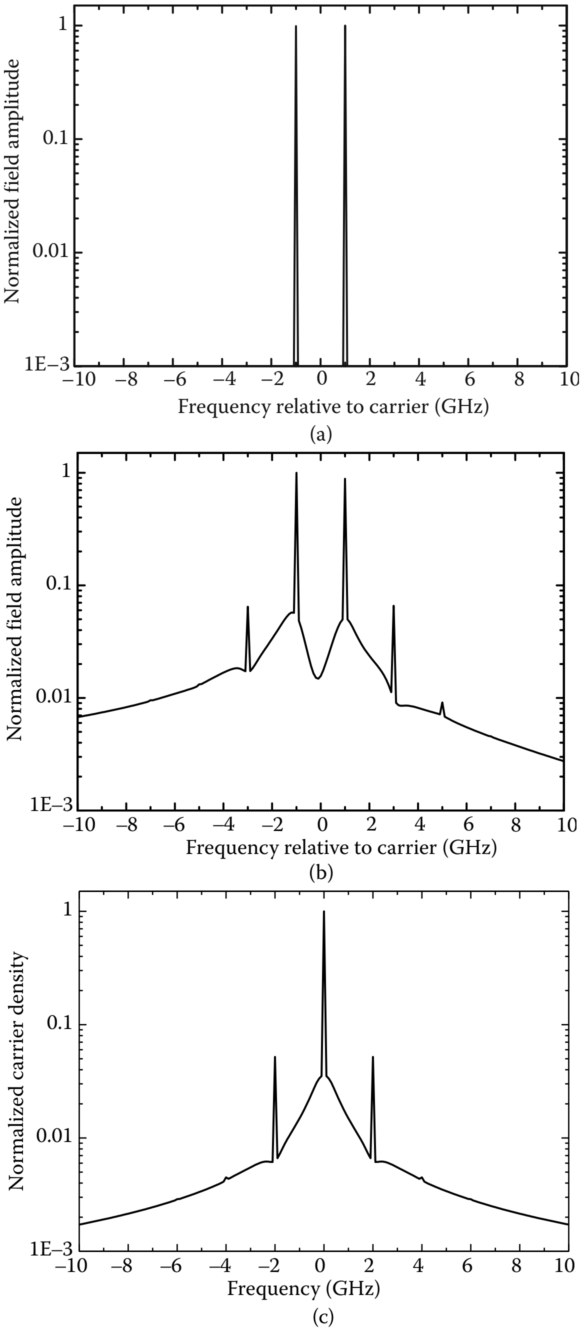
$$\Delta N(z,t) = N(z,t) - N_{\text{SS}}(z)\,, \tag{24.29}$$

where N_{SS} is the carrier density determined by the carrier injection, nonradiative recombination, and the amplified spontaneous emission (ASE) rates in the absence of external optical injection. The spectrum of the CDM amplitude is obtained by taking the discrete Fourier transform of Equation 24.29. Modulation of the carrier density results in modulation of the quasi-Fermi levels—and hence of the voltage drop across the active region (Equation 24.17). The spectrum of the injection current comes from DFT of Equation 24.16, while that of the voltage from Equation 24.17.

Simulated input and output optical spectra (from discrete Fourier transforms of the optical wave time series), and carrier density spectra (from discrete Fourier transform of Equation 24.29) are shown in Figure 24.4. The detuning of the two input signals was 2 GHz and the bias voltage was adjusted so that the current was 150 mA. The CW input optical fields had power levels of −0.7 and −0.8 dBm, respectively. The reference frequency was taken as the average of the pump and probe frequencies. An identical CDM spectra to Figure 24.4c was obtained when the reference frequency coincided with the pump frequency, confirming that the CDM spectra do not depend on the choice of the reference frequency used for the slowly varying envelope approximation. For the rest of the results presented below, the reference frequency coincided with the frequency of the pump. Figure 24.4 shows that the carrier density pulsates at the detuning frequency (beat frequency) of the pump and probe. The central peak in the carrier density spectrum is the static carrier density. The carrier population pulsations observed in Figure 24.4 are consistent with the experimental results (Nietzke et al., 1989).

The earlier simulations were repeated but with one of the optical input signals increased from −0.8 dBm to +4.7 dBm. Smaller secondary peaks are now also observed in Figure 24.4c at the multiples of the detuning frequency and these are enhanced as one of the input powers is increased (Figure 24.5).

The higher-order components (small secondary peaks at multiples of the detuning frequency) in Figure 24.5 are due to beating between the newly generated FWM products and the input beams or to higher-order harmonics in the CDM due to the nonlinearity of the carrier equation. Nonlinearities in the carrier equation are caused by the carrier density-dependent carrier lifetime and the saturation of the stimulated recombination rate. The peaks at the detuning frequency disappeared in the carrier density



(a)

(b)

(c)

FIGURE 24.4 Simulated spectra (a), input optical field (b), output optical field, and (c) carrier density. Parameters: detuning 2 GHz, pump power −0.7 dBm, probe power −0.8 dBm, bias current 150 mA.

spectrum when one of the input signals was set to zero, confirming that the CDM is caused by the interaction of the input waves. This is consistent with the experimental results reported in (Dlubek et al., 2010). In addition to the primary peak, a weaker peak was observed in the measured CDM spectrum at twice the detuning frequency. This second harmonic had a bandwidth larger than the primary harmonic by a factor

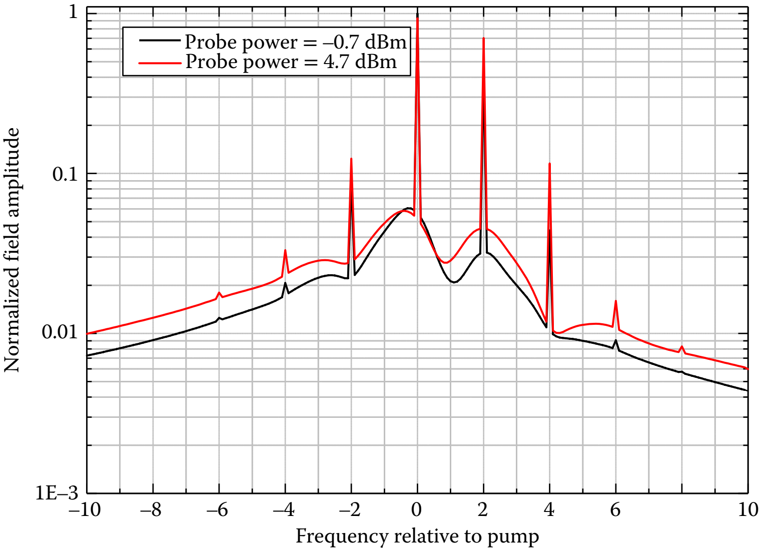


FIGURE 24.5 Optical spectra for the same parameters as in Figure 24.4 for two different probe powers. The pump signal is aligned to the reference signal, chosen around the gain peak.

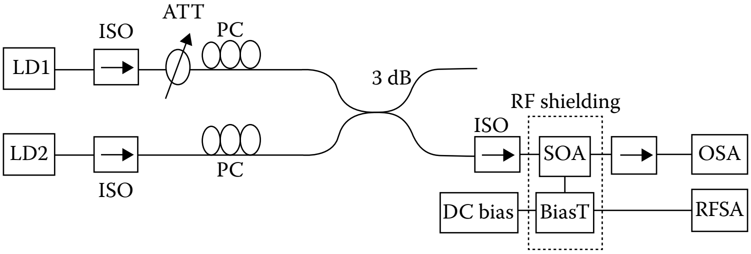


FIGURE 24.6 Carrier density modulation (CDM) measurement setup. ISO, isolator; ATT, variable optical attenuator; PC, polarization controller; 3 dB–50:50 coupler; OSA, optical spectrum analyzer; RFSA, radio frequency spectrum analyzer (From Dlubek, M.P. et al., *Optics Communications*, 283, 1481–1484, 2010).

of 1.5, indicating that it is the result of beating between the original waves and a new wave, which is itself broadened compared to the input waves.

24.5.2 Experimental Procedure for Measuring CDM

The experimental setup is shown in Figure 24.6 (Dlubek et al., 2010). The SOA used in the experiment was a nonlinear SOA from Centre of Integrated Photonics (CIP) (SOA-NL-OEC-1550 operating in 1550 nm region) with an active region consisting of InGaAs unstrained quantum wells and InGaAs barriers (Kelly et al., 1996). Two standard telecommunications diode lasers, LD1 and LD2, both a nominal wavelength of 1547.02 nm, were used as probe and pump, respectively. The frequency detuning between the two lasers was obtained by temperature tuning, so that it did not exceed 3 GHz (the bandwidth of the radio frequency (RF) spectrum analyzer). An optical spectrum analyzer was used to monitor the frequency detuning. A BiasT connector was used to uncouple the constant bias current to the SOA from the RF component from the device. Polarization controllers were used to adjust the polarization states of the interacting waves. Further experimental details are reported in Dlubek et al. (2010).

Before measuring the CDM spectrum, the reference beating spectrum of the two lasers (LD1 and LD2) was measured by heterodyning them on a standard 10 Gb/s photodetector. The 3 dB bandwidth of the heterodyne spectrum was ~40 MHz, suggesting that the laser linewidths were ~20 MHz. The spectral shape was approximately Lorentzian, as expected (Dlubek et al., 2010).

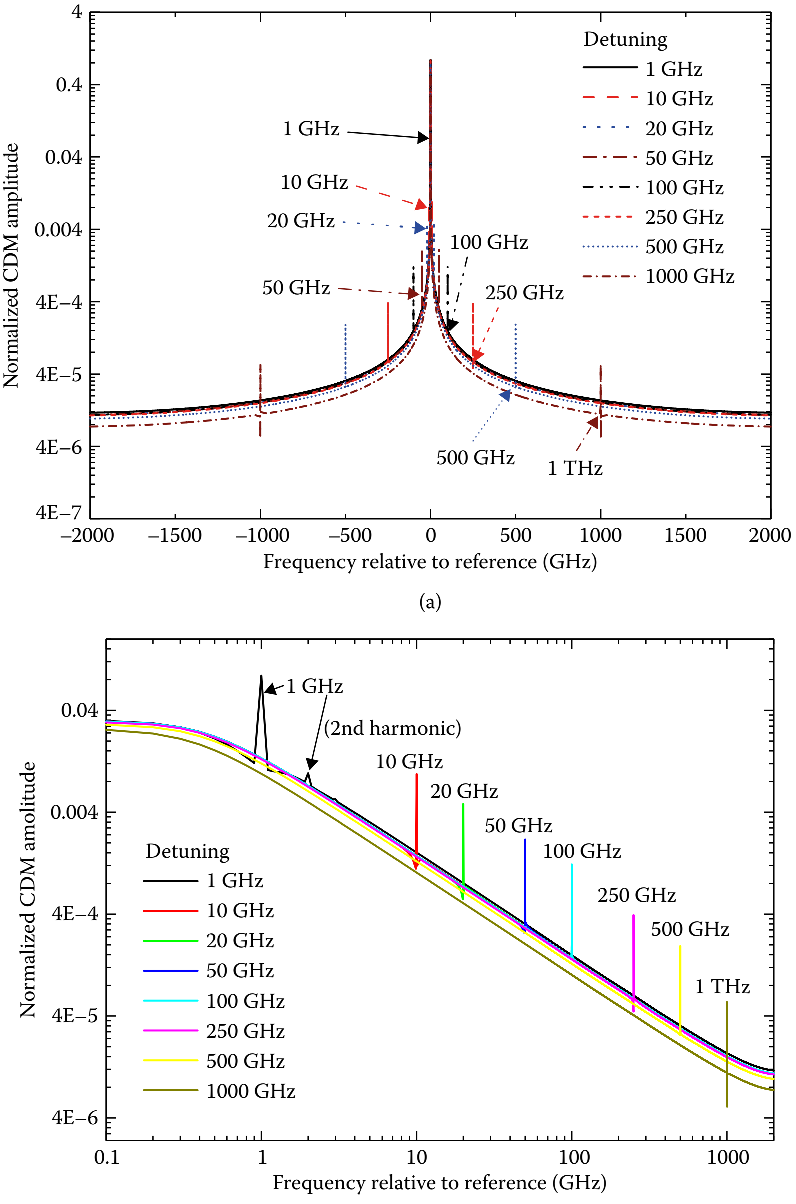
24.5.3 Dependence on Operating Conditions

The strength of interband effects depends on the rate of carrier replenishment into the active region. Thus, the CDM amplitude should decrease as the frequency detuning increases, since the carrier density cannot follow the fast oscillations of the stimulated emission rate. This was tested by simulating the FWM in the SOA for detuning frequencies up to 1 THz (by varying f_d in Equation 24.28), with all other parameters the same as in earlier simulations. Figure 24.7 shows the CDM spectra for different detunings. The peak at 2 GHz is a second harmonic for the 1 GHz detuning.

To obtain a clearer view of the frequency dependence of the CDM amplitude, the information in Figure 24.7 is extracted by obtaining the amplitude of the first harmonic (indicated against the detuning in Figure 24.7b) and plotting it against the detuning frequency (log scale), as shown in Figure 24.8. The CDM amplitude decreases sharply as the detuning frequency increases. The CDM amplitude at 1 THz is almost three orders of magnitude smaller than the value for a few GHz. The stimulated emission rate fluctuates at a much faster rate than the carriers injected into the active region. As a result, the carrier density cannot respond and the modulation efficiency decreases as the frequency detuning increases. Bream (2006) has also shown that the contribution of CDM to the refractive index perturbation is still significant even at frequencies well above the modulation bandwidths of the QW carrier distributions. As observed in Figure 24.8, although CDM continues to decrease with increasing detuning, it still contributes significantly to very fast nonlinear optical processes (even for detuning up to 1 THz) (Bream, 2006). Dynamic CH and SHB have not been taken into account in the model. It is interesting to note the similarity in the dependence of the CDM amplitude on detuning frequency and the dependence of amplitude of the conjugate signal on the detuning frequency observed in Figure 24.5a of Uskov et al. (1994). However, it is also worth noting that above 1 THz (~4 meV), acoustic phonon scattering should become relevant, while longitudinal optical (LO) phonon scattering should also be important above 9 THz. Thus, care is needed when extrapolating results to 1 THz and beyond, as other carrier scattering mechanisms can be resonantly excited.

Figure 24.9 shows the variation of the normalized CDM amplitude with bias current. The CDM in each case were normalized with respect to their maximum amplitude. As expected, the CDM amplitude increases with bias current. The detuning was adjusted to 1.23 GHz to match the experiments. In Agrawal's seminal paper on FWM (Agrawal, 1988), the carrier density rate equation is expressed as a power series involving the harmonics of the frequency detuning. The truncated series approximation (TSA1) curve in Figure 24.9 was calculated using Agrawal's model, where the power series is truncated at the first harmonic (Agrawal, 1988). The simulated and the experimental results both show that the CDM amplitude saturates as the bias current is increased further. Furthermore, a comparison of the experimental and simulated curves shows that the CDM approaches saturation at a faster rate in the experimental measurements than is observed in simulations. The CDM from measurements and truncated series assumption show the largest amplitude at 0 mA (experiment) and very close to 0 mA (TSA1).

The qualitative behavior is similar in both experiment and simulation. The model predicts the behavior of CDM better than the TSA1 in the gain region. At 0 mA, the rate of absorption is at its maximum and the net gain is equal to the absorption coefficient. The SOA behaves like an unbiased diode with an internal field and hence acts as a square-law detector. As the bias current increases from 0 mA toward transparency at around 35 mA, the rate of absorption decreases while the rate of emission increases. This results in an overall decrease in the net absorption and explains why the CDM amplitude decreases as the current increases from 0 mA toward transparency. Above transparency, the net gain increases. This leads to an increased rate of stimulated emission relative to absorption, which results in an increase in the CDM amplitude, saturating at about 150 mA. This is consistent with the measured gain-current dependence



(b)

FIGURE 24.7 Normalized CDM amplitude for different detunings (a) full spectra and (b) positive spectra (log scale).

shown in Figure 24.10. The discrepancy between simulations and experiments in the absorption region and low bias current above transparency (0–50 mA) in Figure 24.9 can be attributed to the use of the unipolar model and the effects of ASE. The *I*–*V* characteristics of a diode appear most nonlinear near zero bias. The unipolar model is based on the linear approximation of the current injection (Equations 24.16 and 24.17), which does not include the nonlinear *I*–*V* characteristics, hence in the deviation for currents between zero and transparency. Spontaneous emission, and the subsequent ASE, have not been included in the propagating field. The effects of ASE on carrier dynamics are more pronounced at low (electrical)

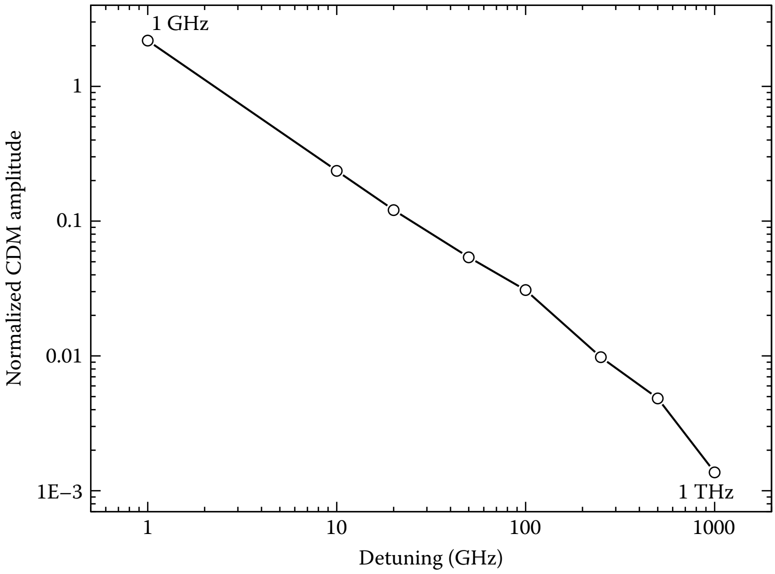


FIGURE 24.8 Normalized CDM amplitude at the first harmonic for different detunings (frequency in log scale).

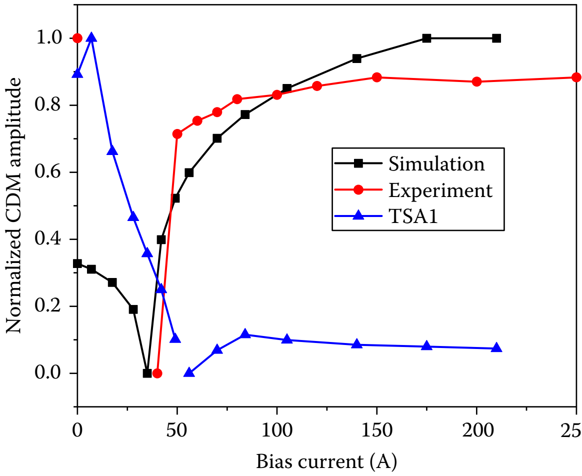


FIGURE 24.9 Normalized CDM amplitude at the first harmonic for different bias current, detuning 1.23 GHz, pump power −0.8 dBm, probe power −0.7 dBm. Truncated series approximation (TSA1) based on Agrawal's model (Agrawal, 1988), truncated at the first harmonic.

bias and optical input power level, since the rate of spontaneous emission can be comparable to the rate of stimulated emission.

The absence of experimental data points between 0 and approximately 50 mA in Figure 24.9 can be attributed to the low level of CDM, such that its amplitude was below the noise level. Although there is significant CDM in the absorption regime, it is not useful for many applications (apart from heterodyne detection, as discussed in Section 24.5.2) and both experiments and simulations confirm that there are no detectable/measurable optical mixing products.

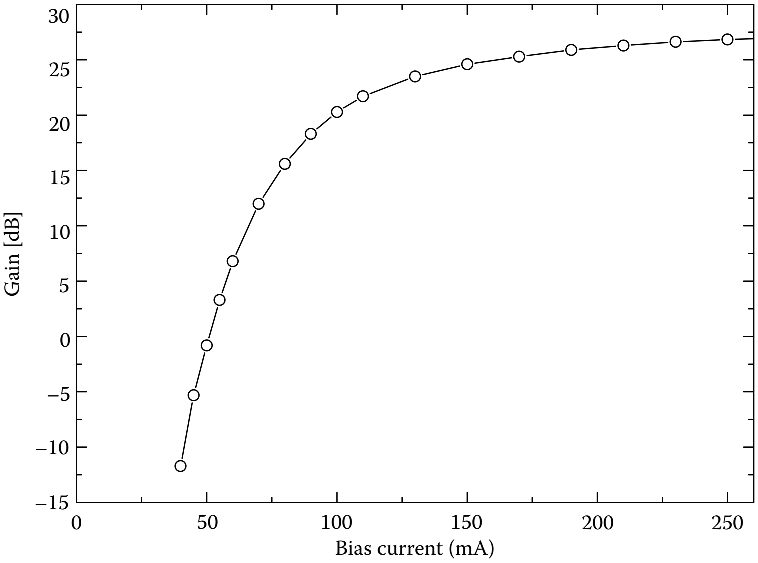


FIGURE 24.10 Measured small signal gain. Dependence on bias current for input power −28 dBm.

The effectiveness of CDM is expected to increase with input power, as the interaction between the different signals should be enhanced. As the optical power intensity increases in the active region, carrier depletion, and gain saturation effects should ultimately prevent any further increase in the CDM amplitude. (Signal distortion due to clipping as the carrier density approaches its transparency value should also create more signal components in the frequency spectrum.) Figure 24.11a and b show the simulated and measured CDM amplitude dependence on pump power for two probe powers (−0.8 dBm and +4.7 dBm). The detuning between the pump and probe was 1.23 GHz and the bias voltage was adjusted to give a bias current of 150 mA. The CDM amplitude increased with input power up to a point. The figures also reveal that the largest CDM is produced when the pump and probe signals have approximately equal power. This is a key result that cannot be determined from previous works, which rely on the assumption that the pump power is much larger than the probe power.

There is good qualitative agreement between simulation and experiment. However, there is some discrepancy in the amplitude of the CDM obtained using simulations and experiment, as shown in Figures 24.9 and 24.11. Apart from the limitations of the unipolar electrical and 1D traveling wave optical models, some of these differences are due to errors in the material parameters of the SOA, as these were not provided by the manufacturer. This discrepancy may also result from a lack of knowledge of the device's parasitic impedances and layer resistivities, which will also affect the RF signal measured at the device terminals (after the BiasT in Figure 24.6).

The normalized CDM amplitude is generally higher for the lower probe power (−0.8 dBm), as shown in Figure 24.12, where the dependence of the simulated CDM amplitude on the pump power is plotted for two probe powers (−0.8 dBm, +4.7 dBm). The other parameters are the same as those used in the simulations for Figure 24.11. When one beam is much weaker than the other, the CDM amplitude decreases as the beating is reduced due to carrier depletion and gain saturation by the strong beam.

24.6 Summary and Conclusions

We have developed a suitable time-domain model to investigate dynamic nonlinear optical effects in SOAs. The model takes into account the most important interband processes and minimizes the number of phenomenological parameters. This simulation tool was used to show how wave mixing processes manifest

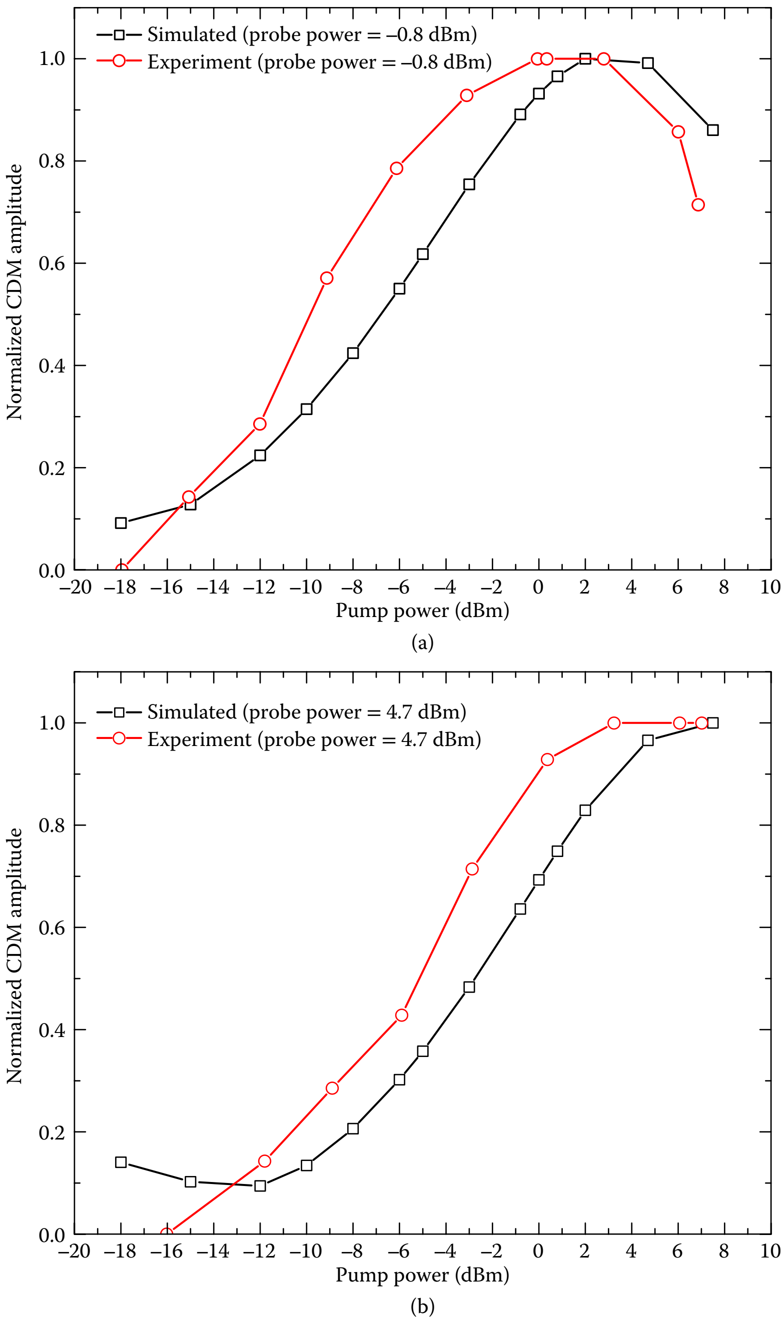


FIGURE 24.11 Normalized CDM amplitude dependence on pump power for different probe power levels: (a) probe power −0.8 dBm and (b) probe power +4.7 dBm. Detuning 1.23 GHz, bias current 150 mA.

themselves in SOAs, and discuss their impact on existing and potential applications. This is key to understanding and modeling nonlinear optical signal impairments. Furthermore, the knowledge obtained from these studies can be used in the design of new devices and/or for optimizing existing ones that rely on similar nonlinear optical effects, e.g., high-density PICs. The nonlinear phenomena are also accessible

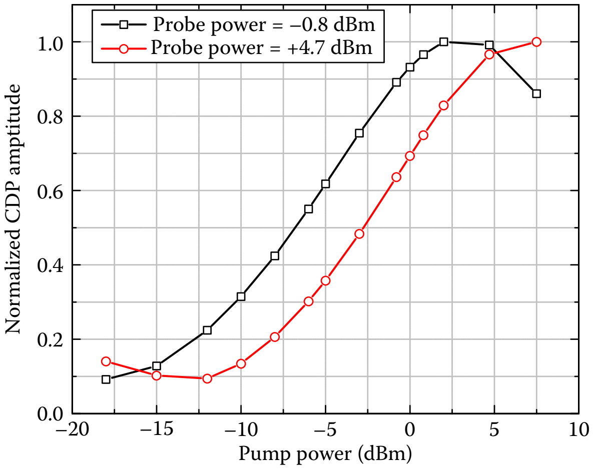


FIGURE 24.12 Simulated CDM amplitude (normalized) for two probe levels as in Figure 24.11a and b.

through electrical measurements, as well as through FWM and SHG measurements, providing additional opportunities for the validation of the simulation tool.

Other key outcomes of this work are the general description of the nonlinear polarization and the demonstration of the inclusion of nonlinear optical effects using the recursive filter functions. These can be used in more advanced tools, such as those with higher dimensionality (2D or 3D), those with a dynamic bipolar electrical model, and those with more accurate time domain-optical models (e.g., which properly include the spectral effects in the propagation and waveguiding). Finally, the general description of the nonlinear polarization means this approach can be readily adapted to include other contributions to the nonlinear polarization.

Acknowledgments

S.N. Kaunga-Nyirenda would like to thank the Commonwealth Scholarship Commission (United Kingdom) for financial support. This work was also supported in part by the Higher Education Funding Council of England (HEFCE) with a HERMES Fellowship and the UK Engineering and Physical Sciences Research Council (EPRSC) under grant GR/T09309/01. The authors are also grateful to the CIP for useful information and discussion about the SOA. The authors would also like to thank Prof. M. Wale (Oclaro) and Prof. K. Williams (TU Eindhoven) for useful discussions on the simulation requirements for active devices in PICs.

References

Agrawal, G.P. (1988) Population pulsations and nondegenerate four-wave mixing in semiconductor lasers and amplifiers. *Journal of the Optical Society of America B*, **5**, 147–159.

Agrawal, G.P. and Dutta, N.K. (1986) *Long-Wavelength Semiconductor Lasers*. New York, NY: Van Nostrand Reinhold.

Boyd, R.W. (2007) *Nonlinear Optics*. Cambridge, MA: Academic Press.

Bream, P.J. (2006) *Nonequilibrium Carrier Dynamics and Gain in Semiconductor Quantum Wells*. Nottingham: School of Electrical and Electronic Engineering.

Dai, Z., Michalzik, R., Unger, P., and Ebeling, K.J. (1997) Numerical simulation of broad-area high-power semiconductor laser amplifiers. *IEEE Journal of Quantum Electronics*, **33**, 2240–2254.

Darwish, A.M., Ippen, E.P., Le, H.Q., Donnelly, J.P., and Groves, S.H. (1996) Optimization of four-wave conversion efficiency in the presence of nonlinear loss. *Applied Physics Letters*, **69**, 737–739.

Das, N.K., Yamayoshi, Y., and Kawagushi, H. (2000) Analysis of basic four-wave mixing characteristics in a semiconductor optical amplifier by finite-difference beam propagation method. *IEEE Journal of Quantum Electronics*, **36**, 1184–1192.

Dlubek, M.P. (2008) *Optical Performance Monitoring for Amplified Spontaneous Emission Related Issues in Transparent Optical Networks*. Nottingham: School of Electrical and Electronic Engineering.

Dlubek, M.P., Kaunga-Nyirenda, S.N., Phillips, A.J., Sujecki, S., Harrison, I., and Larkins, E.C. (2010) Experimental verification of the existence of optically induced carrier pulsations in SOAs. *Optics Communications*, **283**, 1481–1484.

Durhuus, T., Mikkelsenm, B., and Stubkjaer, K.E. (1992) Detailed dynamic model for semiconductor optical amplifiers and their crosstalk and intermodulation distortion. *Journal of Lightwave Technology*, **10**, 1056–1064.

Gong, P.-M., Hsieh, J.-T., Lee, S.-L., and Wu, I. (2004) Theoretical analysis of wavelength conversion based on four-wave mixing in light-holding SOAs. *IEEE Journal of Quantum Electronics*, **40**, 31–40.

Hutchings, D.C., Sheik-bahae, M., Hagan, D.J., and Stryland, E.W. (1992) Kramers-Kronig relations in nonlinear optics. *Optical and Quantum Electronics*, **24**, 1–30.

Jackson, J.D. (1975) *Classical Electrodynamics*. New York, NY: John Wiley & Sons.

Jones, D.J., Zhang, L.M., Carroll, J.E., and Marcenac, D.D. (1995) Dynamics of monolithic passively mode-locked semiconductor lasers. *IEEE Journal of Quantum Electronics*, **31**, 1051–1058.

Joyce, W.B., and Dixon, R.W. (1977) Analytic approximations for the Fermi energy of an ideal Fermi gas. *Applied Physics Letters*, **31**, 354–356.

Kaunga-Nyirenda, S.N., Dlubek, M.P., Phillips, A.J., Lim, J.J., Larkins, E.C., and Sujecki, S. (2010) Theoretical investigation of the role of optically induced carrier pulsations in wave mixing in semiconductor optical amplifiers. *Journal of Optical Society of America B*, **27**, 168–178.

Kelly, A.E., Lealman, I.F., Rivers, L.J., Perrin, S.D., and Silver, M. (1996) Polarisation insensitive, 25 dB gain semiconductor laser amplifier without antireflection coatings. *Electronics Letters*, **32**, 1835–1836.

Kovačević, M., and Acampora, A. (1995) Benefits of wavelength translation in all-optical clear-channel networks. *IEEE Journal of Selected Areas in Communications*, **14**, 868–880.

Lealman, I.F. (2009) Centre for Integrated Photonics (CIP), B55, Adastral Park (Private Communication).

Lim, J.J., MacKenzie, R., Sujecki, S., Sadeghi, M., Wang, S.M., Wei, Y.Q., Gustavsson, J.S., Larsson, A., Melanen, P., Sipilä, P., Uusimaa, P., George, A.A., Smowton, P.M., and Larkins, E.C. (2007) Simulation of double quantum well GaInNAs laser diodes. *IET Optoelectronics*, **1**, 259–265.

Lim, J.J., Sujecki, S., Lang, L., Zhang, Z., Paboeuf, D., Pauliat, G., Lucas-Leclin, G., Georges, P., MacKenzie, R.C.I., Bream, P., Bull, S., Hasler, K.-H., Sumpf, B., Wenzel, H., Erbert, G., Thestrup, B., Petersen, P.M., Michel, N., Krakowski, M., and Larkins, E.C. (2009) Design and simulation of next-generation high-power, high-brightness laser diodes. *IEEE Journal of Selected Topics in Quantum Electronics*, **15**, 993–1008.

Liu, J. (2005) *Photonic Devices*. Cambridge: Cambridge University Press.

Manning, R.J., Ellis, A.D., Poustie, A.J., and Blow, K.J. (1997) Semiconductor laser amplifiers for ultrafast all-optical signal processing. *Journal of the Optical Society of America B*, **14**, 3204–3216.

Mukai, T., and Saitoh, T. (1990) Detuning characterisitcs and conversion efficiency of nearly degenerate four-wave mixing in a 1.5-μm travelling-wave semiconductor laser amplifier. *IEEE Journal of Quantum Electronics*, **26**, 865–875.

Nietzke, R., Panknin, P., Elsasser, W., and Gobel, E.O. (1989) Four-wave mixing in GaAs/AlGaAs semiconductor lasers. *IEEE Journal of Quantum Electronics*, **25**, 1399–1406.

Ning, C.Z., Indik, R.A., and Moloney, J.V. (1997) Effective Bloch equations for semiconductor lasers and amplifiers. *IEEE Journal of Quantum Electronics*, **33**, 1543–1550.

Park, J., Li, X., and Huang, W.P. (2005) Comparative study of mixed frequency-time-domain models of semiconductor optical amplifiers. *IEE Proceedings in Optoelectronics*, **152**, 151–159.

Pratt, E.M., and Carroll, J.E. (2000) Gain modelling and particle balance in semiconductor lasers. *IEE Proceedings in Optoelectronics*, **147**, 77–82.

Press, W.H., Teukolsky, S.A., Vetterling, W.T., and Flannery, B.P. (2002) *Numerical Recipes in C++: The Art of Scientific Computing*. Cambridge: Cambridge University Press.

Smith, S.W. (2002) *Digital Signal Processing: A Practical Guide for Engineers and Scientists*. Poway, CA: California Technical Publishing.

Tolstikhin, V.I. (2000) Carrier charge imbalance and optical properties of separate confinement heterostructure quantum well lasers. *Journal of Applied Physics*, **87**, 7342–7348.

Toptchiyski, G., Kindt, S., Petermann, K., Hillinger, E., Diez, S., and Weber, H.G. (1999) Time-domain modeling of semiconductor optical amplifiers for OTDM applications. *Journal of Lightwave Technology*, **17**, 2577–2583.

Uskov, A., Mork, J., and Mark, J. (1994) Wave mixing in semiconductor laser amplifiers due to carrier heating and spectral-hole burning. *IEEE Journal of Quantum Electronics*, **30**, 1769–1781.

Vurgaftman, I., Meyer, J.R., and Ram-Mohan, L.R. (2001) Band parameters for III–V compound semiconductors and their alloys. *Applied Physics Review*, **89**, 5815–5875.

White, J.K., Moloney, J.V., Gavrielides, A., Kovanis, V., Hohl, A., and Kalmus, R. (1998) Multilongitudinal-mode dynamics in a semiconductor laser subject to optical injection. *IEEE Journal of Quantum Electronics*, **34**, 1469–1473.

Williams, K.A., Aw, E.T., Wang, H., Penty, R.V., and White, I.H. (2008) Physical layer modelling of semiconductor optical amplifier based terabit/second switch fabrics. *International Conference of Numerical Simulation of Optoelectronic Devices*, Nottingham, United Kingdom.

Wong, Y.L., and Carroll, J.E. (1987) A Travelling-wave rate equation analysis for semiconductor lasers. *Solid-State Electronics*, **30**, 13–19.

Yanhua, H., Bandyopadhyay, S., Spencer, P.S., and Shore, K.A. (2003) Polarization-independent optical spectral inversion without frequency shift using a single semiconductor optical amplifier. *IEEE Journal of Quantum Electronics*, **39**, 1123–1128.

Zory, P.S., Jr. (ed.) (1993) *Quantum Well Lasers*. Cambridge, MA: Academic Press.

25

Semiconductor Optical Amplifier Dynamics and Pattern Effects

Zoe V. Rizou
and
Kyriakos E. Zoiros

25.1 Introduction 771
25.2 SOA Dynamics Background 772
25.3 SOA Model Formulation 776
25.4 SOA Model Simplification 781
25.5 SOA Optical Gain Modulation 784
25.6 Pattern Effect I: Direct Optical Amplification 787
25.7 SOA Electrical Gain Modulation 791
25.8 Pattern Effect II: Direct Current Modulation 792

25.1 Introduction

Since their advent in the second half of the 1980s (O'Mahony, 1988), semiconductor optical amplifiers (SOAs) have evolved technologically to the point that they have become key elements for the development of optical communications circuits, systems, and networks. Due to their attractive features of low power consumption, compactness, broad gain bandwidth, and ability for integration with affordable cost, the multifunctional potential of SOAs has been exploited in data amplification (Zimmerman and Spiekman, 2004) and processing (Mørk et al., 2003) in the optical domain and, more recently, for data encoding as well (Udvary and Berceli, 2010).

Motivated by the widespread employment of SOAs and the concomitant need to assist their design and support the implementation of the diverse applications they are destined to serve, the purpose of this chapter is to present in a concise and comprehensible manner the basic processes which govern the dynamic behavior of SOAs. Furthermore, it aims at describing how these processes can theoretically be taken into account in order to model their impact on the SOA response. It is concerned with the significant phenomenon of the pattern effect which manifests when the SOA gain is modulated, either optically by a single data pulse train or electrically by digital information superimposed on the SOA current. More specifically, the chapter topics which are addressed with regard to the SOA dynamics are outlined as follows:

Section 25.2: SOA Dynamics Background—Types, physical origin, conditions for manifestation, and qualitative impact on SOA response.

Section 25.3: SOA Model Formulation—SOA gain dynamics modeling theoretical formulation and numerical/closed form solutions.

Section 25.4: SOA Model Simplification—Ways of simplification and implications.

Section 25.5: SOA Optical Gain Modulation—Characterization and implications.

Section 25.6: Pattern Effect I—Direct Optical Amplification: Performance limitations and methods for mitigation.

Section 25.7: SOA Electrical Gain Modulation—Characterization and implications.
Section 25.8: Pattern Effect II—Direct Current Modulation: Performance limitations and methods for mitigation.

25.2 SOA Dynamics Background

Figure 25.1a depicts the standard configuration of an SOA (Mørk et al., 2003). Light is coupled via one of the facets into an intrinsic region of semiconductor material, which is sandwiched between a hole-dominant, that is, p-doped, and electron-dominant, that is, n-doped, cladding layer of higher bandgap energy and slightly lower refractive index. This structure allows confinement of a large number of free carriers in a small active volume, which hence exhibit a large density. By forward biasing the heterojunction through the injection of an electric current being below the lasing threshold point, the population of the carriers can be sufficiently inverted so that optical gain is established at a rate that prevails absorption. Then when signal photons of suitable energy travel through the active area, they stimulate radiative recombinations of electrons and holes and are coherently (i.e., with the same polarization, frequency, and phase) amplified as they propagate along the formed waveguide and exit from the other side of the device. In order to avoid unwanted back reflections and oscillations inside the cavity, the reflectivity of the SOA facets is reduced to very low values of less than 1×10^{-5} using antireflection coatings, tilted waveguides and tapered waveguides (Connelly, 2002). These technological means allow us to achieve a single-pass controllable amplification as high as 30 dB and with ripples-free parabolic-like spectrum profile, which can extend from 1300 to 1600 nm by changing the chemical composition of III–V group semiconductor materials. In particular, in this chapter we consider InGaAsP/InP-based SOAs, which operate in the 1550 nm telecommunications window where the attenuation of signals transmitted via optical fibers is lowest.

When a lightwave beam of optical frequency ν is injected into an SOA with a photon energy, hν, being larger than the bandgap energy, $E_g = E_c$-E_v, which, as shown in Figure 25.1b, separates the conduction and valence bands having energy at their bottom and top E_c and E_v, respectively, it triggers through band-to-band transitions stimulated emission and depletes the pumped carriers (Diez, 2000; Vacondio, 2011).

The reduction in the density of the excited carriers has then two consequences. First, the produced gain is reduced up to the point where it becomes saturated. Second, a set of dynamic processes tend to

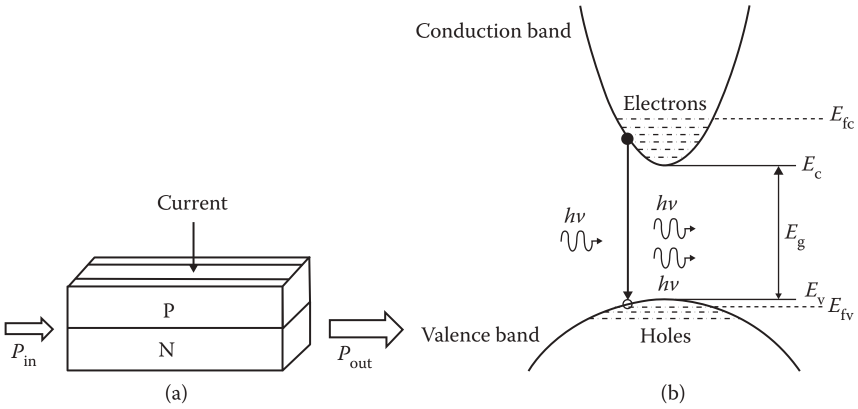


FIGURE 25.1 (a) Schematic representation of typical semiconductor optical amplifier (SOA) structure. (b) Simplified band diagram of semiconductor material. E_g, bandgap energy given by the difference between the energy at the bottom of the conduction band, E_c, and the energy at the top of the valence band, E_v; E_{fc}, quasi-Fermi level in the conduction band, and E_{fv}, quasi-Fermi level in the valence band.

bring the SOA gain back to its original level. These processes, which govern the SOA gain recovery, can be categorized in two cases of transitions (Occhi, 2002):

- Intraband transitions, which affect the distribution of carriers in the energy bands but leave the overall carrier concentrations unchanged. Varying the carrier density affects this distribution, which is further modified by spectral hole burning (SHB), free carrier absorption (FCA), carrier heating (CH), and carrier cooling.
- Interband transitions, which are related to the exchange of carriers between the energy bands and hence result in changes in the carrier density. These transitions are determined by electrical pumping, stimulated emission, absorption, spontaneous emission, nonradiative recombinations and two-photon absorption (TPA).

As electrons need not change band but are rearranged within the latter, the time constants of intraband transitions are much shorter than those of interband transitions. For this reason, the former are referred to as ultrafast effects. Given the time constants of these effects, they tend to be important compared with interband dynamics when the spectrum of the input signal is very wide, as in short pulse amplification, or when many channels are amplified in the context of wavelength division multiplexing (WDM) (Vacondio, 2011).

The SOA gain dynamics are directly coupled with those of free carriers (Occhi, 2002). Figure 25.2 qualitatively illustrates the temporal evolution of the free carrier distribution (energy versus density) in the conduction band (CB) of an SOA, which operates in the gain regime (Hall et al., 1994) subject to an ultrashort optical pulse that belongs to a time-varying signal (Occhi, 2002; Mørk et al., 2003; Wang, 2008). Prior to the pulse arrival, the carrier density within the CB is in quasi-equilibrium (Saleh and Teich, 1991), which occurs when the relaxation times for transitions within an energy band are much shorter than the relaxation time versus the other energy band. In this situation, the carrier distribution within the considered band is graphically shown in Figure 25.2a and is described by the product of the density of states, which is increased away from the band edge ($E = E_c$) at a rate that depends on the effective masses of electrons, with the probability that a given energy level, $E > E_c$, is occupied by an electron (Diez, 2000; Connelly, 2002). This probability is described by the Fermi function, $1/\left[\exp\left((E\text{-}E_{fc})/k_B T\right) + 1\right]$, where E_{fc} is the quasi-Fermi level, T is the temperature of the medium at thermal equilibrium and k_B is Boltzmann's constant (so that $k_B T$ defines the thermal energy, which equals 26 meV at $T = 300$ K). The position of the quasi-Fermi levels in the conduction and valence bands, E_{fc} and E_{fv}, respectively, is determined by the pumping rate (current injection), and if the latter is sufficiently large so that their separation exceeds the bandgap energy then the semiconductor medium provides gain and hence acts as an amplifier (Diez, 2000). Such a condition can be satisfied only if the quasi-Fermi levels lie, as shown in Figure 25.1, inside the conduction and valence bands, respectively, which for a semiconductor happens only under strong biasing (Vacondio, 2011). Note here that since the electrons and holes in the SOA are interrelated because of charge neutrality, it is sufficient to refer only to the conduction band (Schubert, 2004).

The natural quasi-equilibrium state is altered when pulses of duration in a subpicosecond timescale enter the SOA. This happens due to intraband effects which manifest according to the following phenomena: First, the induced stimulated emission responsible for signal amplification causes carriers to recombine around a narrow range of energies defined by pulse photon energy. This opens a hole in the carrier distribution, which is referred to as "SHB" and provokes a deviation from the normal Fermi distribution (Figure 25.2b). As a consequence, a localized reduction in the number of carriers at the transition energies occurs, while the total (electron/hole) carrier density within the band is reduced by the stimulated emission. For high-input pulse energies (>1 pJ) (Tang and Shore, 1998) TPA may also take place where, as the name implies, two photons are simultaneously absorbed and an electron is transferred from the valence band to a high energy level in the conduction band due to the high photon density in the active region (Hall et al., 1994; Mørk et al., 2003). In addition, a free carrier absorbs a photon and moves to a higher energy within the same band, according to the so-called FCA (Hall et al., 1994). After these effects have manifested on a timescale practically of few tens of femtoseconds (fs) (Diez, 2000) and the optical pulse has left

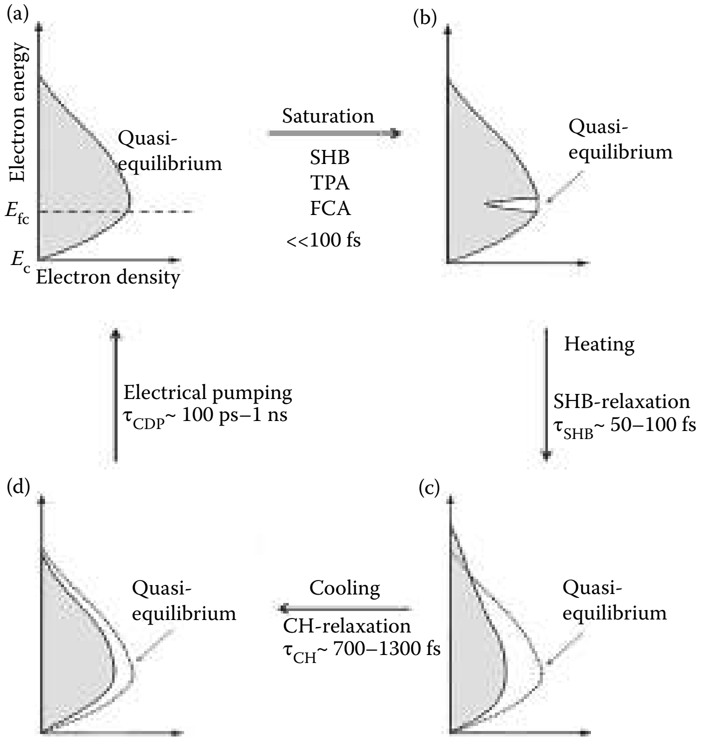


FIGURE 25.2 Evolution of carrier distribution induced in the conduction band (CB) of SOA by an ultrashort strong optical pulse. The shaded areas under the solid lines are proportional to the total carrier concentration in the CB (Wiesenfeld, 1996). The quasi-equilibrium is described by a Fermi distribution of quasi-Fermi level E_{fc}, which exceeds the minimum energy of the CB, E_c, in accordance with Figure 25.1b. SHB, spectral hole burning; CH, carrier heating; FCA, free carrier absorption; TPA, two-photon absorption, which occurs only for high pulse energies. The time constants values are typical for bulk SOAs, and may vary by more than a factor of two. (From Diez S, 2000, All-optical signal processing by gain-transparent semiconductor switches. PhD dissertation, Technical University of Berlin.)

the SOA, the Fermi distribution is restored via a process known as carrier–carrier scattering (Mørk et al., 2003), and the time required for this to happen is determined by the constant τ_{SHB}, which is known as SHB relaxation time, with values typically several tens of femtoseconds (50–100 fs). Since this parameter is finite, the governing SHB essentially sets the timescale during which the quasi-equilibrium Fermi distribution is established among the carriers in a band. Additionally, free carriers at energy levels lower than the average carrier density in the band are removed by stimulated emission or are transferred to higher levels due to the contribution of FCA and (possibly) TPA. As a result, the average carrier temperature is elevated and becomes higher than before the pulse arrival (Hall et al., 1994; Mørk et al., 2003). This transient increase of the electron and hole temperatures is called "CH" and impacts the carrier distribution (Figure 25.2c). The characteristic time, τ_{CH}, required for carriers to release their excess energy through phonon emission and cool down to the lattice temperature is of the order of several hundred femtoseconds (700–1300 fs). After this temperature relaxation time has elapsed, the starting Fermi distribution has been reestablished (Figure 25.2d), but the carrier density is still reduced compared to the stationary state (Figure 25.2a). Then the interband effect of carrier density pulsation (CDP) takes the lead and lasts between some hundreds of picoseconds and a few nanoseconds, depending on the dimensions and material of the active region as well as the SOA operating conditions. During this interval, which is defined by the respective time constant, τ_{CDP}, the carrier density, which had been decaying due to the stimulated emission, is increased again toward its original level owing to the resupply of electrons to the SOA via electrical pumping and the refilling of the respective bands. Finally, the carrier distribution is recovered to the quasi-equilibrium state (Figure 25.2a).

The distinctive evolution of the above processes can be conveniently viewed on the temporal profile of the SOA gain response, which is coupled to the carrier density. In practice, the SOA intraband and interband gain dynamics temporal characteristics and their dependence on operating parameters are directly disclosed by means of the pump-probe technique (Hall et al., 1994). This is a time-domain measurement technique where the SOA gain is perturbed by a strong optical pulse called "pump" and sampled at different delays, τ_{delay}, by a weak pulse called "probe," as illustrated in Figure 25.3a. The strong pump pulse drives the SOA under test into nonlinearity. The weak probe pulse cannot affect but instead experiences the altered SOA properties (Vallaitis, 2010). In order to capture the full extent of the SOA gain dynamics, the pulse repetition rate is of the order of several tens of MHz so that only one pump pulse at a time travels through the SOA and the gain of the latter is fully recovered before the arrival of the next pulse. Furthermore, ultrashort, that is, femtosecond-wide (Mørk et al., 2003) optical pulses are employed because their width determines the measurement's temporal resolution, which must be as high as possible to be able to extract the maximum information. A typical outcome of such measurement is graphically shown in Figure 25.3b, where the probe transmission has been plotted against the pump–probe delay, which is varied between a few ps to a few hundreds of ps (positive delay means that the pump precedes the probe pulse) [Occhi, 2002; Mørk et al., 2003; Wang, 2008; Vacondio, 2011]. The probe transmission is proportional to the SOA gain, so the curve monitors the SOA gain dynamic behavior.

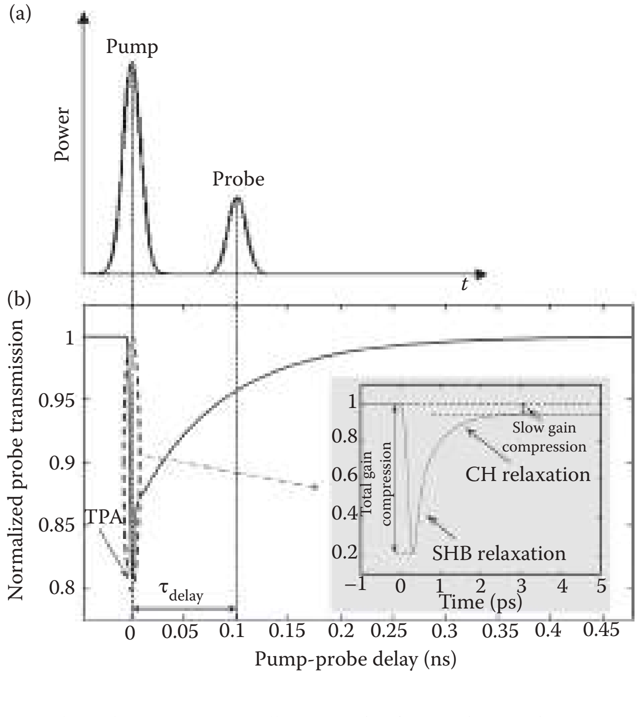


FIGURE 25.3 (a) Pump-probe technique. (From Vallaitis T., *Ultrafast Nonlinear Silicon Waveguides and Quantum Dot Semiconductor Optical Amplifiers*, Germany: Univ Karlsruhe, 2010.) (b) SOA typical evolution of (normalized) probe transmission as a function of pump–probe delay. The depicted curve does not represent exactly the defined SOA gain but rather its convolution with the pump and probe pulses. The timescale of the intraband gain dynamics is depicted in the zoomed-in inset. (From Occhi L, *Semiconductor optical amplifiers made of ridge waveguide bulk InGaAsP/InP: Experimental characterisation and numerical modelling of gain, phase, and noise, PhD dissertation*, Swiss Federal Institute of Technology in Zurich, 2002)

Initially, the pump pulse compresses the gain by an amount defined by the ratio of the unsaturated to the minimum level of the probe transmission, which occurs at the point where the carrier depletion is maximized by the pump peak power (Occhi, 2002). This gain compression is governed by SHB and (if enough photons are present in the SOA) TPA (Hall et al., 1994). Due to the finite resolution of the employed measurement technique (Hall et al., 1994), these ultrafast physical mechanisms are interpreted as instantaneous and thus overlap at zero time delay in the probe transmission (Schreieck, 2001). Afterward, the gain recovers within the interval required for the probe transmission to rise from 10% to 90% of the asymptotically approached maximum level. This recovery occurs according to the following separate processes. More specifically, in the first hundred femtoseconds timescale after the lowest compressed gain point SHB becomes the dominant nonlinear effect. Then CH takes over for a timescale of ~1 ps, which reestablishes the Fermi carrier distribution in the two energy bands. In the last time regime, the interband effects become pronounced and lead to slow gain compression defined by the unsaturated probe level, which is restored due to the applied electrical pumping, divided by the long-lasting probe transmission level (Occhi, 2002).

The impact of different SOA and signal operating conditions on the aforementioned gain dynamics has been investigated and assessed through experimental observations and numerical simulations. From the interpretation and analysis of the results obtained from both approaches, the following conclusions have been drawn (Occhi, 2002):

1. Both total and slow gain compression increase with the pump energy. For the total gain compression, this happens because the lattice temperature in the active region is elevated with higher pump energy. For the slow gain compression on the other hand, this variation is attributed to the reduction of the free carrier density as a by-product of the increased stimulated emission. Moreover, when the pulse energy is increased, it takes longer time for the gain compression associated with the intraband effects to recover.
2. The magnitudes of the two defined types of gain compression are affected differently by the width of the pump pulses. More specifically, the narrower the pump pulses, the larger the compression associated with the intraband effects. In contrast, the slow gain compression becomes smaller, even for input pulse energies higher than the SOA input saturation energy, for which the SOA gain is dropped by 3 dB against the small-signal gain. However, this is not possible for wider pulses, for example in the 100 ps range, since then the slow gain compression is much larger due to the much higher contribution of the interband effects to the SOA saturation.
3. Both gain compressions increase with current density because of the corresponding increase in gain. However, the fast gain recovery time behaves differently from the recovery time resulting from interband mechanisms. In fact, the former becomes longer for higher bias currents, while the latter is decreased.
4. The total gain compression is practically the same for all investigated device lengths in the case of an input energy equal to the SOA input saturation energy. The increase of the input energy beyond this point causes a larger gain reduction in longer SOAs, since their gain is higher compared to the shorter ones, and accordingly results in higher total gain compression. On the contrary, the fast gain recovery time is not influenced by the SOA longitudinal dimension. This finding does not hold too for the recovery time associated with the interband dynamics, since it decreases appreciably with the increase of the SOA length (Girardin et al., 1998).

25.3 SOA Model Formulation

Many theoretical treatments have been conducted aiming at explaining the operation, understanding the dynamics, predicting the behavior, evaluating the performance, and optimizing the working conditions of SOAs. For this purpose, both analytical and numerical models of different complexity level have been formulated. The more elaborate ones provide a thorough insight into what happens inside the SOA. This is done by microscopically describing the propagation of optical pulses and the change of the SOA carrier dynamics due to interaction with light. This approach provides extensive information at the expense of increased hardware and software resources. On the other hand, it is also practically desirable to maintain a good balance between mathematics and computer power. This can be achieved by viewing and treating the physical processes that manifest within an SOA from a phenomenological perspective (Toptchiyski, 2002).

The theoretical description of the operation of an SOA that is driven by a strong, ultrafast data pulse stream should be as realistic as possible while simultaneously being characterized by versatility and computational efficiency. In this effort, it is necessary to set the valid framework within which this task will be accomplished, which is based on the following key points. First, for SOA applications in which the employed pulse widths are below the critical value of 10 ps the influence of the intraband carrier processes on nonlinear gain compression is significant and must be included in the description of the SOA saturation under pulsed operation (Borri et al., 1999). Moreover, of the possible relevant phenomena, the effect of TPA can be disregarded because it becomes significant for pulses that are hundreds of femtoseconds wide and have energy larger than 1 pJ (Tang and Shore, 1998). In fact, in most cases of interest the width and energy of pulses launched into an SOA >0.1 ps and <1 pJ, respectively. For this reason, the main effects whose contribution is taken into account are CH and SHB. Second, the SOA small-signal gain, internal loss, and saturation energy are treated as wavelength independent parameters, since usually the spectral width of picosecond optical pulses is at least two orders of magnitude smaller than the gain bandwidth (Ning et al., 1997). The wavelength dependence of these SOA parameters must be taken into account when SOAs are employed for multichannel amplification (Jennen et al., 2001) or for nonlinear optical signal processing purposes, where more than one signal propagates and/or interacts within the active medium (Gutiérrez-Castrejón et al., 2000). The modeling approach that is usually adopted in this case is to approximate the spectral profile of the material gain coefficient per unit length with a parabolic function of wavelength, where the peak of the latter is a linear function of the carrier density (O'Mahony, 1988). The scope of this chapter, however, concerns single signal amplification and SOA linear applications, so high-order polynomial approximations of the gain spectrum would only increase the complexity of SOA modeling without offering essential information for the SOA performance. In addition to this reasoning, the central wavelength of the input signal launched into the SOA is assumed to be set at the peak of the SOA gain spectrum (in the vicinity of 1550 nm), which renders the modal gain (defined in Equation 25.2) independent of the wavelength. Third, the gain and group-velocity dispersion are neglected for pulse widths in the picosecond range and SOAs that are several hundreds of micrometers long (Agrawal and Olsson, 1989a; Tang and Shore, 1998). Fourth, the polarization of the input light is linear and preserved during propagation through the SOA, which is treated as a polarization independent element since the polarization sensitivity of practical devices is below 0.5 dB (Morito et al., 2003). Fifth, the residual facet reflectivities are practically negligible and cannot cause gain ripples, as enabled by the relevant technological evolution in the design and fabrication of SOAs devices (Zimmerman and Spiekman, 2004). Sixth, the SOA has the form of a traveling wave amplifier whose active region dimensions allow the support of only a single wave-guide, while it behaves as an isotropic device, which means that the susceptibility tensor is scalar (Gutiérrez-Castrejón et al., 2000).

The SOA response to an optical beam of electrical field $\vec{E}$ that propagates along the SOA largest dimension (that is, its length), which defines the z-axis in a Cartesian system of coordinates, is described by the carrier density rate equation (Tang and Shore, 1998)

$$\frac{\partial N}{\partial \tau} = \frac{I}{qV} - \frac{N}{\tau_{\text{carrier}}} - \frac{1}{\Gamma\sigma\hbar\omega}\frac{g\left|\vec{E}\right|^2}{1+\varepsilon\left|\vec{E}\right|^2}, \tag{25.1}$$

where $N = N(z, \tau)$ is the carrier density, I is the injection current, q is the electron charge, $V = wdL$ is the volume of the active region whose characteristic geometries have a cross section (wd) of the order of 1 μm^2 and a length (L) of 0.5 to 2 mm, τ_{carrier} is the carrier lifetime that is linked to the total recombination rate of carriers as specified in Gutiérrez-Castrejón and Duelk (2007), Γ is the confinement factor that accounts for the transverse character of the optical waves and the spread of the optical mode outside the active region of the amplifier (Agrawal and Olsson, 1989a), σ is the mode cross section ($= wd/\Gamma$), $\hbar$ is the reduced Planck's constant, ω is the angular frequency of the electromagnetic radiation, and ε is the gain compression factor, described right below in Equation 25.13. The parameter $g = g(z, \tau)$ stands for the modal gain that is assumed to vary linearly with N as

$$g = \Gamma a_N \left(N - N_{\mathrm{tr}}\right), \tag{25.2}$$

where a_N is the SOA differential gain and N_{tr} is the carrier density at transparency. Since, from Equation 25.2, $N = g/\Gamma a_N + N_{\mathrm{tr}}$, Equation 25.1 can be transformed, after some algebraic manipulations, into the following gain equation

$$\frac{\partial g}{\partial \tau} = \frac{g_{\mathrm{ss}} - g}{\tau_{\mathrm{carrier}}} - \frac{g\left|\vec{\mathrm{E}}\right|^2}{U_{\mathrm{sat}}\left(1 + \varepsilon\left|\vec{\mathrm{E}}\right|^2\right)}, \tag{25.3}$$

where $g_{\mathrm{ss}} = \Gamma a_N N_{tr}(I/I_{tr} - 1) = \frac{\ln(G_{\mathrm{ss}})}{L}$ is the coefficient of the nominal SOA small-signal power gain, G_{ss}, per unit length, $I_{\mathrm{tr}} = qVN_{\mathrm{tr}}/\tau_{\mathrm{carrier}}$ is the injected current required for transparency, and $U_{\mathrm{sat}} = \hbar\omega\sigma/a_N$ is the SOA intrinsic saturation energy, which for typical values $d = 250$ nm, $w = 2$ μm, $\Gamma = 0.48$ and $a_N = 3.3 \times 10^{-20}$ m^2 at an operating wavelength $\lambda = 1550$ nm is approximately 1 pJ. Alternatively, $U_{\mathrm{sat}} = P_{\mathrm{sat}}\tau_{\mathrm{carrier}}$, where P_{sat} is the saturation power of the SOA material, which in steady-state operation mode is approximately 0.7 times the SOA output saturation power (Connelly, 2002).

The dynamical evolution of an input light whose polarization is linear and maintained as it traverses the SOA obeys Maxwell's wave equation

$$\nabla^2\vec{\mathrm{E}} - \frac{1}{c^2}\left(n_b^2 + \chi\right)\frac{\partial^2\vec{\mathrm{E}}}{\partial\tau^2} = 0 \tag{25.4}$$

where c is the speed of light in vacuum, and $n_{\mathrm{b}}^2 \approx 1 + \Re\left\{\chi_0\right\}$ is the background refractive index, with $\Re\left\{\chi_0\right\}$ denoting the real part of the medium susceptibility in the absence of external pumping by current injection so that it accounts for material absorption (Schubert, 2004), while

$$\chi = -\frac{\bar{n}\,c}{\omega\Gamma}g\left(\frac{\iota}{1 + \varepsilon\left|\vec{\mathrm{E}}\right|^2} + \alpha_{\mathrm{CDP}} + \alpha_{\mathrm{CH}}\frac{\varepsilon\left|\vec{\mathrm{E}}\right|^2}{1 + \varepsilon\left|\vec{\mathrm{E}}\right|^2}\right) \tag{25.5}$$

is the susceptibility parameter (Gutiérrez-Castrejón et al., 2000), which represents the contribution of the charge carriers inside the active region, where $\bar{n}$ is the effective mode index, $\iota = \sqrt{-1}$ is the imaginary number, and the parameters α_{CDP} and α_{CH} are described below in Equation 25.22.

The electrical field of the linearly polarized light that is inserted in the SOA and propagates along its longitudinal axis, z, is $E(x, y, z, \tau) = \hat{x}\frac{1}{2}\mathrm{F}(x, y)\mathrm{A}(z, \tau)\exp\left[\iota(\beta z - \omega\tau)\right]$, where $\hat{x}$ is the polarization unit vector, $\mathrm{F}(x, y)$ is the transverse mode distribution, $A(z, \tau)$ is the complex envelope of the optical pulse and $\beta = \bar{n}\omega/c$ represents the propagation constant. According to the procedure in Tang and Shore (1998) and Gutiérrez-Castrejón et al. (2000), which is based on the slowly varying envelope approximation and the integration over the transverse dimensions, an equation that describes the propagation of the optical pulse in the SOA can be obtained from Equations 25.4 and 25.5

$$\frac{\partial\mathrm{A}(z, \tau)}{\partial z} + \frac{1}{\upsilon_{\mathrm{SOA}}}\frac{\partial\mathrm{A}(z, \tau)}{\partial\tau} = \frac{1}{2}\frac{g\mathrm{A}(z, \tau)}{1 + \varepsilon\left|\mathrm{A}\right|^2} - \frac{\iota}{2}\left[\alpha_{\mathrm{CDP}}g - \alpha_{\mathrm{CH}}\frac{\varepsilon g\left|\mathrm{A}\right|^2}{1 + \varepsilon\left|\mathrm{A}\right|^2}\right]\mathrm{A}(z, \tau). \tag{25.6}$$

This equation has actually been derived from semiclassical density-matrix equations by adiabatical approximation of the interband polarization dynamics (Mørk and Mecozzi, 1996).

The time measured in a static reference (laboratory) coordinate system, τ, is transformed into the local time t measured in a reference frame moving with the pulse at the group velocity inside the amplifier, υ_{SOA}:

$$t = \tau - \frac{z}{\upsilon_{\mathrm{SOA}}}. \tag{25.7}$$

Thus, in this moving coordinate system, the pulse is centered on $t = 0$ at every plane along the amplifier. According to Equation 25.7, the transformation of a general function $\mathrm{f}(z, \tau)$ to $\mathrm{f}(z, t)$ leads to the following expressions for the time and spatial derivatives (Jennen et al., 2001):

$$\frac{\partial \mathrm{f}(z,\tau)}{\partial \tau} = \frac{\partial \mathrm{f}(z,t)}{\partial t} \cdot \frac{\partial t}{\partial \tau} = \frac{\partial \mathrm{f}(z,t)}{\partial t} \tag{25.8}$$

$$\frac{\partial \mathrm{f}(z,\tau)}{\partial z} = \frac{\partial \mathrm{f}(z,t)}{\partial z} + \frac{\partial \mathrm{f}(z,t)}{\partial t} \cdot \frac{\partial t}{\partial z} = \frac{\partial \mathrm{f}(z,t)}{\partial z} - \frac{1}{\upsilon_{\mathrm{SOA}}} \cdot \frac{\partial \mathrm{f}(z,t)}{\partial t}. \tag{25.9}$$

Separating further “A” into amplitude and phase terms,

$$\mathrm{A} = \sqrt{\mathrm{Const} \times P}\exp(\iota\varphi) \Rightarrow P = |\mathrm{A}|^2 \big/ \mathrm{Const}. \tag{25.10}$$

“Const” is a normalization coefficient that relates the units between amplitude and power so that $\mathrm{A}(z, t)$ is expressed in units of $\mathrm{W}^{1/2}$, while $P = P(z, t)$ and $\varphi = \varphi(z, t)$ are the power and the phase of the traveling optical pulse, respectively. Applying Equations 25.8 and 25.9 together in Equation 25.6 and replacing in Equations 25.3 and 25.6 the squared modulus of the total electrical field, $|\vec{\mathrm{A}}|^2 = \vec{\mathrm{E}} \cdot \vec{\mathrm{E}}^* = |\mathrm{A}|^2$, where the symbol $*$ denotes the complex conjugate, the following set of partial coupled differential equations that govern the power, P, gain, g, and phase, φ, evolution of a strong optical signal that propagates and is amplified in an SOA is obtained (Tang and Shore, 1998)

$$\frac{\partial P(z,t)}{\partial z} = \frac{g(z,t)}{1 + \varepsilon P(z,t)} P(z,t) \tag{25.11}$$

$$\frac{\partial g(z,t)}{\partial t} = \frac{g_{\mathrm{ss}} - g(z,t)}{\tau_{\mathrm{carrier}}} - \frac{g(z,t)}{1 + \varepsilon P(z,t)} \frac{P(z,t)}{U_{\mathrm{sat}}} \tag{25.12}$$

$$\frac{\partial \varphi(z,t)}{\partial z} = -\frac{1}{2}\left[\alpha_{\mathrm{CDP}} g(z,t) - \alpha_{\mathrm{CH}} \frac{\varepsilon g(z,t) P(z,t)}{1 + \varepsilon P(z,t)}\right] \tag{25.13}$$

The term $\varepsilon = \varepsilon_{\mathrm{CH}} + \varepsilon_{\mathrm{SHB}}$ combines the contribution to the SOA gain compression from CH and SHB when these two effects are considered instantaneous. This assumption holds when the width of the launched optical pulses exceeds 1 ps, and is therefore valid under most practical situations encountered in optical communications. In this case, the impact of CH and SHB on the gain compression is taken into account through the term $1 + \varepsilon P$, where the factor εP is connected to a stationary carrier density above transparency, at which the SOA gain is unity and hence leaves intact a signal that passes through (Occhi, 2002; Wang, 2008).

Equations 25.11 and 25.12 form a system with unknowns the power and gain of the signal that perturbs the SOA nonlinear optical properties. This system cannot be solved in closed form, but only numerically. For this purpose, and in order to account for the spatial and temporal dependence denoted by the variables (z, t), L is divided into discrete segments, each of length $\Delta z = \frac{L}{m}$, as shown in Figure 25.4, while the pulses are sliced over their period, τ_{period}, at uniform intervals $\Delta t = \frac{\tau_{\mathrm{period}}}{k}$, where k is a non-zero integer so that their profile can be reconstructed at the output. Now, since the driving signal propagates through one amplifier segment during the sampling time, the space and time infinitesimal elements are linked to each

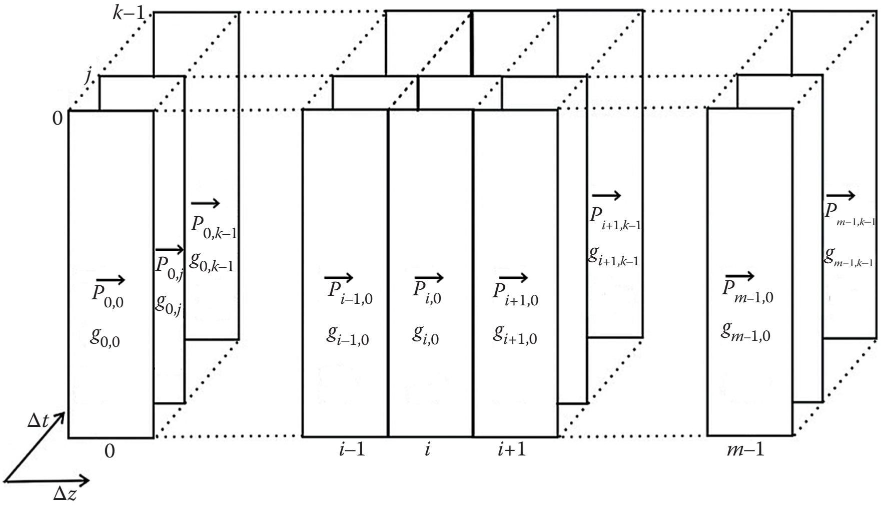


FIGURE 25.4 Dividing the SOA into discrete longitudinal (i) sections and calculating in each one of them the value of the signal power (P) and gain (g) for each sample of the pulse (j) based on the knowledge of the preceding value. This stepwise process is repeated in the formed spatio-temporal grid in an iterative manner until all required values are calculated.

other through $\Delta z = \upsilon_{\text{SOA}}\Delta t$ (Toptchiyski, 2002). The direct expansion gives

$$\Delta t = \frac{\Delta z}{\upsilon_{\text{SOA}}} = \frac{L}{m \cdot \upsilon_{\text{SOA}}} = \frac{\tau_{\text{transit}}}{m} = \frac{\tau_{\text{period}}}{k} \Rightarrow \frac{k}{m} = \frac{\tau_{\text{period}}}{\tau_{\text{transit}}}, \tag{25.14}$$

which means that the number k cannot be chosen independently from m. For example, if $\tau_{\text{period}} = 100$ ps and $\tau_{\text{transit}} = \frac{L}{\upsilon_{\text{SOA}}} = 12$ ps, the specific ratio equals 100:12. This means that the combination of k and m, or their selection, in order to satisfy this requirement, is not unique. Still, this task can be facilitated by taking into account that the number of the longitudinal sections must be larger than 10 (Wong and Blow, 2003). Otherwise, the SOA gain dynamics induced by the input lightwave signal are erroneously quantified. Thus, if $m = 120$, which is a more than sufficient value, then $k = 1000$. This pair allows obtaining a fairly accurate solution of the above equations with the use of simple first-order Euler differentiation. The latter is an efficient arithmetic method when studying in the time domain the nonlinear interactions between picosecond optical pulses whose envelope has a piecewise varying nature, as is the case for a Gaussian shape, and semiconductor active waveguide devices, provided that the sampling rate is high enough (Chi et al., 2001). More precisely, for each discrete space, i, and time section, j, the solution is expressed by

$$P_{\text{i}+1,j} = P_{i,j} + \Delta z \frac{\text{d}P}{\text{d}z}\bigg|_{i,j}, \quad \text{for} \quad i = 0,1,2,\ldots,m-1 \quad \text{and} \quad j = 0,1,2,\ldots,k-1 \tag{25.15}$$

$$g_{i,j+1} = g_{i,j} + \Delta t \frac{\text{d}g}{\text{d}t}\bigg|_{i,j}, \quad \text{for} \quad i = 0,1,2,\ldots,m-1 \quad \text{and} \quad j = 0,1,2,\ldots,k-1. \tag{25.16}$$

In other words, for an incoming signal traveling in the longitudinal direction measured from the left facet of the SOA, the knowledge of the values of $P(z,t)$ and $g(z,t)$ as well as of their derivatives (denoted by $|_{i,j}$) at a given step (i,j) of the algorithm, together with the fact that Equations 25.11 and 25.12 are essentially coupled in terms of the power and gain, allows us to calculate the next values $P(z + \Delta z, t)$ and $g(z, t + \Delta t)$ required at steps $(i + 1, j)$ and $(i, j + 1)$, respectively. This process is depicted in Figure 25.4 and is repeated in an iterative manner until all values are calculated in the spatio-temporal grid $i\,(\Delta z)\times j\,(\Delta t)$. For this purpose, it is necessary to define and apply the appropriate boundary and initial conditions. The setting of the boundary condition is straightforward since the profile of the launched optical power is expressed in analytical form depending on the particular pulse data format and shape. For example, for a return-to-zero on-off keying signal, the power of each temporal sample at the SOA input facet is given by (Pauer et al., 2001) $P_{0,j} = P(t) = \sum_n \left\{C_n\right\} p\left(j\Delta t - n\tau_{\text{period}}\right)$, where n is the number of bits contained in the data sequence length, the code $\left\{C_n\right\}$ is either "1" or "0" with equal probability ½, and $p(t)$ is the pulse shape. For Gaussian pulses (Agrawal and Olsson, 1989a), for example, $p(t) = P_{\text{peak}} \exp\left[-4\ln 2(t - \tau_{\text{period}}/2)^2/\tau_{\text{FWHM}}^2\right]$, where P_{peak} is the peak power and τ_{FWHM} is the full-width at half-maximum (FWHM), which occupies some fraction of the operating period that defines the duty cycle (Pauer et al., 2001). The setting of the initial condition, on the other hand, is more elaborate, since it requires considering the flow of the SOA gain dynamics change by the continuously arriving optical pulses (Tang et al., 2000). More specifically, in each distance segment of the SOA, the first temporal segment in the leading edge of the first arriving pulse experiences an unsaturated, small-signal gain, so that $g_{i,0}\left.\right|_1 = g_{\text{ss}}$. Then, by induction, in each distance segment, the first temporal segment in the leading edge of the following pulse, which reaches the SOA after τ_{period}, sees an initial, recovered gain given by $g_{i,0}\left.\right|_n = g_{\text{ss}} + \left(g_{i,k-1}\left.\right|_{n-1} - g_{\text{ss}}\right)\exp\left(-\frac{\tau_{\text{period}}}{\tau_{\text{carrier}}}\right)$, where $g_{i,k-1}\left.\right|_{n-1}$ is the gain perturbed by the last temporal segment in the trailing edge of the preceding pulse at the same distance segment.

Apart from this arithmetic procedure, other more elaborate techniques can also be employed for numerically solving modified-type nonlinear Schrödinger equations as Equation 25.6, such as the well-known fourth-order Runge-Kutta, or the finite-difference beam propagation method (FDBPM), where the central-difference approximation is applied in the time domain and trapezoidal integration is executed over the spatial section in an iterative manner (Razaghi et al., 2009). This method has enabled us to accurately acquire information for the SOA characteristics, which include the pulse shape, spectrum, and chirp, in the presence of the main ultrafast nonlinear phenomena that critically affect the SOA gain dynamics (Hosseini et al., 2011).

25.4 SOA Model Simplification

The model deployed for simulating the operation of an SOA has been formulated on the basis of the partial coupled differential Equations 25.11 through 25.13. These constitute a system which, due to the existence of the temporal and longitudinal variables, t and z, respectively, is two-dimensional (2D) so that its solution is rather cumbersome from a computational standpoint. However, if the SOA is treated as a "black box" characterized by its impulse response to any arbitrary input optical signal, the dependence on the spatial variable can be dropped and the problem can be reduced to a one-dimensional (1D) one described by an ordinary differential equation in which time is the only independent variable, enabling us to greatly simplify the computational complexity (Cassioli et al., 2000). This reduction is realized starting from the local changes of the carrier density distribution, which occur due to the interaction between the propagating field and the semiconductor material. These changes are the result of the intraband (CH and SHB) and interband (CDP) effects, while nonlinear processes with characteristic times shorter than SHB, such as TPA, are ignored because their effect is comparatively not significant. The gain and the refractive index of the semiconductor depend on these locally perturbed densities, so each local density distribution is associated with a distinct term that contributes to the total gain. The gain then can be expanded to the sum of the respective contributions of the three different physical processes considered. The temporal evolution of the three quantities is described by equal number of rate equations, which in turn depend on the photon density profile and therefore are coupled to the equation that describes the propagation of the field along the waveguide. The key thus for reducing the order of the modeling system to 1D is the integration of both sides of the partial differential equations over the entire SOA length. This transformation leads to the following compact expression for the total device gain (Cassioli et al., 2000)

$$G(t) = \exp[\mathrm{h}(t)], \tag{25.17}$$

where at each point of the pulse profile, $\mathrm{h}(z,t) = \int_0^z \frac{g(z',t)}{1+\varepsilon P(z',t)} dz'$ represents the power gain integrated over the length of the SOA, which hence is treated as a spatially concentrated device. This dimensionless coefficient incorporates the contribution of each mechanism that affects the SOA gain:

$$\mathrm{h}(t) = \mathrm{h}_{\mathrm{CDP}}(t) + \mathrm{h}_{\mathrm{CH}}(t) + \mathrm{h}_{\mathrm{SHB}}(t). \tag{25.18}$$

More specifically, $\mathrm{h}_{\mathrm{CDP}}$ represents the gain associated with the interband effects and its temporal evolution is described by the following first-order ordinary differential equation

$$\frac{\mathrm{dh}_{\mathrm{CDP}}}{\mathrm{d}t} = -\frac{\mathrm{h}_{\mathrm{CDP}}}{\tau_{\mathrm{carrier}}} - \frac{1}{P_{\mathrm{sat}}\tau_{\mathrm{carrier}}}\,[G(t)-1]\,P(t) + \frac{\ln G_{\mathrm{ss}}}{\tau_{\mathrm{carrier}}}. \tag{25.19}$$

h_{CH} is the gain associated with CH, which means that it describes the gain variation due to the mechanisms that heat up the carrier distribution in the conduction band, such as FCA and TPA, as well as due to changes of the carrier density. The relevant rate equation has the form

$$\frac{\mathrm{dh}_{\mathrm{CH}}}{\mathrm{d}t} = -\frac{\mathrm{h}_{\mathrm{CH}}}{\tau_{\mathrm{CH}}} - \frac{\varepsilon_{\mathrm{CH}}}{\tau_{\mathrm{CH}}}\,[G(t)-1]\,P(t) \tag{25.20}$$

where $\varepsilon_{\mathrm{CH}}$ is the CH nonlinear gain compression factor. The first term on the right-hand side characterizes the relaxation to the lattice temperature once the exciting pulse has left the SOA and before the arrival of the next pulse. The second term combines all the effects that heat-up the carrier distribution. The typical values of τ_{CH} for InGaAsP/InP-based SOAs with gain peak around 1.55 μm lie between 0.5 and 1 ps, while $\varepsilon_{\mathrm{CH}}$ is between 0.28×10^{-23} and 4.4×10^{-23} m^3 (Occhi, 2002).

Finally, the rate equation for the SHB contribution is

$$\frac{\mathrm{dh}_{\mathrm{SHB}}}{\mathrm{d}t} = -\frac{\mathrm{h}_{\mathrm{SHB}}}{\tau_{\mathrm{SHB}}} - \frac{\varepsilon_{\mathrm{SHB}}}{\tau_{\mathrm{SHB}}}\,[G(t)-1]\,P(t) - \frac{\mathrm{dh}_{\mathrm{CH}}}{\mathrm{d}t} - \frac{\mathrm{dh}_{\mathrm{CDP}}}{\mathrm{d}t}, \tag{25.21}$$

where $\varepsilon_{\mathrm{SHB}}$ is the SHB nonlinear gain compression factor. Reported values of τ_{SHB} in the conduction band for InGaAsP/InP-based SOAs with maximum gain centered around 1.55 μm are between 30 and 250 fs, while $\varepsilon_{\mathrm{SHB}}$ has been found to be between 0.14×10^{-23} and 1.7×10^{-23} m^3 (Occhi, 2002).

The above dynamic processes affect not only the gain properties but also the refractive index of the SOA within the active region (Wiesenfeld, 1996). This happens because the free carriers contribute substantially to the SOA optical susceptibility, which is complex in nature. The imaginary part of this parameter is related to the gain of the SOA, and the real part is related to the refractive index (Eiselt et al., 1995). The gain and refractive index variations are not independent but coupled via the Kramers–Kronig relationship, according to which the refractive index modulation can be calculated from the gain variation (Hutchings et al., 1992). However, this calculation requires us to integrate over the whole wavelength spectrum the gain coefficient, which means that the latter must be known for a wide range of optical frequencies. Furthermore, the specific integral must be solved numerically, which increases the model computational time significantly. A more convenient manner to describe the refractive index behavior relies on the alpha factor, also known as the Henry factor (Henry, 1982), which links the changes of the active layer refractive index to those of the gain when the carrier density is varied. The important benefit offered by the employment of this factor is that the refractive index dynamics are given directly by the gain dynamics and hence are quantified much faster than with the Kramers–Kronig relations. Moreover, the adoption of the alpha-factor approach is practically realizable since it links the gain and the phase variations associated with the changes of the refractive index, where both are directly measurable. The phase of the optical field at the SOA output can be expressed as a linear function of the same variables which describe the optical gain:

$$\varphi(t) = -0.5\,[\alpha_{\mathrm{CDP}}\mathrm{h}_{\mathrm{CDP}}(t) + \alpha_{\mathrm{CH}}\,\mathrm{h}_{\mathrm{CH}}(t) + \alpha_{\mathrm{SHB}}\,\mathrm{h}_{\mathrm{SHB}}(t)], \tag{25.22}$$

where α_{CDP}, α_{CH}, and α_{SHB} are the phase-amplitude coupling coefficients of CDP, CH, and SHB, respectively. Among these alpha factors, α_{CDP} is the traditional linewidth enhancement factor related to CDP, with typical values 3–12 (Agrawal and Olsson, 1989a). α_{CH} is the alpha factor associated with CH and varies especially near the band edge, where the depletion of cooler carriers takes place (Giller et al., 2006). However, usually signal photon energies influence carriers well inside the band and so, α_{CH} can be approximated by a constant with indicative values ranging between 1 and 4.5, as reported in the literature. Finally, α_{SHB} is the alpha factor linked to SHB and is nearly zero because SHB happens symmetrically around the central wavelength (Mecozzi and Mørk, 1997). Thus, the Kramers–Krönig integral over the whole spectrum is nearly zero and the change of the refractive index is very small near the peak wavelength.

Every phase variation at the SOA output is always accompanied by a certain amount of chirp (Agrawal and Olsson, 1989a). This is the instantaneous frequency deviation relative to the optical beam frequency, $\Delta\nu$, which is provoked because the phase, φ, of the amplified signal is not constant in time. Mathematically, it is given by

$$\Delta\nu(t) = -\frac{1}{2\pi}\frac{\partial\varphi}{\partial t}. \tag{25.23}$$

The set of rate Equations 25.19 through 25.21 for the h_j (j = CDP, CH, SHB) functions is numerically solved for a given optical power to the SOA input, $P(t)$, which is the only excitation of the system. Then, Equations 25.17, 25.22 and 25.23 give the overall power gain, phase variation, and chirp that the input signal experiences through the SOA, respectively. In this manner, results have been obtained for the temporal, spectral, and chirp profile of the amplified pulses (Hussain et al., 2010).

Equations 25.11 through 25.13 can be further combined and reduced for optical pulses going into the SOA, whose width exceeds the critical limit of 10 ps, above which the influence of the CH and the SHB effects is not significant (Borri et al., 1999). This means that in this case it is not necessary to take into account these intraband processes. Then, by following the algebraic steps detailed in Mecozzi and Mørk (1997), but setting, ε to null and $\mathrm{h}_{\mathrm{CDP}} \equiv \mathrm{h}(t)$, the differential equation for the SOA response is simplified to the well-known form that holds when the SOA gain saturation is caused only by the depletion of the carrier density due to stimulated emission (Agrawal and Olsson, 1989a)

$$\frac{\mathrm{dh}(t)}{\mathrm{d}t} = \frac{\ln(G_{\mathrm{ss}}) - \mathrm{h}(t)}{\tau_{\mathrm{carrier}}} - \frac{P(t)}{U_{\mathrm{sat}}}\left\{\exp\left[\mathrm{h}(t)\right] - 1\right\}. \tag{25.24}$$

This equation can be solved analytically, provided that the width of the launched optical pulses is much shorter than the SOA carrier lifetime. This condition holds in practice, since in real SOA devices, the values of the latter parameter are typically of the order of 100 ps and above. Thus, closed-form expressions can be obtained for the variation of the SOA gain in the rapid saturation and slow recovery regions. These expressions allow us to explicitly describe the pulse profile at the SOA output when this element is driven by multiple pulses of alternating binary content, as it happens in real optical networks that are based on optical time division multiplexing (OTDM) (Hamilton et al., 2002).

25.5 SOA Optical Gain Modulation

The influence of critical parameters, which include the SOA small-signal gain and carrier lifetime as well as the FWHM and energy of the input pulses, on the SOA output characteristics has been investigated and evaluated for the case where the change of the SOA gain dynamics is provoked by a 10-Gb/s data-modulated signal onto itself in the context of straightforward amplification (Zoiros et al., 2007). The results reveal that due to the continuous arrival of pulses, which cause the SOA saturation properties to be more strongly modified than for a single pulse, the requirements for the SOA small-signal gain and the input pulse energy are more stringent than those for the isolated pulse amplification studied elsewhere (Agrawal and Olsson, 1989a), while those for the FWHM and carrier lifetime are also tight but to a lesser extent. For example, Figure 25.5 shows that the SOA small-signal gain must be an order of magnitude lower than that for a single pulse or else the output pulse deviates from its normal symmetrical form (Zoiros et al., 2007). Moreover, Figure 25.6 shows that the pulse energy, U_{in}, must be a small fraction of the SOA saturation energy, U_{sat} (Zoiros et al., 2007). These trends together suggest that the SOA must be biased to operate in the low-saturation regime in order to preserve the shape of the input signal at the SOA exit. This condition essentially determines the dynamic range of the inserted optical signal, which is allowed for distortionless pulse amplification, and its validity is further supported by evidence given in Section 25.6. In contrast, when the focus is on the chirp and how it can be tailored in order to be exploited for pulse compression (Agrawal and Olsson, 1989b; Zoiros et al., 2005), the investigation that was conducted reveals that the

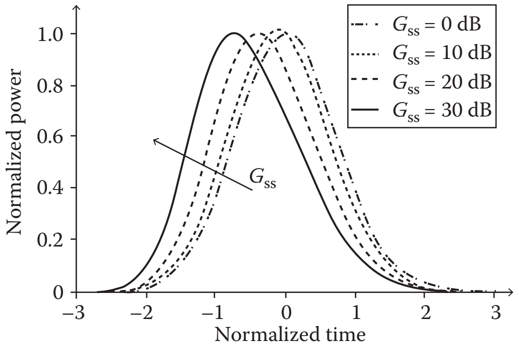


FIGURE 25.5 Shape of amplified pulses versus SOA small-signal gain, G_{ss}. The arrow indicates the G_{ss} direction of increase.

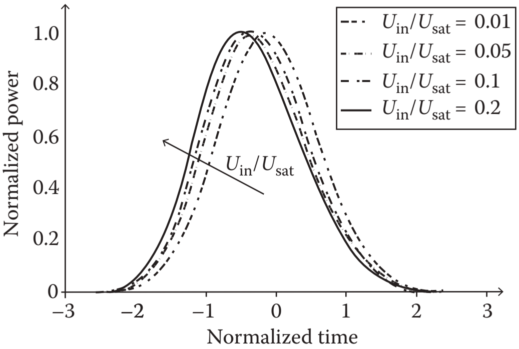


FIGURE 25.6 Shape of amplified pulses versus input pulse energy U_{in}, normalized to SOA saturation energy U_{sat}. The arrow indicates the $U_{\text{in}}/U_{\text{sat}}$ direction of increase.

SOA small-signal gain and the energy of the incoming pulses must be increased to such extent so that together they heavily saturate the SOA. Then, the reduction of the instantaneous phase shift (Ueno et al., 2002) and of the accompanying chirp due to the alternating bit pulses and the concomitant inability of the SOA to timely recover between them can be efficiently mitigated. In this case, the chirp acquires the form shown in Figure 25.7, which is the required one for the intended purpose (Agrawal and Olsson, 1989a; Zoiros et al., 2007). In fact, the linearly increasing part of the chirp curve in the time domain is situated in the center of the pulse where most of its power is contained, while the chirp magnitude is sufficiently enhanced. Therefore, if the SOA is operated according to these conditions, then it can be ensured that the amplified optical pulses exhibit an appropriate power and chirp profile.

On the other hand, if the FWHM is such that its difference from the pulses repetition interval is smaller than the SOA gain recovery time, then the strain imposed on the SOA gain dynamics is heavier. This happens because the SOA responds more slowly to the rapidly varying logical content of the wider pulses, which worsens the pattern effect described in the next section (Kim et al., 2007). In this demanding case, Equation 25.24 must obligatorily be solved by a numerical approach. This can be done in a stepwise manner by sampling the optical pulse over its period at discrete temporal intervals, approximating the time derivative by a finite difference and applying the appropriate initial conditions to account for the different gain at the beginning of each new bit depending on how it has been perturbed during the previous one (Botsiaris et al., 2007). The knowledge then of $\mathrm{h}(t)$ allows us to calculate the electric field of the amplified data, $E_{\mathrm{SOA}}(t) = E_{\mathrm{data}}(t)\exp\left[\frac{1}{2}\left(1 - j\alpha_{\mathrm{CDP}}\right)\mathrm{h}(t)\right]$, where, according to Equation 25.10, $|E_{\mathrm{SOA}}(t)|^2 = P_{\mathrm{SOA}}(t)$ and $|E_{\mathrm{data}}(t)|^2 = P(t)$ (Agrawal and Olsson, 1989a), as well as the chirp from $\Delta\nu(t) = -(1/4\pi)\left[\alpha_{\mathrm{CDP}}\mathrm{dh}(t)/\mathrm{d}t\mathrm{d}t\right]$, where the derivative of $\mathrm{h}(t)$ is directly taken from the right-hand of Equation 25.24. With this approach, the changes that are incurred on the amplified pulse trains and eye diagrams for different SOA saturation conditions can be correctly captured, as verified by comparison to experiments. For this purpose, the degree of pattern effect is modified by altering the maximum launched signal power and the SOA small-signal gain. Figure 25.8 depicts the amplified pulse patterns when the peak power of the marks successively drives the SOA into the low, medium, and deep saturation region, respectively. As the input peak power is progressively doubled, the peak amplitude fluctuations become more pronounced to an extent that depends on whether the marks are preceded either by one or more spaces or by other marks (Zoiros et al., 2008). In contrast, as the input peak power is decreased, these variations are smoothed and the associated pattern effect de-escalates. Similarly, when the SOA small-signal gain is halved, the peak amplitudes become less uneven, as illustrated by comparing Figure 25.9 to Figure 25.8c. The accurate reproduction of the form of the pattern effect which manifests on RZ pulses (Zoiros et al., 2009, 2010a,b), together with the good qualitative matching between the left- and right-hand sides of these figures, designates that the experimental trend for the SOA response is properly modeled. This also holds

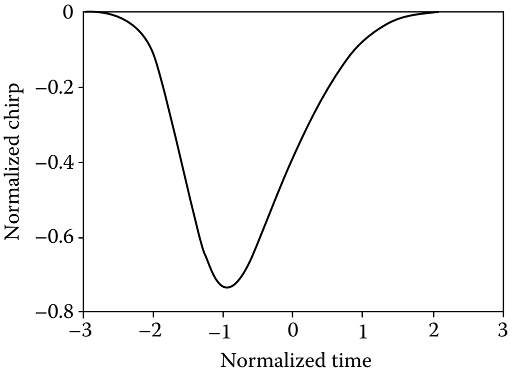


FIGURE 25.7 Profile of SOA output chirp for heavily saturated SOA.

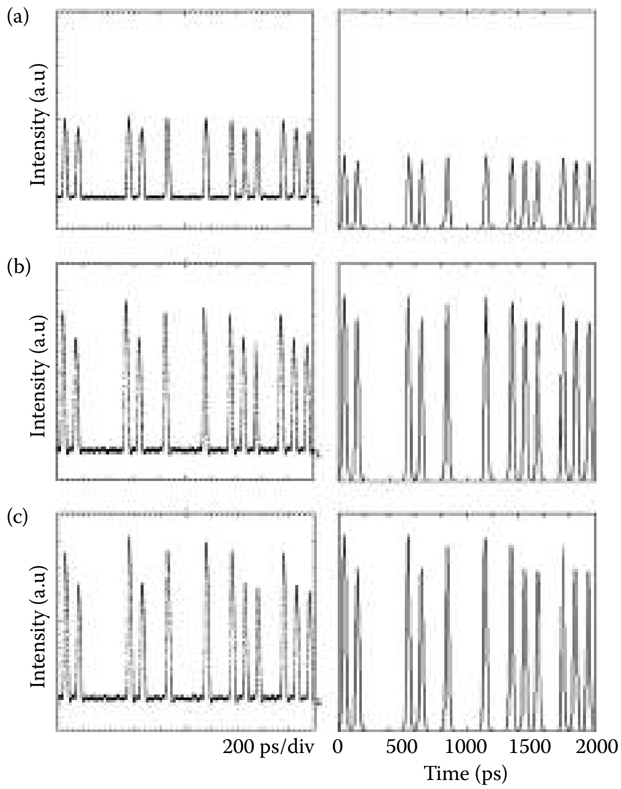
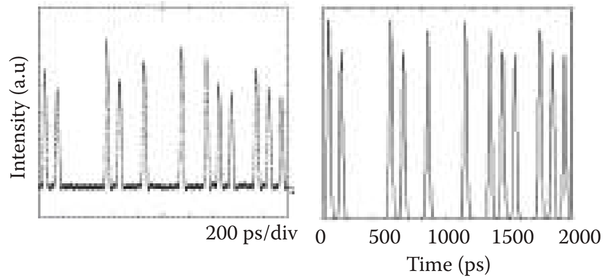


FIGURE 25.8 Experimental (left) and theoretical (right) amplified data pulse trains for (a) low, (b) medium, and (c) deep SOA saturation.



(a) (b)

FIGURE 25.9 Experimental (a) and theoretical (b) amplified data pulse trains for halved SOA small-signal gain compared to Figure 25.8(c).

for the respective eye diagrams shown in Figure 25.10. In this case, the asymmetric subenvelopes that occur for a high-input peak power in Figure 25.10c gradually disappear as this parameter is decreased, in Figure 25.10a and b, and the eye diagrams become more uniform, which is an improvement that can be monitored too in the simulated pseudo-eye diagrams obtained as detailed in Gutiérrez-Castrejón et al. (2001).

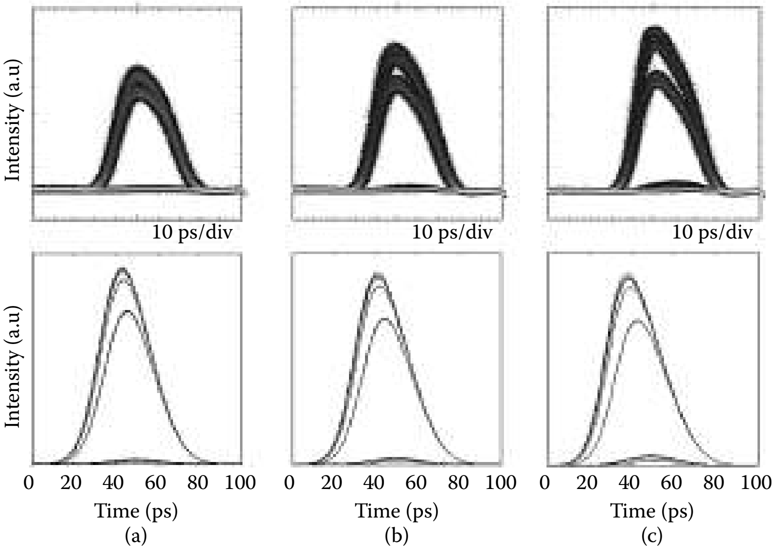


FIGURE 25.10 Experimental (top) and theoretical (bottom) eye diagrams at SOA output for (a) low, (b) medium, and (c) deep SOA saturation.

25.6 Pattern Effect I: Direct Optical Amplification

The distinctive properties of SOAs have revived interest for direct signal amplification purposes, which have traditionally been the primary target of SOAs (Connelly, 2002). As shown in Figure 25.11, the intention is to exploit SOAs as booster amplifiers to enhance the signal power before being launched into an optical link (Kim et al., 2003) as in-line amplifiers to compensate for transmission losses along the link (Boscolo et al., 2006) and as preamplifiers so that the received signal has sufficient power and can be correctly detected (Mynbaev and Scheiner, 2001; Singh, 2011).

In these applications, the profile of the output signal should ideally be an amplified replica of the input signal. In practice, the combination of the power and duration of the input data signal can be such that the SOA is heavily saturated. Concurrently, the SOA gain recovery time is finite with typical values of the order of few hundreds of picoseconds. Thus, in most practical cases, it exceeds the pulses repetition period, which scales inversely with the data rate. Under these driving conditions, the SOA gain is not perturbed in a regular fashion but according to the binary content of each bit slot. As can be schematically seen in Figure 25.12, if the SOA is excited by a train of distinguishable pulses, that is, "1" and "0," whose period, τ_{period}, is such that they arrive faster than the interval available for carrier replenishment, τ_{recovery}, then the gain variation is not uniform since it is dropped for a "1" and partially recovered for a "0" (Manning et al., 1997). This mode of operation has a negative impact on the SOA response, as it can be highlighted by means of the following representative scenarios which concern two "1"s which are separated by a "0." First, in case of a row of marks, the "1" which follows immediately after the "1" that has initially been inserted in the SOA and hence modified its optical properties from a higher level, suffers a reduced gain and this continues to happen for every subsequent "1" with respect to the preceding "1" until an equilibrium is reached or the sequence is broken by a "0." Second, in case of multiple "0"s, which allow the gain to rise substantially, the leading "1" experiences a higher gain than the trailing "1," given that the latter is preceded by only one "0," while it encounters a comparatively lower gain if there are continuous "1"s that prevent the gain from completely recovering. In either case, the amplified output suffers from peak-to-peak amplitude fluctuations, as can characteristically be seen in Figure 25.13.

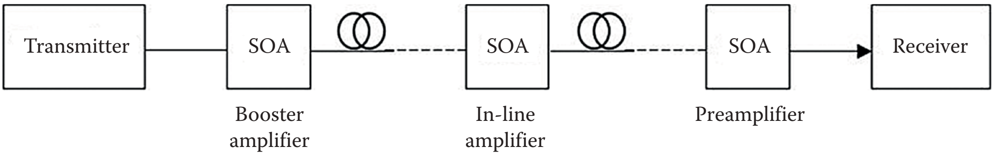


FIGURE 25.11 SOA direct amplification applications.

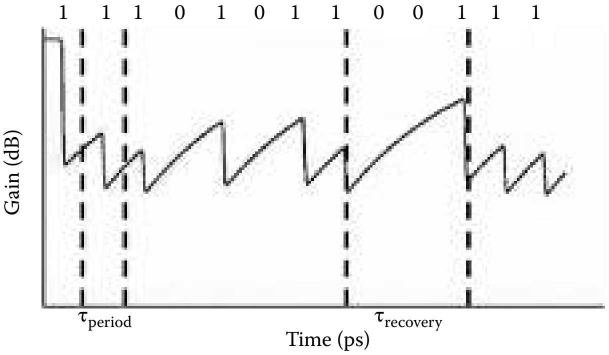


FIGURE 25.12 Evolution of SOA gain in response to data train input.

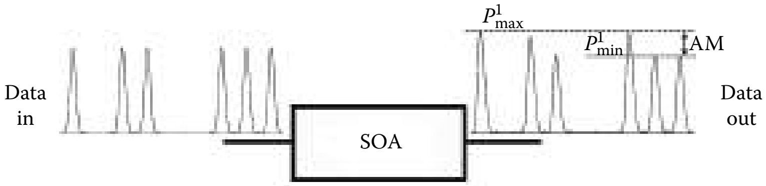


FIGURE 25.13 Data amplification in SOA and pattern-dependent output distortion, with amplitude modulation (AM) definition.

This undesirable phenomenon is called “pattern effect” (Manning et al., 1997) and constitutes a major obstacle in the effort to employ conventional SOAs in their classical role. This happens due to the significant deterioration of several metrics that characterize the quality of the amplified signal, which include the amplitude modulation (AM), the extinction ratio (ER), and the input power dynamic range (IPDR). The AM is defined as AM (dB) $= 10\log\left(P^1_{max}/P^1_{min}\right)$, where P^1_{max} and P^1_{min} are the maximum and minimum peak power of the marks in the amplified data stream, respectively (Figure 25.13). The AM quantifies the degree of uniformity of the marks and should be as low as possible but it is considered acceptable for lightwave telecommunications systems if it is lower than 1 dB (Zoiros et al., 2008). The ER is defined as $\mathrm{ER}\,(dB) = 10\log\left(P^1_{avg}/P^0_{avg}\right)$, where P^1_{avg} and P^0_{avg} are the average power of the marks and spaces in the amplified data stream, respectively (Figure 25.16). The ER quantifies how distinct the marks are from the spaces and it should be as high as possible but it is considered acceptable if it is well over 10 dB (Hinton et al., 2008). Finally, the IPDR is defined as the range where the Q-factor, which in the thermal noise limit, where the amplitude fluctuations due to the pattern effect act as noise variance on the marks, is defined as $\mathrm{Q} = P^1_{avg}/\sigma^1_{std}$, with σ^1_{std} being the standard deviation of the peak powers of the marks, is over six (Agrawal, 2002). The negative impact of the SOA pattern effect on these metrics has been investigated and evaluated and the main outcome is compiled and presented in the following in Figures 25.14 and 25.15 (Rizou et al., 2015a). In Figure 25.14, the SOA output suffers from intense peak amplitude fluctuations

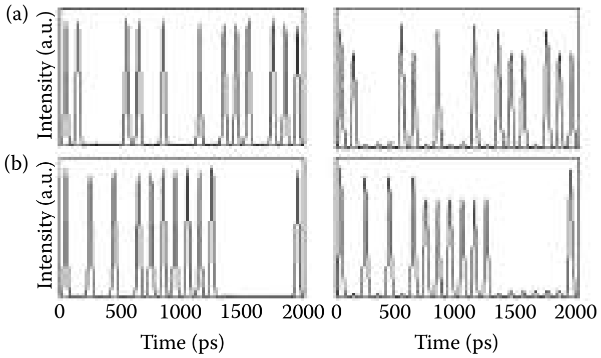


FIGURE 25.14 (a) SOA input data pattern with initial AM = 0.35 dB (left) and corresponding output (right). (b) SOA input data pattern which contains a string of consecutive “0”s (left) and corresponding output (right).

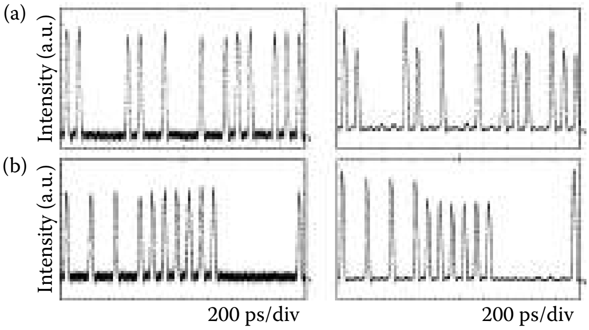


FIGURE 25.15 Corresponding experimental patterns of Figure 25.14.

both when the framed input data waveform has an initial AM = 0.35 dB (Figure 25.14a, left) or contains a string of consecutive “0”s (Figure 25.14b, left). In both cases, the AM level at the SOA exit is intolerable, as it amounts to AM = 1.3 dB (Figure 25.14, right). This behavior is comparable to the experimental one in Figure 25.15 with regard to pulse profile and amplitude wandering. Moreover, in Figure 25.16, the eye diagram at the SOA output is degenerated into secondary traces whose borders deviate from the principal one both in magnitude and shape, while the baseline defined by the spaces is less than 11 dB away from the average level of the marks (Rizou et al., 2013).

In Figure 25.17, on the other hand, the part of the chirp at the SOA output which is negative at the leading edge of each mark, that is, ‘red chirp’ (Agrawal and Olsson, 1989a), suffers from strong peak-to-peak excursions similar to those observed at the corresponding pulse positions, while the part of the chirp which is positive at the trailing edge of each mark, that is, ‘blue chirp’, exhibits a much weaker dependence than its red chirp counterpart. Physically, this happens because the red chirp is related to carrier depletion and gain compression, while the blue chirp is linked to carrier regeneration and gain recovery (Girault et al., 2008). This means that the magnitude of the induced chirp is proportional to the rate of change of the carrier concentration and subsequently of the gain. For this reason, it would be rationally expectable that the pattern effect due to the irregular gain variation would be imposed on both types of chirp. However, as can be seen in Figure 25.17, this holds almost exclusively for the red chirp, which exhibits strong peak fluctuations. For the blue chirp, in contrast, an increase in the peaks occurs only for successive “1”s, since then the carrier recombination rate is enhanced due to the weaker carrier depletion, but still this increase is so small that it is hardly noticed. This pronounced discrepancy in the pattern dependence is attributed to

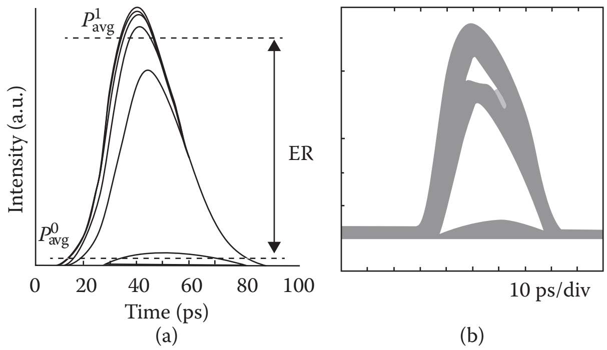


FIGURE 25.16 Theoretical (a) and experimental (b) eye diagrams at SOA output, with extinction ratio (ER) definition.

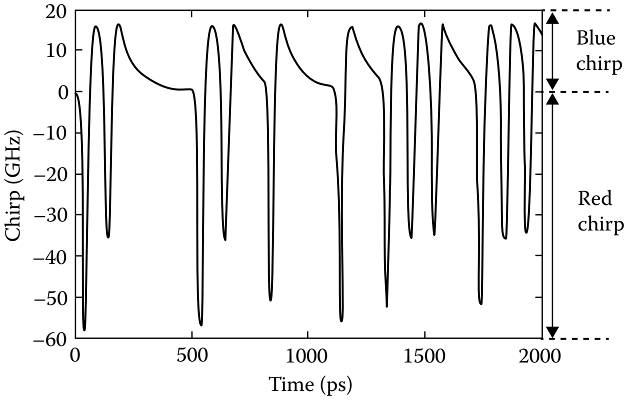


FIGURE 25.17 Chirp variation at SOA output.

the different timescale on which the two chirp phenomena take place. In fact, the carrier depletion and the accompanying phase increase in the leading edge of the pulse is somewhat faster compared to the much slower carrier replenishment and the concomitant phase decrease in its trailing edge, which gives rise to an approximately constant amount of blue chirp. Therefore, the explanation of the pattern effect can also be given in the frequency domain through the red chirp. For this purpose, its form in Figure 25.17 is correlated (Wang et al., 2007) with the corresponding gain change in Figure 25.18.

More specifically, the rising edge of the first "1" at the left-most edge of the pattern window acquires the largest red chirp among all the marks owing to the fact that, for the same energy, its gain is dropped from the maximum possible level, that is, that of the unsaturated state, and as a by-product the incurred gain difference is highest. Similarly, the next largest peak is that of the third "1," since it is preceded by a run of three "0"s. In contrast, the lower the carrier density, or equivalently the greater the extent of gain compression, the slower the additional carriers are depleted, thereby resulting in decreased gain variation and accordingly less red chirp, as it is the case for consecutive "1"s where the SOA cannot fully recover in between them. Finally, in Figure 25.19, the Q-factor is permissible only in saturation region A, where the SOA gain (inset) is reduced up to just 3 dB (Rizou et al., 2013). Thus, the IPDR cannot be extended to regions B and C, since the SOA pattern effect is aggravated as the gain is decreased by more than 3 and 6 dB, respectively. This limitation deteriorates the optical signal-to-noise ratio (OSNR), impedes the maximization of the output power and results in closer amplifier spacings. In order to combat all these pattern-induced

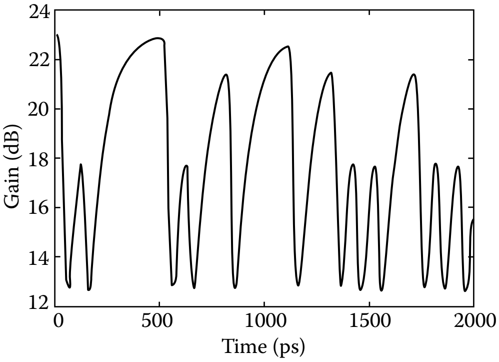


FIGURE 25.18 Instantaneous SOA gain variation.

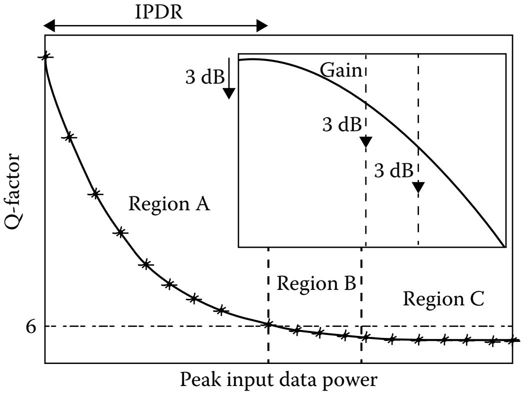


FIGURE 25.19 Variation of Q-factor and corresponding SOA gain (inset) versus peak input data power. The horizontal dotted line denotes the lower limit set for the Q-factor. Defined SOA saturation regions: A = low, B = medium, and C = deep.

performance degradations and restore the quality of the amplified signal, various methods have been proposed and employed. These methods involve intervening on the SOA design, material, and structure and applying sophisticated line coding and interferometric and filtering configurations (see relevant references in Rizou et al., 2013). However, this is a special subject that should be separately addressed elsewhere.

25.7 SOA Electrical Gain Modulation

In addition to straightforward amplification and all-optical signal processing, there has recently been intense research interest in SOAs for use as external modulators. The motivation behind this activity is to be able to provide both data amplification and modulation so as to overcome limitations imposed by other optical modulators (Udvary and Berceli, 2010) while enabling the implementation of various applications (see relevant references in Zoiros et al., 2015). The basic configuration that is exploited for this purpose is shown in Figure 25.20 (Rizou et al., 2015b). The SOA bias current, which is fixed at some point where the *P*–*I* curve is linear (Udvary and Berceli, 2010), is modulated by an electrical data signal. This modulation alters the SOA gain analogously, which is perceived by a lightwave beam of constant power over time (CW). In this manner, the SOA current modulation is mapped on the CW signal, which thus carries the original information that has now been converted into optical form. According to this physical process, an

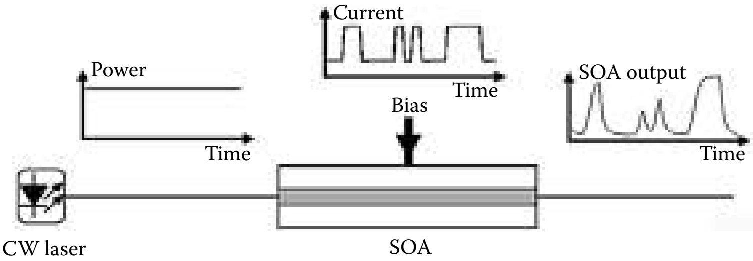


FIGURE 25.20 Basic configuration of directly modulated SOA.

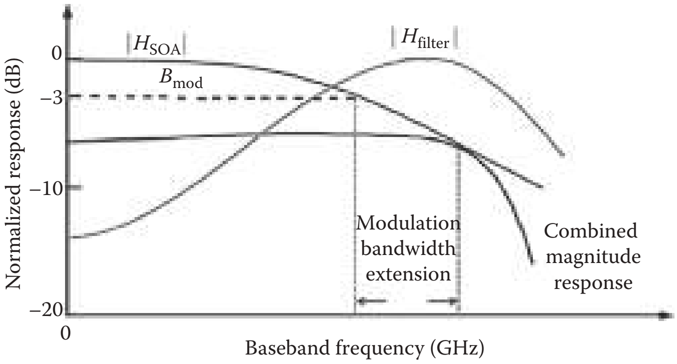


FIGURE 25.21 Schematic description of SOA electrical response characteristic $|H_{SOA}|$, transfer function of a filter $|H_{filter}|$ and the combined magnitude response.

exact copy of the applied electrical signal should be transferred at the SOA output. However, this does not happen when the SOA modulation bandwidth, B_{mod}, is smaller than the bandwidth that corresponds to the rate and format of the applied excitation, B_{exc}. More specifically, due to the SOA finite differential carrier lifetime, τ_d, B_{mod} is limited to the order of 1 GHz. In fact, the small-signal frequency response analysis procedure, where the SOA rate equations are linearized and solved analytically by assuming that there are only small perturbations around an operating point (König, 2014), reveals that B_{mod} scales inversely with τ_d according to $B_{mod} = 1/(2\pi\tau_d)$ and the SOA electrical response exhibits the low-pass filter characteristic qualitatively shown in Figure 25.21 (Wang, 2008). On the other hand, the latter parameter depends on the data format (Agrawal, 2002), and for non-return-to-zero (NRZ) pulses is $B_{exc} = B_{rep}/2$, where B_{rep} is the repetition rate. Thus, if B_{rep} is increased so that B_{exc} exceeds B_{mod}, the performance of the directly modulated SOA becomes progressively pattern-dependent and eventually poor.

25.8 Pattern Effect II: Direct Current Modulation

The characteristics of the encoded pulse stream are degraded as the rate of SOA direct modulation is increased due to the concomitant aggravation of the associated pattern effect. In fact, Figure 25.22 shows that with a five-fold increase from 1 Gb/s to 5 Gb/s the amplitude deviations between marks, between spaces, and between marks and spaces are intensified (compare left- to right-hand side). Moreover, the pseudo-eye diagram becomes asymmetrical due to the long rise and fall times and tends to close (compare left to right-hand side of Figure 25.23). Finally, the chirp suffers from analogous peak amplitude fluctuations and transient distortions (compare left to right-hand side of Figure 25.24). Nevertheless, these impairments can be mitigated and the SOA be directly modulated at an extended data rate than that enabled

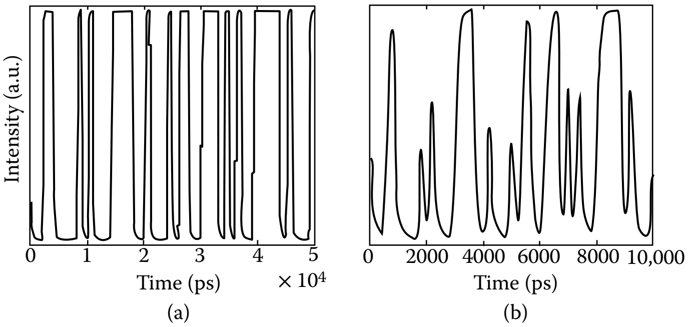


FIGURE 25.22 Encoded data pattern at directly modulated SOA output for (a) 1 Gb/s, (b) 5 Gb/s.

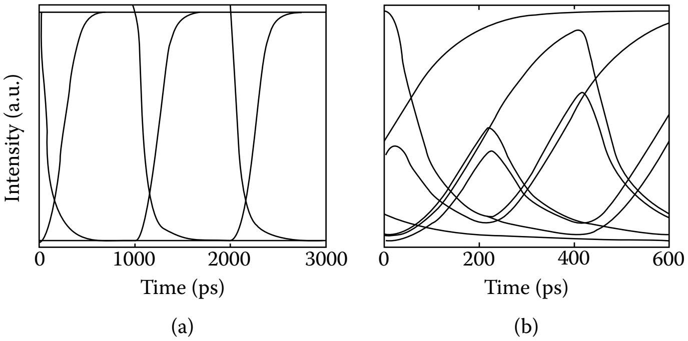


FIGURE 25.23 Pseudo-eye diagram of encoded signal at directly modulated SOA output for (a) 1 Gb/s, (b) 5 Gb/s.

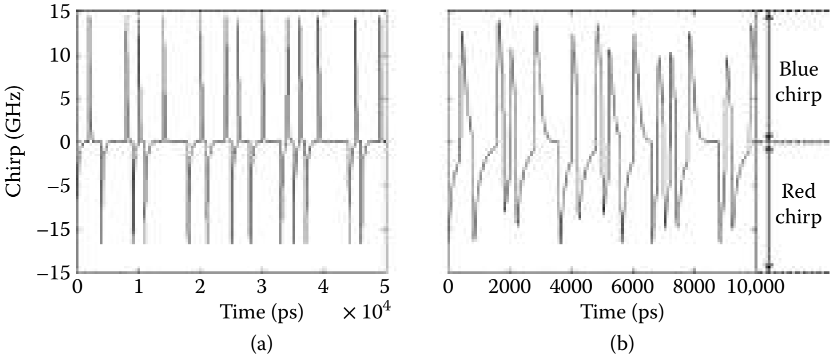


FIGURE 25.24 Chirp of encoded signal at directly modulated SOA output for (a) 1 Gb/s, (b) 5 Gb/s.

by its limited modulation bandwidth by means of post optical notch filtering. This technique can be implemented with different technologies (see relevant references in Zoiros and Morel, 2014), which have in common the fact that they produce a filter characteristic whose slope is in the opposite direction to that of the SOA, as seen in Figure 25.21. This counteracts the SOA response and so the combined transfer function becomes independent of the direct modulation frequency for a wider range than for the SOA only. In this manner, the quality characteristics of the encoded signal at the directly modulated SOA output can be significantly improved so that they are appropriate for supporting the target applications.

References

Agrawal GP (2002) *Fiber-Optic Communication Systems*. New York, NY: Wiley.

Agrawal GP and Olsson NA (1989a) Self-phase modulation and spectral broadening of optical pulses in semiconductor laser amplifiers. *IEEE J Quantum Electron*. 25:2297–2306.

Agrawal GP and Olsson NA (1989b) Amplification and compression of weak picosecond optical pulses by using semiconductor-laser amplifiers. *Opt Lett*. 14:500–502.

Borri P et al. (1999) Measurement and calculation of the critical pulsewidth for gain saturation in semiconductor optical amplifiers. *Opt Commun*. 164:51–55.

Boscolo S, Bhamber R, Turitsyn SK, Mezentsev VK, and Grigoryan VS (2006) RZ-DPSK transmission at 80 Gbit/s channel rate using in-line semiconductor optical amplifiers. *Opt Commun*. 266:656–659.

Botsiaris C, Zoiros KE, Chasioti R, and Koukourlis CS (2007) Q-factor assessment of SOA-based ultrafast nonlinear interferometer. *Opt Commun*. 278:291–302.

Cassioli D, Scotti S, and Mecozzi A (2000) A time-domain computer simulator of the nonlinear response of semiconductor optical amplifiers. *IEEE J Quantum Electron*. 36:1072–1080.

Chi JWD, Chao L, and Rao MK (2001) Time-domain large-signal investigation on nonlinear interactions between an optical pulse and semiconductor waveguides. *IEEE J Quantum Electron*. 37:1329–1336.

Connelly MJ (2002) *Semiconductor Optical Amplifiers*. Dordrecht: Kluwer Academic Publishers.

Diez S (2000) All-optical signal processing by gain-transparent semiconductor switches. PhD dissertation, Technical University of Berlin.

Eiselt M, Pieper W, and Weber HG (1995) SLALOM: Semiconductor laser amplifier in a loop mirror. *J Lightwave Technol*. 13:2099–2112.

Giller R, Manning RJ, and Cotter D (2006) Gain and phase recovery of optically excited semiconductor optical amplifiers. *IEEE Photon Technol Lett*. 18:1061–1063.

Girardin F, Guekos G, and Houbavlis A (1998) Gain recovery of bulk semiconductor optical amplifiers. *IEEE Photon Technol Lett*. 10:784–786.

Girault G et al. (2008) Analysis of bit rate dependence up to 80 Gbit/s of a simple wavelength converter based on XPM in a SOA and a shifted filtering. *Opt Commun*. 281:5731–5738.

Gutiérrez-Castrejón R and Duelk M (2007) Modeling and simulation of semiconductor optical amplifier dynamics for telecommunication applications. In: *Computer Physics Research Trends* (Bianco SJ, ed.), pp. 89–124. New York, NY: Nova Science Publishers.

Gutiérrez-Castrejón R, Occhi L, Schares L, and Guekos G (2001) Recovery dynamics of cross-modulated beam phase in semiconductor amplifiers and applications to all-optical signal processing. *Opt Commun*. 195:167–177.

Gutiérrez-Castrejón R, Schares L, Occhi L, and Guekos G (2000) Modeling and measurement of longitudinal gain dynamics in saturated semiconductor optical amplifiers of different length. *IEEE J Quantum Electron*. 36:1476–1484.

Hall KL, Lenz G, Darwish AM, and Ippen EP (1994) Subpicosecond gain and index nonlinearities in InGaAsP diode lasers. *Opt Commun*. 111:589–612.

Hamilton SA, Robinson BS, Murphy TE, Savage SJ, and Ippen EP (2002) 100 Gb/s optical time-division multiplexed networks. *J Lightwave Technol*. 20:2086–2100.

Henry CH (1982) Theory of the linewidth of semiconductor lasers. *IEEE J Quantum Electron*. 18:259–264.

Hinton K, Raskutti G, Farrell PM, and Tucker RS (2008) Switching energy and device size limits on digital photonic signal processing technologies. *IEEE J Sel Top Quantum Electron*. 14:938–945.

Hosseini SR, Razaghi M, and Das NK (2011) Analysis of ultrafast nonlinear phenomena's influences on output optical pulses and four-wave mixing characteristics in semiconductor optical amplifiers. *Opt Quant Electron*. 42:729–737.

Hussain K, Pradhan R, and Datta PK (2010) Patterning characteristics and its alleviation in high bit-rate amplification of bulk semiconductor optical amplifier. *Opt Quant Electron*. 42:29–43.

Hutchings DC, Sheik-Bahae M, Hagan DJ, and Van Stryland EW (1992) Kramers–Kronig relations in nonlinear optics. *Opt Quant Electron*. 24:1–30.

Jennen J, De Waardt H, and Acket G (2001) Modeling and performance analysis of WDM transmission links employing semiconductor optical amplifiers. *J Lightwave Technol.* 19:1116–1124.

Kim NY, Tang X, Cartledge JC, and Atieh AK (2007) Design and performance of an all-optical wavelength converter based on a semiconductor optical amplifier and delay interferometer. *J Lightwave Technol.* 25:3730–3738.

Kim Y et al. (2003) Transmission performance of 10-Gb/s 1550-nm transmitters using semiconductor optical amplifiers as booster amplifiers. *J Lightwave Technol.* 21:476–481.

König S (2014) *Semiconductor Optical Amplifiers and mm-Wave Wireless Links for Converged Access Networks*. Karlsruhe Series in Photon & Communication (Leuthold J, Freude W, Koos C, eds.), Vol. 14. Karlsruhe: University of Karlsruhe.

Manning RJ, Ellis AD, Poustie AJ, and Blow KJ (1997) Semiconductor laser amplifiers for ultrafast all-optical signal processing. *J Opt Soc Am B*. 14:3204–3216.

Mecozzi A and Mørk J (1997) Saturation effects in nondegenerate four-wave mixing between short optical pulses in semiconductor laser amplifiers. *IEEE J Sel Top Quantum Electron.* 3:1190–1207.

Morito K, Ekawa M, Watanabe T, and Kotaki Y (2003) High-output-power polarization-insensitive semiconductor optical amplifier. *J Lightwave Technol.* 21:176–181.

Mørk J and Mecozzi A (1996) Theory of the ultrafast response of active semiconductor waveguides. *J Opt Soc Am B*. 13:1803–1816.

Mørk J, Nielsen ML, and Berg TW (2003) The dynamics of semiconductor optical amplifiers: Modeling and applications. *Opt Photon News*. 14:42–48.

Mynbaev DK and Scheiner LL (2001) *Fiber-Optic Communications Technology*. New Jersey, NJ: Prentice-Hall.

Ning CZ, Indik RA, and Moloney JV (1997) Effective Bloch equations for semiconductor lasers and amplifiers. *IEEE J Quantum Electron*. 33:1543–1550.

Occhi L (2002) Semiconductor optical amplifiers made of ridge waveguide bulk InGaAsP/InP: Experimental characterisation and numerical modelling of gain, phase, and noise. PhD dissertation, Swiss Federal Institute of Technology, Zurich.

O'Mahony MJ (1988) Semiconductor laser optical amplifiers for use in future fiber systems. *J Lightwave Technol.* 6:531–544.

Pauer M, Winzer PJ, and Leeb WR (2001) Bit error probability reduction in direct detection optical receivers using RZ coding. *J Lightwave Technol.* 19:1255–1262.

Razaghi M, Ahmadi V, and Connelly MJ (2009) Comprehensive finite difference time-dependent beam propagation model of counter-propagation picosecond pulses in a semiconductor optical amplifier. *J Lightwave Technol.* 27:3162–3174.

Rizou ZV, Zoiros KE, Hatziefremidis A, and Connelly MJ (2013) Design analysis and performance optimization of a Lyot filter for semiconductor optical amplifier pattern effect suppression. *IEEE J Sel Top Quantum Electron*. 19:1–9.

Rizou ZV, Zoiros KE, Hatziefremidis A, and Connelly MJ (2015a) Performance tolerance analysis of birefringent fiber loop for semiconductor optical amplifier pattern effect suppression. *Appl Phys B*. 119:247–257.

Rizou ZV, Zoiros KE, and Houbavlis T (2015b) Operating speed extension of SOA external modulator using microring resonator. *Proceedings of 36th Progress in Electromagnetics Research Symposium*, Prague, Czech Republic, July 6–9.

Saleh BEA and Teich MC (1991) *Fundamentals of Photonics*. New York, NY: John Wiley & Sons.

Schreieck RP (2001) Ultrafast dynamics in InGaAsP/InP optical amplifiers and mode locked laser diodes. PhD dissertation, Swiss Federal Institute of Technology in Zurich.

Schubert C (2004) Interferometric gates for all-optical signal processing. PhD dissertation, Technical University of Berlin.

Singh S (2011) An approach to enhance the receiver sensitivity with SOA for optical communication systems. *Opt Commun*. 284:828–832.

Tang JM and Shore KA (1998) Strong picosecond optical pulse propagation in semiconductor optical amplifiers at transparency. *IEEE J Quant Electron.* 34:1263–1269.

Tang JM, Spencer PS, Rees P, and Shore KA (2000) Ultrafast optical packet switching using low optical pulse energies in a self-synchronization scheme. *J Lightwave Technol.* 18:1757–1764.

Toptchiyski GO (2002) Analysis of all-optical interferometric switches based on semiconductor optical amplifiers. PhD dissertation, Technical University of Berlin.

Udvary E and Berceli T (2010) Improvements in the linearity of semiconductor optical amplifiers as external modulators. *IEEE Trans Microw Theory Tech.* 58:3161–3166.

Ueno Y, Nakamura S, and Tajima K (2002) Nonlinear phase shifts induced by semiconductor optical amplifiers with control pulses at repetition frequencies in the 40–160-GHz range for use in ultrahigh-speed all-optical signal processing. *J Opt Soc Am B.* 19:2573–2589.

Vacondio F (2011) On the benefits of phase shift keying to optical telecommunication systems. PhD dissertation, Laval University of Canada.

Vallaitis T. (2010) *Ultrafast Nonlinear Silicon Waveguides and Quantum Dot Semiconductor Optical Amplifiers*. Karlsruhe Series in Photon & Communication (Leuthold J, Freude W, eds.), Vol. 7. Karlsruhe: University of Karlsruhe.

Wang J (2008) *Pattern Effect Mitigation Techniques for All-Optical Wavelength Converters Based on Semiconductor Optical Amplifiers. Karlsruhe Series in Photon & Communication* (Leuthold J and Freude W, eds.), Vol. 3. Germany: Univ Karlsruhe.

Wang J et al. (2007) Pattern effect removal technique for semiconductor optical amplifier based wavelength conversion. *IEEE Photon Technol Lett.* 19:1955–1957.

Wiesenfeld JM (1996) Gain dynamics and associated nonlinearities in semiconductor optical amplifiers. In: *Current Trends in Optical Amplifiers and Their Applications* (Lee TP, ed.), pp. 179–222. Singapore: World Scientific.

Wong WM and Blow KJ (2003) Travelling-wave model of semiconductor optical amplifier based non-linear loop mirror. *Opt Commun.* 215:169–184.

Zimmerman DR and Spiekman LH (2004) Amplifiers for the masses: EDFA, EDWA, and SOA amplest for metro and access applications. *J Lightwave Technol.* 22:63–70.

Zoiros KE, Chasioti R, Koukourlis CS, and Houbavlis T (2007) On the output characteristics of a semiconductor optical amplifier driven by an ultrafast optical time division multiplexing pulse train. *Optik.* 118:134–146.

Zoiros KE, Houbavlis T, and Moyssidis M (2005) Complete theoretical analysis of actively mode-locked fiber ring laser with external optical modulation of a semiconductor optical amplifier. *Opt Commun.* 254:310–329.

Zoiros KE, Janer CL, and Connelly MJ (2010a) Semiconductor optical amplifier pattern effect suppression for return-to-zero data using an optical delay interferometer. *Opt Eng.* 49:085005-085005-4. doi:10.1117/1.3481137.

Zoiros KE and Morel P (2014) Enhanced performance of semiconductor optical amplifier at high direct modulation speed with birefringent fiber loop. *AIP Adv.* 4:077107. doi:10.1063/1.4889869.

Zoiros KE, Morel P, Hamze M (2015) Performance improvement of directly modulated semiconductor optical amplifier with filter-assisted birefringent fiber loop. *Microw Opt Technol Lett.* 57:2247–2251.

Zoiros KE, O'Riordan C, and Connelly MJ (2009) Semiconductor optical amplifier pattern effect suppression using Lyot filter. *Electron Lett.* 45:1187–1189.

Zoiros KE, O'Riordan C, and Connelly MJ (2010b) Semiconductor optical amplifier pattern effect suppression using a birefringent fiber loop. *IEEE Photon Technol Lett.* 22:221–223.

Zoiros KE, Siarkos T, and Koukourlis CS (2008) Theoretical analysis of pattern effect suppression in semiconductor optical amplifier utilizing optical delay interferometer. *Opt Commun.* 281:3648–3657.

Index

A

Absorption, 83, 97
Abstraction reaction, 172
Acoustic phonons, interactions, 137
Adsorption reaction, 172
Advanced Physical Models of Semiconductor Devices (APSYS), 524–525
AIM, *see* Anderson impurity model
AlGaInP LEDs, 462, 467
Alloys, 237
  disorder, 264–269
  piezoelectric properties, 239, 240
  spontaneous polarization, 238
  schemes for dealing with, 237–239
AM, *see* Amplitude modulation
Amplified spontaneous emission (ASE), 647–652, 716, 760, 764–765
  noise, 633, 649
  regime, 589
  spectra, 734
Amplitude modulation (AM), 788–789
Anderson impurity model (AIM), 288–296
Antenna effect, 412–413
Antireflection (AR) coatings, 612
Antireflection resonant optical waveguides (ARROWs), 116
Antiresonant Fabry–Pérot resonators, 116
Antiresonant frequencies, 649, 653, 662
APDs, *see* Avalanche photodiodes
APSYS, *see* Advanced Physical Models of Semiconductor Devices
Arbitrary crystal orientation
  interband optical-matrix elements for QW, 386–387
  wurtzite structure with, 383–385
  zinc-blende structure with, 379–383
ARROWs, *see* Antireflection resonant optical waveguides
ASE, *see* Amplified spontaneous emission
Auger coefficients, 561, 579, 642
Auger recombination, 580, 619
  coefficients, 706–707
  process, 49, 58
Avalanche photodiodes (APDs), 283–285
Axial NW heterostructures, 402–405

B

Background refractive index, 57, 778
Backward-propagating photon densities, 594, 668
BAC model, 265
  Anderson impurity model, 287–295
  CB DOS prediction, 287
  dilute nitride CB structure based on, 254
  GaN$_x$As$_{1-x}$ alloys, electron transport, 296–297, 300, 302
  (In)GaN$_x$As$_{1-x}$ band structure calculation, 259–260
  N-related parameters of, 261
  parameterization of, 260
  parameterized explicitly via TB supercell calculations, 262
  two-band, 259, 267–269
Band-anticrossing, dilute nitride alloys, 259–264
Band-filling, 644
Bandgap shrinkage, 644
Band mixing model, 702
Band structure
  electronic, 5–7
  first Brillouin zone, 16–17
  organic materials, 203–205
  semiconductor, 7–8
Basic rate equations, 642–647
BBR, *see* Black-body radiation
Beam propagation method (BPM), 117–120, 700
Bell-shaped (Gaussian) ASE spectra, SLEDs with, 601
Bent waveguides, 123–124
Benzene, p-orbitals of, 196
Biaxial strain, nonlinearities under, 228–229
Bimolecular recombination
  coefficient, 619
  mechanisms, 479
Bipolar carrier transport models, 700
Bit error rate (BER) estimation, 751
Black-body radiation (BBR), 82, 468

Bloch equation
  in HF approximation, 57
  optical, 88, 95
  semiconductor, 55–57, 95
Bloch's theorem, 4, 5, 7, 8, 191
Blocking layers, OLED working principle, 476
Blue/green nitride LEDs, 462
Blue tunnel-junction LED, 532–536
  electrical and optical characteristics of, 530, 531
Boltzmann constant, 84, 141, 457, 478, 568, 635, 755
Boltzmann (Fermi–Dirac) distribution, GaN_xAs_{1-x} alloys, 299
Boltzmann equation, for carrier transport, 140–142
Boltzmann inversion, 491
Boltzmann-like equations, 717
Boltzmann–Peierls (BP) equation, 136, 143
Boltzmann transport equation (BTE), 35–37, 287
  drift-diffusion model, 47–48
    boundary conditions and numerical solution, 49–50
    generation–recombination models, 48–49
  energy balance model, 45–47
  hierarchical approximation, 38–39
    carrier conservation, 39–40
    collision terms, 43–44
    energy conservation, 42–43
    momentum conservation, 40–42
  hydrodynamic model, 44–45
  Monte Carlo transport simulation, 37–38
Boltzmann-type equation, 724
Booster amplifiers, 787, 788
Born–Markov approximation, 724, 733
Born–Oppenheimer approximation, 195
Botz–Kalos–Lebowitz algorithm, *see* Kinetic Monte Carlo method
Boundary condition (BC)
  Dirichlet, 49, 51
  homogeneous Neumann, 49
Boundary self-energies, 68–69
Box–Muller algorithm, 731
BPM, *see* Beam propagation method
Bragg couplings, 701
Bravais lattice, 130, 131
Brillouin zone, first
  band structure, 16–17
  crystal, 6
Broad-bandwidth SLED designs, 600–602
BTE, *see* Boltzmann transport equation
Buckminsterfullerene, 196
Bulk semiconductors, band structures of
  with arbitrary crystal orientation, 379–387
  wurtzite Hamiltonian of (001) orientation, 368–370
  Zinc-Blende Hamiltonian of (001) orientation, 366–368
Buried heterostructure (BH) semiconductor optical amplifier (SOA), 752
Burt–Foreman Hermitianization, 636

C

Carbon-carbon bond, potential energy surface for, 200
Carrier densities for MQW-based SOAs, spatial distributions of, 669
Carrier densities spatial distribution, 667–669
Carrier density, 777
  distribution, 678
  equation, 679
Carrier density modulation (CDM), 749–750
  amplitude, 763–765
  measurement setup, 762–763
Carrier density pulsation (CDP), 774
Carrier density rate equations, 615, 617–619
Carrier density spatial distribution, 618, 620, 652, 674
Carrier-dependent quasi-Fermi levels, 635
Carrier diffusion, 652–653
Carrier heating (CH), 655–656, 750, 773–774, 776, 782
Carrier injection, 760
Carrier lifetime, 777
Carrier-photon self-energy, 70
Carrier rate equation, 681
Carrier transport
  Boltzmann equation for, 140–142
  drift-diffusion, 39
  effects in MQW SOA, 653
  models, 50, 699–700
    for MQW and QD SOAs, 653–655
    for QD SOAs, 653–655
  in optoelectronic devices, 35
Carrier transport theory, in dilute nitride alloys, 287
  Anderson impurity model, 288–296
  electron transport in GaN_xAs_{1-x} alloys, 296–303
Carrier trapping, in organic device models, 208–210
Cartesian system of coordinates, 777
Cd-based chalcogenide quantum dots
  CdSe/CdTe CQDs, radiative lifetimes in, 437–438
  CdSe/CdTe Type II CQDs, absorption edge wavelength, 434–437
  CdSe QD arrays, charge carrier mobility, 442–443
  correlation energy, 438–441
  equivalent CdTe and CdTe/CdSe QD, absorption spectra, 441–442
CDM, *see* Correlated disorder model
CDP, *see* Carrier density pulsation
CdSe/CdTe CQDs, radiative lifetimes in, 437–438
CdSe/CdTe Type II CQDs, absorption edge wavelength, 434–437
CdSe quantum dot arrays, charge carrier mobility, 442–443

CdTe/CdSe quantum dot
  electron shell localization effect, 438
  equivalent, absorption spectra, 441–442
  hole shell localization effect, 438–441
CdTe quantum dot, equivalent, absorption spectra, 441–442
Centre of Integrated Photonics (CIP), 762
Charge-carrier distributions, 722
Charge-carrier dynamics, 716
Charge carrier mobility, in CdSe quantum dot arrays, 442–443
Charge conservation, for OLED, 478–479
Charge Extraction by Linearly Increasing Voltage (CELIV) measurement, 213
Charges from electrostatic potentials using a grid-based method (CHELPG) scheme, 489
Charge transfer (CT) rate, for OLED, 476, 483–484
Charge transport layers, OLED working principle, 475–476
CHCC Auger recombination, 461
"CHCC" pathway, 279
CHSH Auger recombination mechanism, 279, 283
"CHSH" pathway, 279
CIP, *see* Centre of Integrated Photonics
Circular optical waveguide, 110–112
Classical force fields, in OLED, 489
Classical particle densities, 552
Coherence function, 602–604
Coherent pulse propagation in macroscopic semiconductor devices, 739
Collision operator, 139
  generation-recombination, 141
Colloidal nanocrystal quantum dots (NQD), 420
Colloidal quantum dots (CQDs), 420
  CdSe/CdTe CQDs, radiative lifetimes in, 437–438
  CdSe/CdTe Type II CQDs, absorption edge wavelength, 434–437
Color rendering index (CRI), of white light source, 468–470
Composition-dependent band gap, 255
  bowing, in GaN_xAs_{1-x}, 256
Computer-aided design (CAD) tool, 167
Conduction band (CB), strained-layer QW
  wurtzite semiconductors, 375–376
  zinc-blende semiconductors, 371
Configuration interaction (CI) method, 420
Confinement factor and taper angle, 710–712
Continuity equation, in superlattices, 236–237
Conventional InP-based lasers, nonradiative recombination processes in, 277
Conventional semiconductor alloys, 254
  alloy composition in, 255
  versus highly mismatched alloys, 256
Core-shell NW, *see* Radial semiconductor NW heterostructures
Core/shell QD structures, 420–421
Corpuscular theory of light, 81
Correlated color temperature (CCT), white light source, 468–470
Correlated disorder model (CDM), 486
Coulomb integrals
  NQD, configuration interaction Hamiltonian, 429–430
  PW implementation, quantum dots, 423–424
Coulomb interaction, 8, 9
  matrix element, 723
Coulomb repulsion, 245
Counter propagating photon density distributions, 675–677
Courant–Friedrichs–Lewy (CFL) condition, 151, 678
Covariance matrix, 41
C-plane Hamiltonian, wurtzite semiconductors, 368–370
c-plane LED, simulation result of, 576
Cross-gain modulation (XGM)-based wavelength converter, 623
Crystal
  first Brillouin zone, 6
  photonic, 116
  temperature, 131
  vibrations and phonons, 131–136
Crystal coordinate system (CCS), of light-emitting diodes, 453
Crystal lattice, 129–131
  Lagrangian of, 133
Crystalline semiconductors, 191–193, 480
Crystal-phase bandgap engineering, semiconductor nanowires, 410–412
Crystal-phase quantum structures, 410, 412
Crystals
  vector, 223
  wurtzite
    crystal direction, 224
    piezoelectric tensor in, 223
  zincblende
    crystal direction, 224
    piezoelectric tensor in, 223
CSUPREM process simulation, 178, 186
Current crowding, in LED dice, 466–467
Current spreading, in LED dice, 464–466
Current-voltage characteristics, OLED case studies, 495

D

DBRs, *see* Distributed Bragg reflectors
DCV4T, *see* Dicyanovinyl-substituted quaterthiophe
DD, *see* Drift-diffusion
DDM, *see* Drift-diffusion one
Deal–Grove model, oxidation, 184
Delay-differential-equation model, 716, 717–718, 731
Density functional perturbation theory (DFPT), 232

Density functional theory (DFT), 8–9, 16, 195, 254
Density gradient quantum correction model, 543
Density matrix (DM), 36, 59
Density of states (DOS)
  in dilute nitride alloys, 287
    Anderson impurity model, 288–296
    electron transport in GaN_xAs_{1-x} alloys, 296–303
  MQW model, 525
  for OLED, 485–487
Desorption reaction, 173
Deterministic approach, 647–648
Detuning frequency, 759
Device geometry and simulation parameters, 706–707
DFB laser, *see* Distributed feedback laser
DFT, *see* Density functional theory
Dicyanovinyl-substituted quaterthiophe (DCV4T), 493, 494
Dielectric confinement effect, quantum dots, 420, 424–426
Dielectric constant, of cubic crystals, 458
Diffusion constant, OLED master equation, 481–482
Diffusion model, impurity, 181
Dilute nitride alloys, 252–254, 303–305
  alternative applications of, 281
    avalanche photodiodes, 283–285
    electro-absorption modulators, 283–285
    GaAs-based 1.55-μm quantum well lasers, 281–283
    III-V lasers on silicon, monolithic integration, 286–287
    semiconductor optical amplifiers, 283–285
  density of states and carrier transport theory in, 287
    Anderson impurity model, 288–296
    electron transport in GaN_xAs_{1-x} alloys, 296–303
  electronic structure theory of
    band-anticrossing, 259–264
    first principles and empirical pseudopotential calculations, 254–258
    tight-binding and alloy disorder, 264–269
  TB model for, 261
Dilute nitride quantum well lasers, theoretical model for, 270–271
Dirac delta function, 651
Dirac equation, Foldy–Wouthuysen transformation of, 26
Direct current modulation, 792–793
Direct finite-difference time domain (FDTD) simulations, 466
Direct optical amplification
  applications, 787–788
  chirp variation, 789–790
  gain variation, 790–791
  optical signal-to-noise ratio, 790
  pattern-dependent output distortion, 787–788
  pattern effect, 788
  pulse profile and amplitude wandering, 789
  Q-factor, 790–791
  SOA gain in response to data train input, 787–788
Direct Simulation Monte Carlo scheme (DSMC), 160
Dirichlet boundary condition, 49, 51, 640
Discrete Fourier transform, 760
Distributed bias current, 656
Distributed Bragg reflectors (DBRs), 50, 315, 697
Distributed feedback (DFB) laser, 661, 697
Dopant diffusion, 181
  active impurities, 183
  inactive impurities, 183
  interstitials, 182
  neutral impurities, 183
  vacancies, 181–182
Dopants
  dielectrically enhanced ionization energies of, 409–410
  nanowires, discrete nature of, 407–409
Doping, semiconductor nanowires
  continuous, homogeneous doping-related background charge, 406–407
  dielectrically enhanced ionization energies of dopants, 409–410
  random dopant fluctuations, 407–409
DOS, *see* Density of states
Dot-in-a-well (DWELL) structures, 719, 724
Drift-diffusion (DD)
  carrier transport, 39
  equations, for OLED, 478–480
  model, 58, 142, 546, 549, 551
    Boltzmann transport equation, 47–48
    energy balance, 46–47
    generation–recombination models, 48–49
  quantum-corrected, *see* Quantum-corrected drift-diffusion
Drift-diffusion one (DDM), 456–457
Drift-diffusion solver, random alloy fluctuation with, 568
Drude model, 644
DSMC, *see* Direct Simulation Monte Carlo scheme
Dual Pearson model, ion implantation, 180–181
Dynamic models, 659–660
Dynamic Monte Carlo, *see* Kinetic Monte Carlo method
Dynamic propagation model, 677–678
  upwind scheme numerical implementation, 678–680
Dynamics and pattern effects, SOA
  free carrier distribution, 773–774
  interband transitions, 773
  pump-probe technique, 775
  schematic representation of, 772
  and signal operating conditions, 776
Dynamic TW-SOA models, 656–657, 659–660

E

EAMs, *see* Electro-absorption modulators
EBE, *see* Effective Bloch equations
EBL region, *see* Electron-blocking layer region
ECDM, *see* Extended correlated disorder model
EET, *see* Electronic excitation (energy) transfer
EFA, *see* Envelope function approximation
Effective Bloch equations (EBE), 751–752
Effective diffusion coefficient, 699
Effective index method (EIM), 114, 115, 117
Effective-mass approximation, 92
EGDM, *see* Extended Gaussian disorder model
EIM, *see* Effective index method
Einstein coefficients, 82–84
Elasticity, linear, 239, 242, 542
Electrical equations, solver for, 704–705
Electrical gain modulation, 791–792
Electrical-to-optical conversion efficiency, 707
Electric field energy density, 730
Electric field envelope model, 643–645
Electric field propagation, 717–718, 719, 729
Electric potential distribution, in LED structure, 457
Electro-absorption modulators (EAMs), 283–285
Electrodes, OLED working principle, 474–475
Electroluminescence, of white OLED, 503–506
Electromagnetic environment role, 99–101
Electromechanical coupling
  light-emitting diodes, 454–456
  in wurtzite semiconductors, 239–240
Electron-blocking layer (EBL) region, 458, 504, 530, 561
Electron density, periodic, 5
Electron-electron interaction, 9, 15
Electron gas, drifted Gaussian distribution, 41
Electron-hole pair, 35, 141, 205–206, 271, 274, 315, 425, 504
Electron-hole recombination, for OLED, 484–485, 601
Electronic band, 634
  materials, 8–9
  structure, 5–7
  wurtzite InN, 25
Electronic coupling elements, for OLED, 488, 491
Electronic excitation (energy) transfer (EET), 485
Electronic materials, piezoelectricity in, 220–222
Electronic polarization, tunnel-junction light-emitting diodes, 527
Electronic structure, organic semiconductors, 194–196
Electronic transport in NW LED, 551
Electron orbital of fullerene cage, 197
Electron-photon interaction
  electromagnetic environment, 99–101
  light field, quantization of, 97–99
  solid-state system, 90–97
  in two-level system, 82–90
Electron quantum densities, 549, 550
Electrons, in valence bands, 138
Electron temperature, 41
Electron transporting layers (ETLs), 474
Emission
  layer, OLED working principle, 476–477
  spectrum
    influence of indium fluctuation on, 574–576
    shift, 561
  stimulated, 83
  wavelength, 4
Empirical pseudopotential method (EPM), 28–29, 254, 259–264
  spin-orbit coupling, 29–30
Empirical TB (ETB) model, 21, 22, 25, 26, 28
Empirical tight-binding (ETB) method, 20–26, 543, 548
  approach, 547–549
  Hamiltonian of periodic system, 20
  matrix elements, 23–24
  model, 21, 22, 25, 26
  Slater and Koster approach, 23
  spin-orbit coupling, 26–28
Empty lattice model, 5, 6
Energetic disorder, 486, 492, 493, 504
Energy balance (EB) model, Boltzmann transport equation, 45–47
Energy bands, in silicon, 138–139
Energy conservation expression, 701
Energy-dependent resonant scattering rate, GaN_xAs_{1-x} alloys, electron transport in, 298
Energy model, 142
Energy transfer, for OLED, 485
Energy-transport model, and numerical discretization, 146–152
Envelope-function approaches, 17–20
Envelope function approximation (EFA), 543, 545–547
Epitaxial coordinate system (ECS), of light-emitting diodes, 453
EPM, *see* Empirical pseudopotential method
Equilibrium equations, 565–566
Equilibrium Green's function (EGF) theory, 60
ER, *see* Extinction ratio
ETB, *see* Empirical tight-binding method
Ethene molecule, wave functions of, 195, 196
ETLs, *see* Electron transporting layers
Excitons, 214, 215
  for OLED, 479–480
Extended correlated disorder model (ECDM), 486, 493, 494
Extended Gaussian disorder model (EGDM), 486, 494, 497
Extended rate equation model, 652–656
Extended zone scheme, electron band structure, 6

External quantum efficiency (EQE), of light-emitting diodes, 452–453
Extinction ratio (ER), 284, 591, 625, 751, 788, 790

F

Fabry–Pérot resonators, 116
Fabry–Pérot SOA (FP-SOA), 613, 632
Far-field (FF) pattern, 698
FCA, *see* Free carrier absorption
FDBPM, *see* Finite-difference beam propagation method
FDCNP-BPM, *see* Finite difference, Crank–Nicholson paraxial BPM algorithm
FDCNWA-BPM, *see* Finite difference Crank–Nicolson wide angle beam propagation method
FDTDM, *see* Finite difference time domain method
Fermi–Dirac distribution function, 96, 208, 299, 569, 635, 636, 755
Fermi–Dirac integrals, 48, 656
Fermi distributions, 553, 774
Fermi's Golden Rule, 37, 45, 196
FF pattern, *see* Far-field pattern
Finite-box (FB) approach, 50
Finite charge-carrier densities, 481
Finite-difference beam propagation method (FDBPM), 781
Finite difference, Crank–Nicholson paraxial BPM algorithm (FDCNP-BPM), 120
Finite difference Crank–Nicolson wide angle beam propagation method (FDCNWA-BPM), 122
Finite difference method (FDM), 107
Finite-difference (FD) technique, 49–50
Finite difference time domain method (FDTDM), 117
Finite-element (FE) technique, 49–50
Finite impulse filter (FIR) theory, 751–752
Finite-impulse response (FIR) filter scheme, 660
First-principles molecular dynamics (FPMD), 199
Fit functions for scattering rates, 727–728
Fluctuation-dissipation theorem, 651
Foldy–Wouthuysen transformation, of Dirac equation, 26
Förster theory, 45
Fourier transform (FT), 424, 603, 628, 659, 734, 757, 760
Four-wave mixing (FWM)
  analytic model, 627–628
  CDM characteristics, 758–762
  efficiency, 628
  experimental procedure, CDM measurement, 762–763
  operating conditions, 763–766
  semiconductor laser diodes (SLDs), 757
FPMD, *see* First-principles molecular dynamics
Franck–Condon principle, 199
Free carrier absorption (FCA), 627, 750, 773–774
Free-carrier absorption losses, 702
Free-standing semiconductor NWs, 397
Fullerene, 196
  delocalized electron orbital of, 197
  electron-conducting, 202, 203
Full second-order piezoelectric tensor formalism, 229–232
"Fullwave" model, 659
Full width at half-maximum (FWHM), 576, 603, 624, 633, 711, 784
  of emission spectrum, 575
  pulse, 623
FWHM, *see* Full width at half-maximum
FWM, *see* Four-wave mixing

G

GaAs-based laser structure, 7–8, 252
GaAs-based 1.3-μm InGaNAs quantum well lasers theory, 269–270
  dilute nitride quantum well lasers theoretical model, 270–271
  nitrogen incorporation impact on
    band structure and gain of quantum wells, 271–277
    on carrier recombination, 277–281
GaAs-based 1.55-μm quantum well lasers, 281–283
GaAs/GaP core-shell NWs, 401
Gain and group-velocity dispersion, 777
Gain compression, 756, 776–777, 779, 782
Galliumarsenide (GaAs)/Aluminiumarsenide (AlAs) microcavity, 100
Gallium nitride (GaN) growth, gas phase reaction, 171, 172
GaN-based LEDs, efficiency droop IN, 524
GaN-based visible light-emitting diodes, 58
GaN NW, 403–405
$\text{GaN}_x\text{As}_{1-x}$
  bond lengths analysis in, 257
  CB structure in, 259
  composition-dependent band gap bowing in, 256
  electronic properties of, 265
  electronic structure analysis of, 257
  N-related localized states in, 257
  theoretical model of, 267
  unusual pressure dependence of, 257
$\text{GaN}_x\text{As}_{1-x}$ alloys, electron transport in
  Boltzmann (Fermi-Dirac) distribution, 299
  carrier distribution functions, 301–302
  energy-dependent resonant scattering rate, 298
  negative differential velocity, 296–297
  N-related localized states, 297–299
  quantitative analysis of, 300–301
  single-electron Monte Carlo model, 302–303
$\text{GaN}_x\text{As}_{1-x}\text{P}_x$
  bond lengths analysis in, 257
  electronic structure analysis of, 257

Gap materials, semiconductor band structure, 7–8
Gas, viscosity and thermal conductivity, 173
Gaussian broadening coefficient, 563
Gaussian disorder model (GDM), 486
  parametrization of, 491–493
Gaussian distribution, 720
  of energy levels, 486
  of states, 204–205
Gaussian noise process, 647
Gaussian random variable, 652
Gauss model, ion implantation, 179
Gauss's law, 104
GDM, *see* Gaussian disorder model
GEA, *see* Gradient expansion approximation
Gell–Mann and Low theorem, 59
Geminate recombination, electron, 214
Generalized gradient approximation (GGA), 14
Generation-recombination (GR)
  model, drift-diffusion, 48–49
  transitions, 48
Generation–recombination collision operator, 141
Genuine quantum approach, 58–59
  contour-ordered NEGF, 59–67
  self-energies
    boundary, 68–69
    scattering, 69–72
Geometry and epitaxial design, SLED, 591–592
Ge-Si core-shell nanowire (NW), 401
GGA, *see* Generalized gradient approximation
Gillespie algorithm, *see* Kinetic Monte Carlo method
Gradient expansion approximation (GEA), 13
Graphene, thermal properties
  numerical results, 160–163
  semiclassical description, 152–159
  simulation scheme, 159–160
Green LEDs, carrier transport in, 580–584
Green's function, 59, 60
  chronological and antichronological, 65
  with self-energies, 72
  two-particle, 64
  in weakly confined systems decay, 69
Green's theorem, 39
Green tunnel-junction LED, 535–538
Gummel's method, 215

H

Half-width at half maximum (HWHM), 603, 604
Hamiltonian matrix, 21, 266, 267
  element, ETB method, 24
Hamiltonian of solid-state system, 91
Hamiltonian of two-level system, 86–87
HAROLD 4.0, 703, 704, 706
Hartree–Fock approximation, 94–96
Hartree–Fock (HF) theory, 11, 14, 56
HD model, *see* Hydrodynamic model
Heat flow equation, 705
Heat transfer, in LED dice, 464–466
Heat transport
  macroscopic balance equation for, 142–144
  in semiconductor, 140
Heavy-hole (HH) band, 19, 262, 379, 461
Heisenberg equation
  motion, 56, 64, 86, 94
  polarization, 87
Heisenberg picture, two-level system in, 86–90
"HELLISH" devices, 284
Helmholtz equations, 593, 597, 640, 700
Henderson theorem, 491
Henyey–Greenstein scattering functions, 469
Heyd–Scuseria–Ernzerhof hybrid functional approach, 8, 16
HF theory, *see* Hartree–Fock theory
HH band sub-bands, 639
Highest occupied molecular orbital (HOMO), 203, 204, 207, 476, 505
Highly mismatched semiconductor alloys, 255
  conventional semiconductor alloy versus, 256
High-power SLED designs, 604–605
Hohenberg–Kohn theorem, for nondegenerate ground states, 9–10, 31
Hole/electron blocking layers, OLED working principle, 476
Hole transporting layer (HTL), 474
HOMO, *see* Highest occupied molecular orbital
Homogeneous Neumann boundary condition, 49
Homogeneous wave equation, 99
Hondros–Debye equation, 110, 112
Hopping matrix element, 23, 24
HSE, *see* Screened hybrid functional
HTL, *see* Hole transporting layer
Huawei Marine Networks, 631
HWHM, *see* Half-width at half maximum
Hybrid approaches, 487
Hybrid functionals approach, 14–16
"Hybrid" OLED, 503
Hydrodynamic (HD) model
  Boltzmann transport equation, 44–45
  nonstationary effects, 39
Hydrogen atom, 194

I

Ideal QW, 279, 333, 576, 580
Ideal R-SLED, 596–597
Ideal standard SLED, 596
IFD-BPM, *see* Improved finite-difference beam propagation model
IGVs, *see* Initial guess values

Impedance spectroscopy, OLED case studies, 496–498
Improved finite-difference beam propagation model (IFD-BPM), 660
Impurities, incomplete ionization of, 526–527
Impurity diffusion model, 181
Index guiding waveguides, 113
In-diffused waveguide, 114
Indium, composition distribution of the InGaNMQW, 174–176
Indium fluctuation
  on emission spectrum, influence of, 574–576
  influence of, 576
Indium gallium arsenide (InGaAs), 759
Infinite hierarchy problem, 94
In-fluctuation QW, 580
Infrared (IR) LEDs, 452, 453
InGaAs quantum dot (QD) lasers, development of, 252
InGaNAs/GaAs alloys, limitations of, 282
InGaN-based LEDs, 452, 462, 463, 467, 528
InGaN multiple quantum well growth stimulation, 174–178
(In)GaN$_x$As$_{1-x}$
  band structure calculation, 259–260
  CB structure of, 264
Initial guess values (IGVs), 665, 703
Inorganic semiconductors, 91
  band structure, 203, 204
  versus organic semiconductors, 193, 214
Input power dynamic range (IPDR), 788
Integrated MOPAs, 697, 698
Interface charges, tunnel-junction light-emitting diodes, 527
Internal losses coefficient, 702–703
Internal quantum efficiency (IQE), 560
  curves, 579
  droop behavior of, 578–580
  in LEDs, 232–233
Intervalence band absorption (IVBA), InGaNAs lasers, 283
Intraband transitions approximation, 92
Inverse Fourier transform, 659, 757
Inverse Monte Carlo, 491
Ion implant models, 178–179
  Gauss model, 179
  Pearson model, 179–180
    dual, 180–181
IPDR, *see* Input power dynamic range
IQE, *see* Internal quantum efficiency
Isolated nitrogen states, linear combination of, 264–269
Isotropic quadratic approximation, 138
I-V curve, 570–573, 576–578, 582, 583, 585

J

Jaynes–Cummings model, 98, 99
Jellium model, 92–93

K

Kinetic Monte Carlo (KMC) method, 482–483
Kinetic theory, 173
KMC lifetime simulations, 507, 509–511
KMC method, *see* Kinetic Monte Carlo method
Kohn–Sham equation, 9, 12, 13
Kohn–Sham theory, 10–14
k·p Hamiltonian, PW implementation, 423–424
k·p method for SP electronic structure, quantum dots, 421–423
k·p models, 52, 259–264
k·p theory, 17–20, 274, 422
Kramers–Kronig transformation, 748, 756, 783

L

Lagrange equation motion, 133
Lagrangian of crystal lattice, 133
Landé effective g factor, 267, 323
Langevin noise source, 651
Laser operation, 86, 315, 338
Lattice
  commensurability constraint, 384
  crystal, 129–131
Lattice-mismatched heteroepitaxy, 221
Lattice-mismatched heterostructures, 403, 404
Lattice temperature, 776
Layer transitionmetal dichalcogenides (TMDCs), 152
LCAO, *see* Linear combination of atomic orbitals
LCINS approach, *see* Linear combination of the isolated nitrogen states approach
LDA, *see* Local density approximation
LDOS, *see* Local density of states
Leaky mode, optical waveguide, 116
LEDs, *see* Light-emitting diodes
LEE, *see* Light extraction effiency
LEF, *See* Linewidth enhancement factor
Lennard–Jones potential model, simulation, 169
L-I characteristics of reflecting SLEDs, 593–597
Light conversion, by phosphor, 468–470
Light emission efficiency, light-emitting diodes, 462–463
Light-emitting devices, carrier transport and optical transition, 35
Light-emitting diode (LED) dice
  current spreading, heat transfer, and light extraction in, 464–466
    current crowding, 466–467
    surface recombination, 467

  phosphor-converted white light emission, 467–468
    light conversion by phosphor, 468–470
    white light characteristics, 468
Light-emitting diodes (LEDs), 35, 219
  carrier transport and recombination in, 453–454
    electromechanical coupling, 454–456
    light emission efficiency, 462–463
    recombination models, 458–461
    transport equations, 456–458
  fabrication of, 237
  family of, 452–453
  GaN-based visible, 58
  multiple quantum well, 174
  simulation of, 451
Light extraction efficiency (LEE), 462, 464
Light extraction, in LED dice, 464–466
Light-hole (LH) band, 19, 262, 321
Light–matter coupling, in semiconductor, 92
Light–matter interaction, 81, 86, 101, 154, 717, 748, 755–756
Light, quantization, 97–99
Linear amplification applications, 731
Linear chain, monoatomic and biatomic, 132
Linear combination of atomic orbitals (LCAO), 543
Linear combination of the isolated nitrogen states (LCINS) approach, 254, 265–269
Linear continuum elasticity theory, 399–400
Linear elasticity, 239, 242
Linear piezoelectric coefficients, of III-N semiconductors, 227–228
Linear piezoelectricity, electromechanical coupling with, 240–243
Linewidth enhancement factor (LEF), 615, 702, 748, 783
Lobatto IIIA formula, 684
Local density approximation (LDA), 8–9
  generalized gradient approximation approach, 14
  gradient expansion approximation approach, 13
  Hedin's GW approximation approach, 16
  hybrid functionals approach, 14–16
  Kohn–Sham theory and, 10–13
Local density of states (LDOS), 59, 552
Longitudinal waves, 131
Löwdin orbitals, 21, 22
Löwdin transformation, 21
Lowest unoccupied molecular orbital (LUMO), 203, 204, 207, 476
Luttinger–Kohn Hamiltonian, zinc-blende semiconductors, 366–368
Luttinger–Kohn perturbation theory, 51

M

Macroscopic balance equation, for heat transport, 142–144
Macroscopic evolution equation, closure of system of, 144–146
Macroscopic material polarization, 749
Many-body effects, non-Markovian gain model with, 387–389
Many-body perturbation theory (MBPT), 16
Marcus theory, 197
Martin–Schwinger hierarchy, 64
Master oscillator (MO), 697
Material gain, 634–636
Material parameters, sensitivity to geometrical and, 553–556
Material susceptibility, 701–702
Matsubara Green's function, 60, 289
Maximum entropy principle (MEP), 144–145
Maxwell–Bloch approach, 720
Maxwell–Bloch equations, 88, 716, 721, 722
Maxwell–Bloch framework, 717
Maxwell's equation, 35, 90, 103, 466, 752
Maxwell's wave equation, 716
MBPT, *see* Many-body perturbation theory
MEP, *see* Maximum entropy principle
Merz–Singh–Kollman scheme, 489
Metalorganic chemical vapor deposition (MOCVD), reactor simulation, 167–168
  chemical reaction, 170–172
  inGaN multiple quantum well growth stimulation, 174–178
  surface-gas reactions and film deposition, 172–173
  transport equation, 168–170
Method of moments (MOM), partial differential equation, 38
Microscopic Auger recombination processes, 461
Microscopic charge transport, in organic semiconductors, 196–199
Microscopic polarizations, 749
Mie theory, 468–469
Miller–Abrahams rates, 486
MO, *see* Master oscillator; Molecular orbital
Mobility
  OLED master equation, 481–482
  in organic semiconductors, 205–206
Model formulation, SOA
  carrier dynamics, 776–777
  finite-difference beam propagation method, 781
  linearly polarized light, electrical field, 778
  normalization coefficient, 779
  picosecond optical pulses, 780
  spatio-temporal grid, 780–781
  traveling wave amplifier, 777
Modeling nonlinear effects, challenges, 751–752

Modeling of surface states, 552–553
Model simplification, SOA, 781–783
Modified Euler method, 621
Modulation bandwidth, analytical formulae for, 687–688
Molecular dynamics (MD), organic semiconductors, 199–201
Molecular orbital (MO), 195
Møller–Plesset perturbation theory, 15
Monomolecular degradation, OLED, 507
Monte Carlo simulation technique, 37–38, 178, 287
Morphology simulations, in OLED, 490
Motion
  Heisenberg equation, 56, 86, 94
  Lagrange equation, 133
MQW SOA, 671
  carrier transport effects in, 653
M-shaped ASE spectra, flattening of, 602
M-shaped (Double-Humped) ASE spectra , SLEDs with, 601–602
Multiband effective mass approximation, 19
Multiband envelope function method, 19, 52
Multiband kxp method, *see* kxp perturbation theory
Multiband model, 635–636
Multiple quantum well (MQW), 633
  active region, 614
  carrier transport model for, 653–655
  InGaN, growth stimulation, 174–178
  LEDs, 462, 463
  model, effective mass approximation, 525–526
  structure, 374
Multi-QW NW heterostructures, 402–403

N

Nanocrystal quantum dots (NQD), configuration interaction Hamiltonian, 426–430
Nanorods, 397, 541
Nanoscale random alloy fluctuations, 571
Newton–Raphson algorithm, 705
Newton–Richardson approach, 49
Newton's method, 215
n-fold way algorithm, *see* Kinetic Monte Carlo method
n-GaN/i-InGaN/n-GaN structures, electron transport in, 569–573
n-GaN/InGaN/p-GaN tunnel-junction structures, 528–530
n-GaN/p-GaN tunnel-junction structures, 528–530
Nitride-based blue LED, 560
Nitride LEDs, 452, 453, 462
Nitride SQW LED structures, 463
Noise correlation properties, 730
Noise figure (NF), 652
Nondegenerate ground states, Hohenberg–Kohn theorem for, 9–10
Nonequilibrium Green's function (NEGF)
  algorithm, 72
  contour-ordered, 59–60
    Dyson's equation, 65, 67
    Schwinger–Keldysh contour, 63
    S-matrix operator, 61
  theory, 36, 59
Nonlinear gain suppression, 653, 675, 678
Nonlinear intraband processes, 624
Nonlinearities in III-N semiconductors, 228–229
Nonlinear phenomena in SOAs
  and applications
    bit error rate estimation, 751
    carrier density modulation, 749–750
    carrier heating, 750
    optical–electrical–optical conversion, 751
    optical Kerr effect, 750
    optical performance monitoring, 751
    second harmonic generation, 750
    spectral hole burning, 750
    two-photon absorption, 750
    wavelength conversion, 751
  improved modeling of
    dielectric response, 748–749
    electrical model, 754–755
    light–matter interactions, 755–756
    numerical implementation, 757
    optical model, 752–754
    time-domain model, material dispersion in, 756–757
Nonlinear piezoelectricity, electromechanical coupling with, 243–245
Non-Markovian gain model, with many-body effects, 387–389
Nonpolar orientations, in wurtzite semiconductors, 245
Nonradiative Auger recombination, 354, 453, 460–461, 560
Nonradiative decay, 500, 506
Nonradiative recombination, 461, 463, 760
Nonradiative recombination processes, in conventional InP-based lasers, 277
Nonradiative Shockley–Read–Hall (SRH) recombination, 458–459
Nonstationary effects, HD model, 39
Normalized filtering function, 650
Numerical discretization, energy-transport model and, 146–152

O

OED, *see* Oxidation enhanced diffusion
OFAs, *see* Optical fiber amplifiers
OLEDs, *see* Organic light-emitting diodes
One dimensional (1D) oxidation model, 184, 185
One-dimensional wave equation, 717

1.5-μ InGaAsP/InP tapered amplifier, 706–712
OPD, *see* Optical path length difference
OPLS, *see* Optimized potential for liquid simulations
Optical amplifier devices, 716–717
Optical anisotropy of semiconductor nanowires, 412–413
Optical Bloch equations, 88, 95
Optical–electrical–optical (OEO) conversion, 751
Optical emission spectra, 553, 554
Optical equations, solver for, 705
Optical fiber amplifiers (OFAs), 611, 612
Optical gain model, with many-body effects, 387–389
Optical gain modulation, SOA
  amplified data pulse trains, 785–786
  chirp magnitude, 785
  eye diagrams, 786–787
  pulse energy, 784
  small-signal gain, 784
Optical Kerr effect, 750
Optical models, 700–701
Optical momentum matrix elements, 377, 378
Optical path length difference (OPD), 601, 604
Optical performance monitoring, 751
Optical process, in two-level system, 82–84
Optical solvers, coupling carrier transport models with, 53–55
Optical waveguide, 103
  antireflection resonant, 116
  beam propagation method application to, 117, 118
  circular, 110–112
  effective index method, 114, 115, 117
  modeling, 107
  in rectangular coordinate system, 104
  rib-loaded, 114
  slab, 108–110, 115, 122
  theory, 103
Optimized potential for liquid simulations (OPLS), 200, 489
Optoelectronic devices, carrier transport in, 35
Optoelectronics, 4, 90, 220
Orbitals, Löwdin, 21, 22
Organic devices
  models, carrier trapping in, 208–210
  role of excitons in, 214–215
  simulations, numerical aspects, 215–216
Organic light-emitting diodes (OLEDs), 192, 193, 473–474
  case studies
    current-voltage characteristics, 495
    degradation, 506–511
    efficiency, 499–502
    electroluminescence of white OLED, 503–506
    impedance spectroscopy, 496–498
  density of states, 485–487
  drift-diffusion equations, 478–480
  electronic coupling elements, 488
  kinetic Monte Carlo, 482–483
  master equation, 480–482
  morphology, 488–490
  rates
    charge transfer, 483–484
    electron-hole recombination, 484–485
    energy transfer, 485
  reorganization energy, 487–488
  scale bridging
    Gaussian disorder models, parametrization of, 491–493
    stochastic models, 490–491
    tabulated mobilities, 493–494
  working principles of
    blocking layers, 476
    charge transport layers, 475–476
    electrodes, 474–475
    emission layer, 476–477
Organic materials, band structure, 203–205
Organic semiconductors, 191–193
  charge generation process in, 215
  device models for, 207
  electronic structure, 194–196
  inorganic semiconductors versus, 193, 214
  microscopic charge transport in, 196–199, 207
  mobility in, 205–206
  molecular dynamics, 199–201
  purity and doping, 201
Organic solar cell device models, 208
Ornstein–Uhlenbeck processes, 729
Oxidation enhanced diffusion (OED), 185
Oxidation model, 184–185

P

Partial differential equations (PDEs), 717
  energy, momentum, and density, 44
  infinite hierarchy of, 39
  method of moments, 38
Pauli exclusion principle, 160, 206, 483
PC-LEDs, *see* Phosphor-converted LEDs
PCW, *see* Photonic crystal waveguide
PDEs, *see* Partial differential equations
Pearson model, ion implantation, 179–180
Perfectly matched layers (PML), 705
Periodic electron density, 5
Perturbative approach, 487
Phonon(s), 70, 129
  crystal vibrations and, 131–136
  interactions of, 137
  momentum densities, 143
  transport equation, 136
Phonon–carrier interaction, 138–140
Phonon–electron scatterings, 139
Phosphide LEDs, 452, 453, 462

Phosphor-converted LEDs (PC-LEDs), 467–468
Phosphor-converted white light emission, 467–468
  light conversion by phosphor, 468–470
  white light characteristics, 468
Photon, 70, 81, 86, 141, 215
Photon densities for MQW-based SOAs, spatial distributions of, 669
Photon density model, 645–646
Photonic crystal (PC), 116
Photonic crystal waveguide (PCW), 116
Photonic integrated circuit (PIC) technology, 612
Photonics-based approach, 286
Photon momentum approximation, 92
Photon recycling process, 701
Picosecond optical pulses, 660, 777, 780
Piezoelectric coefficient, of wurtzite III-N semiconductors, 226
Piezoelectricity
  effect, 220
  in electronic materials, 220–222
  linear, electromechanical coupling with, 240–243
  nonlinear, electromechanical coupling with, 243–245
Piezoelectric polarization, 219, 382, 544, 567
  alloys, 239
  of semiconductor nanowires, 405–406
Piezoelectric tensor, 222–223
  formalism, full second-order, 229–232
  in zincblende crystals, 223
Planar heterostructures, 398, 402, 403, 405
Planck's law, 84
Plane wave (PW) implementation, **k**·**p** Hamiltonian and Coulomb integrals, 423–424
PLED, *see* Polymer LED
PML, *see* Perfectly matched layers
Poisson/drift-diffusion calculation, 547
Poisson effect, 403
Poisson–Schrödinger solver, 50
Polarization
  Heisenberg equation, 87
  piezoelectric, 219
    alloys, 239
  spontaneous
    alloys, 238
    III-N semiconductors, 233–236
    wurtzite crystals, 226
  strain-induced, 224–226
    wurtzite crystals, 225–226
    zincblende crystals, 224–225
Polarization-assisted tunneling effect, 523–524
Polarization controllers, 762
Polarization-dependent ASE output spectral density, 620
Polarization-induced electric field, 555
Polarization insensitivity, 636–642
Poly(3-hexylthiophene-2,5-diyl) (P3HT), 202
Polyaniline, 192, 194
Polymer LED (PLED), 473–474
P-orbitals, 195–196
Primitive cell, for two-dimensional lattice, 130, 131
PROCOM simulation, 167, 170
Propagators, 59
p-type doping, 476, 530
Pulse amplification analytic model, 623–626
Pulse photon energy, 773
Pump and probe, 622, 627–628, 759, 766
Pump peak power, 776
Pump-probe technique, 775

Q

QCSE, *see* Quantum-confined Stark effect
QDash, *see* Quantum dash
QDSOAs, *see* Quantum-dot semiconductor optical amplifiers
QD SOAs, carrier transport model for, 653–655
Quantization of light field, 97–99
Quantum-confined Stark effect (QCSE), 284–285, 405, 547
Quantum-corrected drift-diffusion, 50–52
  capture and escape process, 52–53
  coupling carrier transport model, 53–55
  semiconductor Bloch equation, 55–57
Quantum corrections, 50, 58, 549, 552
Quantum dash (QDash), 633
  SOAs, 615
Quantum disk nanowire light-emitting diodes
  EFA models, 545–547
  empirical tight-binding approach, 547–549
  light-emitting nanorod arrays, 541
  modeling of surface states, 552–553
  multiscale simulation approach, 542
  nanostructure based emerging electronic device, 541–542
  numerical methods, 542–543
  self-consistent calculations, 549–552
  sensitivity to geometrical and material parameters, 553–556
  strain maps, 543–545
  transport properties, 545
Quantum-dot charge-carrier
  dynamics, 721
  scattering, 722–727
Quantum-dot epi structures for SLEDs, 591
Quantum-dot ground-state dynamics, 741
Quantum-dot-in-a-well material equations, 718–722
Quantum dots (QDs), 397, 419–420, 550, 551, 615
  Cd-based chalcogenide QDs
    CdSe/CdTe CQDs, radiative lifetimes in, 437–438
    CdSe/CdTe Type II CQDs, absorption edge wavelength, 434–437

        CdSe QD arrays, charge carrier mobility, 442–443
        correlation energy, 438–441
        equivalent CdTe and CdTe/CdSe QD, absorption spectra, 441–442
    configuration interaction Hamiltonian for NQD, 426–430
    core/shell QD structure, 420–421
    dielectric confinement effect, 424–426
    k·p Hamiltonian and Coulomb integrals, PW implementation, 423–424
    k·p method, for SP electronic structure, 421–423
    materials, 633
    nanocrystal arrays, transport properties, 430–434
Quantum-dot scattering rates, 728
Quantum-dot semiconductor amplifiers, rabi oscillations in, 739–741
Quantum-dot semiconductor optical amplifiers (QDSOAs), 715–716, 716
    delay-differential-equation model, 717–718
    fit functions for scattering rates, 727–728
    gain recovery dynamics in, 736–739
    modeling of spontaneous emission, 728–731
    QD charge-carrier scattering, 722–727
    quantum-dot-in-a-well material equations, 718–722
    rabi oscillations in QD semiconductor amplifiers, 739–741
    static characterization of, 731–735
Quantum mechanical
    effects, 716
    electron, 549
    models, 543
Quantum model, 543
Quantum-well charge-carrier density, 722, 726
Quantum wells (QWs), 4, 19, 50, 365, 389–395, 397, 591, 759
    band structures
        bulk semiconductors, 366–370, 379–387
        of strained-layer QW, 371–376
    in Green LEDs, 582
    imperfect, 582–584
    influence of possible imperfection in, 580–584
    optical gain model with many-body effects, 387–389
    optical matrix elements, 377–378
Quasi-analytic static model, 615–617
Quasi-3D algorithm, 703, 704, 706
Quasi-Fermi distribution, 725
QWs, *see* Quantum wells

R

Rabi flops, 85
Rabi oscillations, 739, 740
    in QD semiconductor amplifiers, 739–741
Radial semiconductor NW heterostructures, 400–401
Radiation, blackbody, 82
Radiation mode, optical waveguide, 115
Radiative recombination, 459–460
Radiative spontaneous recombination, 634–636
    rate, 634, 673
Radiative stimulated recombination rate, 634
Random alloy
    approach, 548, 549
    fluctuation with drift-diffusion solver, 568
    modeling, 562–569
    unipolar transport for, 569–573
Random alloy distribution, 559–562, 563–564
    carrier transport in Green LEDs, 580–584
    MQW InGaN LED and comparison to traditional model, 573–580
    random alloy modeling, 562–569
    unipolar transport for random alloy system, 569–573
Random dopant fluctuations (RDFs), 407–409
Random indium fluctuations model, 582–584
Rate equations
    amplified spontaneous emission, 647–652
    basic, 642–647
    extended rate equation model, 652–656
Rayleigh–Ritz variation principle, 10
Ray tracing, in LED dice, 466
RDFs, *see* Random dopant fluctuations
Reactor simulation, MOCVD, 167–168
Reciprocal lattice, Wigner–Seitz cell of, 131
Recombination models, of light-emitting diodes, 458–461
Recombination rate for OLED, 479
Recombination reaction, 173
Rectangular dielectric waveguide, 113
Recursive coefficients, 756–757
Red/amber AlGaInP LEDs, 462, 467
Red LED structure, 462, 463
    near-field emission intensity distribution, 465, 466
Reduced zone scheme, electron band structure, 6
Reflecting SLEDs (R-SLEDs)
    design, 605
    L-I characteristics of, 593–597
Reflective semiconductor optical amplifier (RSOA), 614, 632–633
Refractive index, 105, 108, 110–113, 123, 591–592, 644, 700, 752
Refractive index variation, 634–636
Reorganization energy, for OLED, 487–488
Repeated zone scheme, electron band structure, 6
Residence-time algorithm, *see* Kinetic Monte Carlo method
Residual facet reflectivities, 777
Resonant energy transfer (RET), *see* Electronic excitation (energy) transfer
Resonant frequencies, 649
Rib-loaded optical waveguide, 114

Rotating wave approximation (RWA), 85, 716, 721
RSOA, *see* Reflective semiconductor optical amplifier
RWA, *see* Rotating wave approximation

S

Scattering self-energies, 69–72
SCGF approach, *see* Self-consistent Green's function approach
SCGF-LCINS approach, 293–294
SCH, *see* Separate confinement heterostructure
Scharfetter and Gummel-finite-box (SG-FB) scheme, 50
Scharfetter–Gummel scheme, 146, 149
Schrödinger equation, 8–13, 17, 19, 20, 634
  and emission rate, 568–569
Schrödinger picture, two-level system in, 84–86
Schrödinger's equation, in indefinite potential well, 194
Schwinger–Keldysh contour, 62, 63
SCNM, *see* Self-consistent numerical method
SCR, *see* Strong confinement regime
Screened hybrid functional (HSE), 16
Secondary ion mass spectrometry (SIMS)-interpolation approach, 178
Secondary peak suppression ratio (SPSR), 604
Second harmonic generation (SHG), 750
Second order perturbation theory, 18, 290
Seebeck effect, 143
Self-and cross-gain modulation (SGM/XGM), 749
Self-and cross-phase modulation (SPM/XPM), 749
Self-consistent approach, 543, 657
Self-consistent band profiles, 550
Self-consistent calculations, 549–552
Self-consistent classical/quantum calculation, 551
Self-consistent Green's function (SCGF) approach, 288–296
Self-consistent numerical method (SCNM), 660, 664, 677
Self-consistent quantum densities, 553
Self-consistent quantum/drift-diffusion coupling, 554
Self-energies
  boundary, 68–69
  Green's function with, 72
  scattering, 69–72
Self-phase modulation (SPM) effect, 615
Semianalytical algorithm, implementation of, 675–677
Semiconductor(s)
  band structure, 7–8
  Bloch equations, 55–57, 95
  crystalline, 191–193
  deep-level impurities in, 256
  devices, 697
  direct band-gap, 141
  heat transport in, 140, 142–144
  inorganic, 91
  lasers, development of, 251
  light-matter coupling in, 92
  materials, 3–4
    tetrahedrally bonded, 8
  modeling technology, 167
  organic, *see* Organic semiconductors
  two-band direct-gap, 90
  wurtzite, electromechanical coupling in, 239–240
Semiconductor-based optical amplifiers (SOAs), 715
Semiconductor-light interaction, 90
Semiconductor nanocrystals, *see* Quantum dots (QDs)
Semiconductor nanowires (NWs), 397–398
  charge confining mechanisms in
    approaches to compute elastic properties, 399–400
    bulk electronic properties, 398
    elastic properties of, 400–405
    piezoelectric and spontaneous polarization, 405–406
  continuous, homogeneous doping-related background charge, 406–407
  crystal-phase bandgap engineering, 410–412
  dielectrically enhanced ionization energies of dopants, 409–410
  heterostructures in, 397, 398
  optical anisotropy of, 412–413
  random dopant fluctuations, 407–409
Semiconductor optical amplifiers (SOAs), 283–285, 612, 632, 633, 697
  analysis of reference, 707–709
  basic principles, 612–615
  basic structure, 612
  bulk SOA static model including ASE, 617–621
  carrier density excursions, 758
  dynamics and pattern effects, 772–776
  electrical gain modulation, 791–792
  FWM analytic model, 627–628, 751
  model formulation, 776–781
  model simplification, 781–783
  nonlinear phenomena in, 748–751, *see also* Nonlinear phenomena in SOAs
  optical fiber amplifiers, 611
  optical gain modulation, 784–787
  pulse amplification analytic model, 623–626
  quasi-analytic static model, 615–617
  structure, 643
  time-domain model including ASE, 621–623
  wave mixing in, 759
Semiempirical model, 8, 17, 19–20
Sensitivity to geometrical and material parameters, 553–556
Separate confinement heterostructure (SCH), 752
SHB, *see* Spectral hole burning
SHB effect, *see* Spatial hole burning effect
Shear deformation, 222

Shockley–Read–Hall (SRH), 479, 568, 642
  and Auger recombination, 755
  generation/recombination, 141
  recombination, 210–214
  surface recombination model, 552
  trapping and recombination, 208, 619, 702
Si-based APDs, 285
Si-compatible group-IV materials, 286
Si-compatible III-V materials, 286
Signal distortion, 766
Signal spectral dependence, 646–647
Signal-to-noise ratios (SNRs), 613
Silicon
  band structure, 7–8, 203, 204
  energy bands in, 139
  first Brillouin zone of, 138–139
Simple predictor-corrector scheme, 549
Simulation, 167
  CSUPREM, 178, 186
  MOCVD reactor, 167–168
  PROCOM, 167, 170
Simulink® software, 660
Single-carrier charge transport, 481
Single-electron Monte Carlo model, 302–303
Single LED, 530–532
Single-material spherical QDs, 420
Single-particle (SP) electronic structure, **k·p** method for, 421–423
Single-particle Hamiltonian of periodic system, 20
Single-quantum well (SQW)
  epitaxy, 592
  LED structures, 458
Six-band Pikus–Bir Hamiltonian, 399
SK mode, *see* Stranski–Krastanov mode
Slab optical waveguide, 108–110, 115, 122
Slater and Koster approach, ETB method, 23
SLED, *see* Superluminescent light-emitting diode
Slowly varying envelope approximation (SVEA), 717, 754
Small-signal analysis, framework for, 685–686
Small-signal carrier density, 686
Small-signal model for transparent SOA, 685–688
SNRs, *see* Signal-to-noise ratios
SOA cavity, resonant properties of, 648–651
SOAs, *see* Semiconductor optical amplifiers
SOA XPM-based wavelength converter, 625
SOA XPM Mach-Zehnder wavelength converter, 625–626
SOC, *see* Spin-orbit coupling
Solid-state lighting, 219, 467, 524
Solid-state system, 90–97
Solids, thermal effects in, 129
Spatial hole burning (SHB) effect, 709, 748
Spatio-temporal grid, 780–781
Spectral hole burning (SHB), 624, 627, 628, 653, 750
  gain compression, 77
  nonlinear gain compression factor, 782
  relaxation time, 774
Spectral photon density, 646, 649, 650, 659, 665
Spectrum broadening effect, 560
Spectrum slicing technique (SSM), 751–752
Spin-orbit coupling (SOC)
  empirical pseudopotential method, 29–30
  empirical tight-binding method, 26–28
Spin-orbit interaction, 7
Split-off (SO) band, 19
Spontaneous emission, 82
  rate contribution, 678
Spontaneous emission (SE) spectra, dilute nitride quantum well lasers, 271
Spontaneous polarization
  alloys, 238
  of semiconductor nanowires, 405–406
  III-N semiconductors, 233–236
  wurtzite crystals, 226
SPSR, *see* Secondary peak suppression ratio
Square optical waveguide, 115
SSM, *see* Spectrum slicing technique
Steady-state carrier density rate equation, 674
Steady-state conditions, 185, 464, 616
Steady-state device gain, 669–671
Steady-state models, 658–659
Steady-state semianalytical model, 671–674
  counter propagating photon density distributions, 675–677
  fundamentals of, 674–675
Steady-state signal, 667–669
Steady-state TW-SOA models, 656–657
Steady-state wideband model, 661
Steady-state wideband self-consistent numerical model (SCNM), 660–662, 666
  implementation of, 662–667
  results and discussion, 667–671
Stimulated emission, 83, 84
Stochastic approach, 651–652
Stochastic models, for OLED, 490–491
Strain, 542–543
Strained-layer quantum wells
  wurtzite semiconductor, 373–376
  zinc-blende semiconductor, 371–373
Strain-induced polarization, 224–226
  in wurtzite crystals, 225–226
  in zincblende crystals, 224–225
Strain maps, 543–545
Strain modeling, 564–567
Stranski–Krastanov (SK) mode, 718
Stress-strain-displacement relations, 566–567
Strong confinement regime (SCR), 440–441

Superlattices, continuity equation in, 236–237
Superluminescence regime, 589
Superluminescent light-emitting diode (SLED), 589
  design and modeling of, 591–600
  designs with specified targets, 600–605
  development of, 590
Surface-gas reactions, 172–173
Surface recombination, 461
  in LED dice, 467
SVEA, *see* Slowly varying envelope approximation
Synopsys Sentaurus Device simulator, 54
Sze–Bethe thermionic-drift diffusion model, 49

T

Tabulated mobilities, OLED, 493–494
TADF, *see* Thermally activated delayed fluorescence
Taper angle, confinement factor and, 710–712
Tapered semiconductor optical amplifiers
  far-field (FF) pattern, 698
  model implementation, 703–706
  modeling approaches, 698–703
  1.5-μ InGaAsP/InP tapered amplifier, 706–712
  semiconductor devices, 697
Tapered waveguides, 121–122, 772
TB, *see* Tight-binding
TB model, for dilute nitride alloys, 261, 263
TCAD, *see* Technology computer-aided design
TDs, *see* Threading dislocations
Technology computer-aided design (TCAD) process simulation, 167, 181
TeleGeography, 631
TE polarization, *see* Transverse electric polarization
Tetrahedrally bonded semiconductor materials, 8
Thermal conductivity, gas, 173
Thermal effects, in solids, 129
Thermal equations, solver for, 705–706
Thermally activated delayed fluorescence (TADF), 477
Thermal models, 701
Thermoelectric model, 47
Thin films, transport, trapping, and recombination in, 202–203
"Thinning of a Poisson process", 491
Threading dislocations (TDs), 561
3D drift-diffusion charge control (3D-DDCC) program, 562
3D examination of Green LEDs, 582–384
3D FEM method, calculation of strain with, 564–567
Three-dimensional (3D) atom probe tomography, 560
Three-dimensional (3D) confined systems, 420
3D vectorial model of VCSELs, 54
Three-dimensional (3D) simulation, 167
3D crystal lattice, 129–130
3D KMC OLED lifetime simulations, 507
3D random alloy fluctuation model, 580
3D steady-state heat equation, 701
III-nitride LEDs, efficiency droop IN, 524
III-nitride semiconductor, electronic polarization and interface charges, 527
III-nitride tunnel junction, 528–530
III-N semiconductors
  continuity equation and quantum well structures, 236–237
  dealing with alloys, 237–239
  nonlinearities in, 228–229
  piezoelectric coefficients
    linear, 227–228
    second-order, 232–233
  spontaneous polarization, 233–236
  wurtzite
    numerical approaches, 232–233
    piezoelectric coefficients, 226
    proper versus improper coefficients, 226–227
III-V lasers on silicon, monolithic integration, 286–287
III-V semiconductor NWs, 398
Tight-binding (TB) method, 259–264, 264–269
Time-dependent (nonstationary) input optical signal, 657
Time-domain model including ASE, 621–623
Time-domain simulations, 751–752
Time-independent Schrödinger equation, 9, 19, 28, 568
TLM, *see* Transmission line model
TMM, *see* Transfer matrix method
TM polarization, *see* Transverse magnetic polarization
TMW, *see* Traveling microwave
TMW small-signal model, 680–684
Transfer matrix method (TMM), 658, 659
Transition energy, photons, 141
Transmission line model (TLM), 660
Transport
  of electrons, 543
  equations, of light-emitting diodes, 456–458
  in organic semiconductors, 196–199
  properties, 545
Transversal waves, 131
Transverse electric (TE) polarization
  wurtzite structure, 378
  zinc-blende structure, 377
Transverse electric (TE)-polarized optical gain, dilute nitride quantum well lasers, 271
Transverse magnetic (TM) polarization
  wurtzite structure, 378
  zinc-blende structure, 377–378
Transverse optical confinement factor, 717
Trap-assisted recombination, *see* Shockley–Read–Hall
Traveling microwave (TMW), 656
Traveling-wave approach, 617–619
Traveling-wave models (TWM), 700, 701
Triplet–polaron quenching (TPQ), 477, 500, 507, 509
Triplet–triplet annihilation (TTA), 500

Triplet–triplet quenching, 477
Tunnel-junction light-emitting diodes
  blue tunnel-junction LED, 532–536
  green tunnel-junction LED, 535–538
  physical models and simulation parameters
    bulk band structure, 524–525
    electronic polarization and interface charges, 527
    impurities, incomplete ionization, 526–527
    interband tunneling model, 527–528
    MQW model, 525–526
  polarization-assisted tunneling effect, 523–524
  single LED, 530–533
  III-nitride LEDs, efficiency droop, 524
  III-nitride tunnel junction, 527–530
TWM, *see* Traveling-wave models
Two-band direct-gap semiconductor, 90
"Two-band" model, 634, 635
Two-dimensional lattice, primitive cell for, 130, 131
Two-dimensional (2D) simulation, 167
Two-dimensional space-spectrum mesh, 663
Two-dimensional (2D) spatiotemporal mesh, 680
Two-level system
  in Heisenberg picture, 86–90
  optical process in, 82–84
  in Schrödinger picture, 84–86
Two-photon absorption (TPA), 750, 773
TW-SOA, operation principles of, 632–633

U

Ultraviolet (UV) LEDs, 453
Unipolar models, 700, 702, 748, 764
Upwind scheme numerical implementation, 678–680

V

Vacuum stability condition, 60
Valence bands
  electrons in, 138
  strained-layer QW
    wurtzite semiconductors, 375–376
    zinc-blende semiconductors, 371–373
Valence force field (VFF) method, 399, 401, 543, 548, 551
Variable step size method (VSSM), 482, 483
VCA, *see* Virtual crystal approximation
VCSELs, *see* Vertical-cavity surface emitting lasers
VCSOAs, *see* Vertical-cavity semiconductor optical amplifiers
VECSELs, *see* Vertical-external-cavity surface-emitting lasers
Vector, crystals, 223
Vegard's law, 228, 237, 239, 254, 255, 548
Vertical-cavity devices, 281
Vertical-cavity semiconductor optical amplifiers (VCSOAs), 281, 284
Vertical-cavity surface emitting lasers (VCSELs), 35, 50, 116, 252
  resonance frequency, 53–54
  3D vectorial model, 54
Vertical-external-cavity surface-emitting lasers (VECSELs), 281, 282
VFF method, *see* Valence force field method
Virtual crystal approximation (VCA), 548
Viscosity, gas, 173
VSSM, *see* Variable step size method

W

WA-FDBPM, *see* Wide-angle finite-difference beam propagation method
Wall-plug efficiency (WPE)
  of blue light-emitting diode structures, 535
  droop of, 532
  of green single light-emitting diode, 535, 537, 538
Wave functions, of ethene molecule, 195, 196
Waveguide
  bent, 123–124
  in-diffused, 114
  optical, *see* Optical waveguide
  photonic crystal, 116
  rectangular dielectric, 113
  tapered, 121–122
Waveguide design, 635–636–642
Waveguide polarization sensitivity, 637
Wave-guiding mechanism, 116
Wavelength
  conversion, 751
  emission, 4
Wavelength division multiplexing (WDM), 632, 751, 773
Wavelength independent parameters, 777
Wave mixing effects, in SOA, 747–748
  CDM measurement, 762
  simulation parameters, 759
  time-domain model, 748, 751, 756–757
Wave vector-dependent deformation potentials, 38
WDM, *see* Wavelength division multiplexing
Wentzel–Kramers–Brillouin (WKB) tunneling, 577
Wide amplification spectrum, 632
Wide-angle finite-difference beam propagation method (WA-FDBPM), 703, 704
Wiener–Khinchin theorem, 603, 730
Wigner–Seitz cell, 130
  of reciprocal lattice, 131
WPE, *see* Wall-plug efficiency
Wurtzite crystals, 221
  crystal directions in, 224
  piezoelectric coefficients, second-order, 231
  piezoelectric tensor in, 223
  spontaneous polarization in, 226
  strain-induced polarization in, 225–226

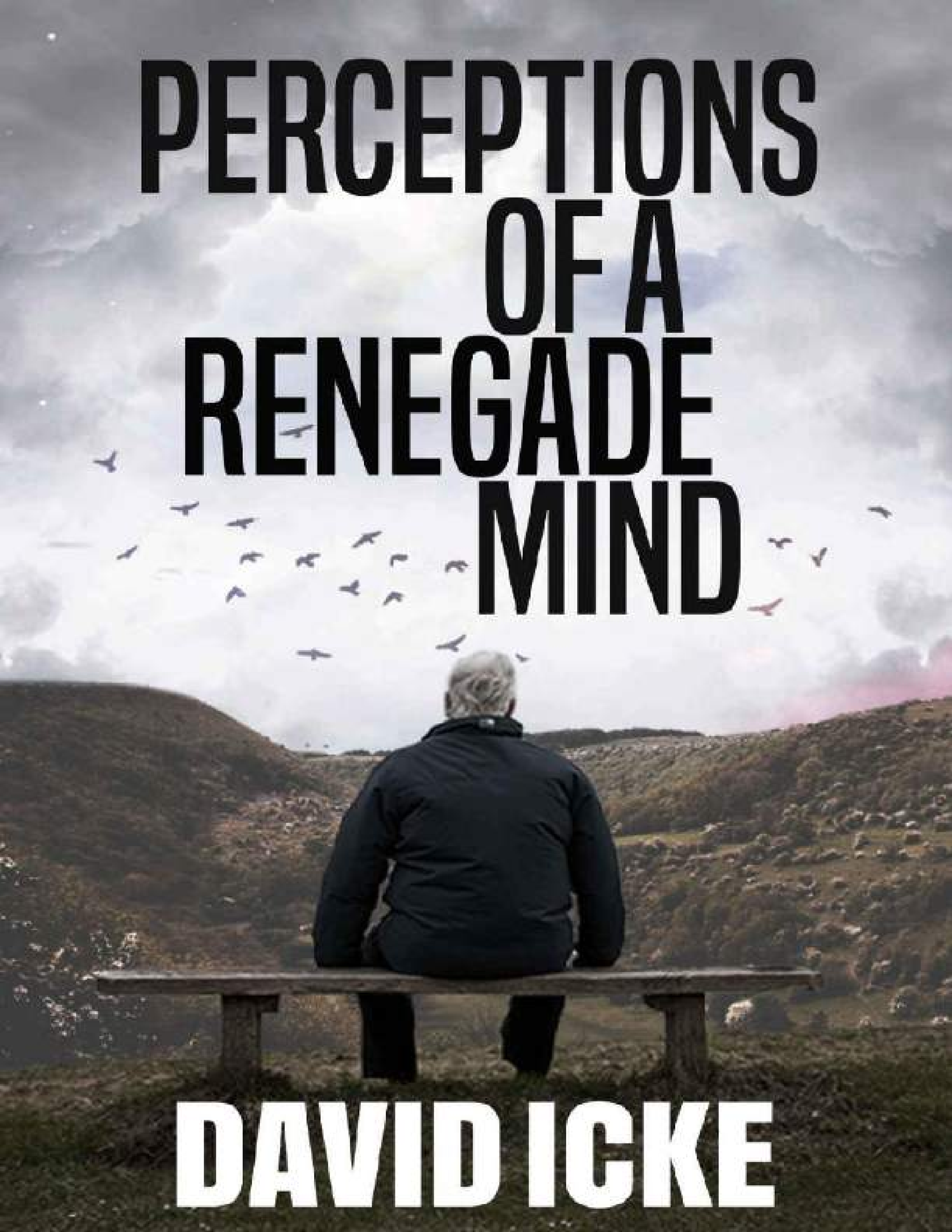
Wurtzite III-N semiconductors
  numerical approaches, 232–233
  piezoelectric coefficients of, 226
  proper versus improper coefficients, 226–227
Wurtzite InN, electronic band structure, 25
Wurtzite nanostructure, multiband envelope-function model for, 52
Wurtzite semiconductors
  with arbitrary crystal orientation, 383–385, 386–387
  electromechanical coupling, 239–240
    with linear piezoelectricity, 240–243
    with nonlinear piezoelectricity, 243–245
  examples of, 392–395
  Luttinger–Kohn Hamiltonian
    block-diagonalized 3 × 3 Hamiltonian, 369–370
    6 × 6 Hamiltonian for valence band, 368–369
  nonpolar orientations in, 245
  strained-layer QW
    eigenvalues and eigenfunctions screening, 375–376
    energy splitting, 374–375
    internal field, 373–374
  structure, optical matrix elements, 378
WZ III-V semiconductors, 412

X

XGM wavelength converter, 622–623

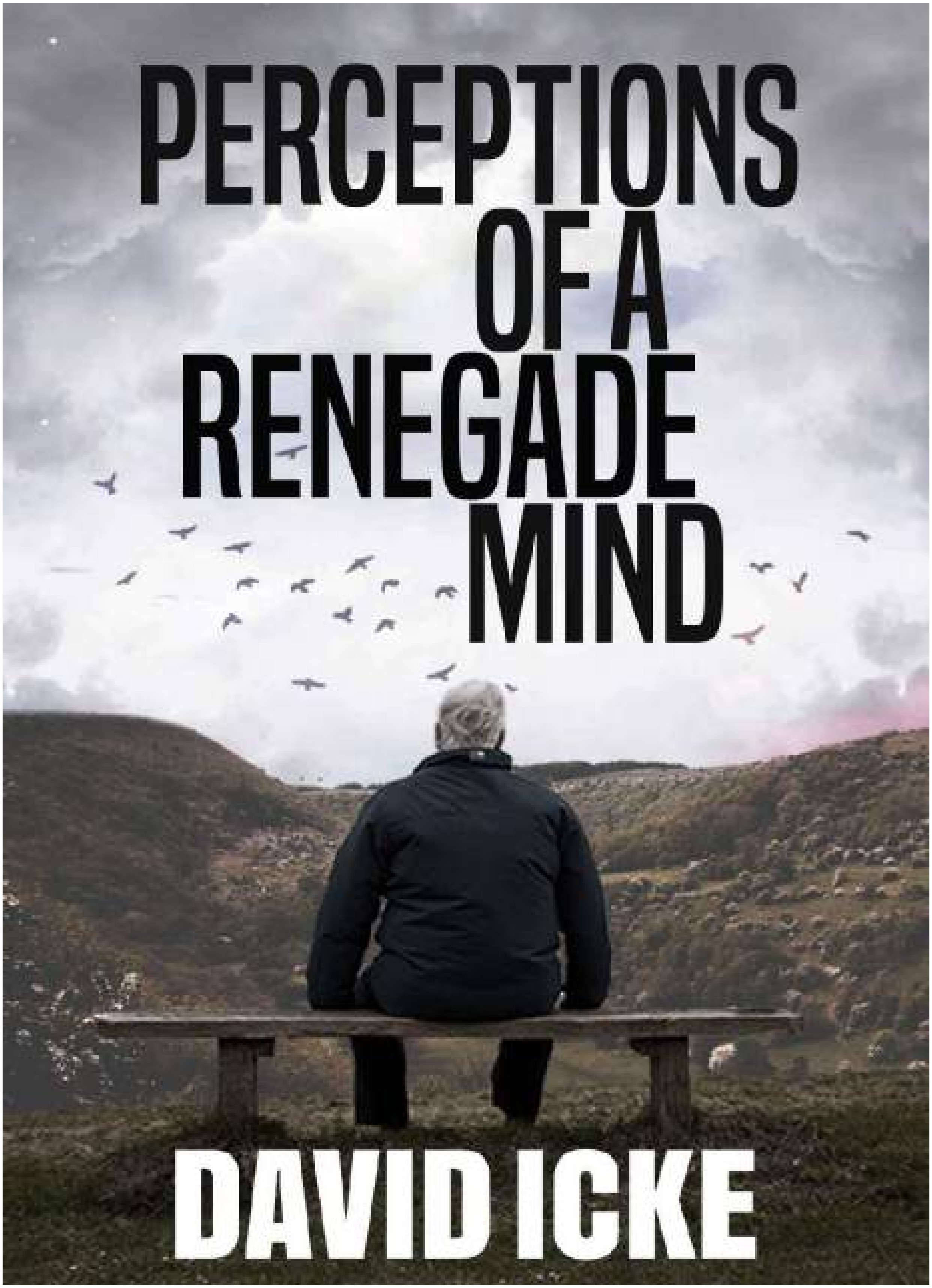
Z

Zerner's intermediate neglect of differential overlap (ZINDO), 198, 488
Zero facet reflectivities, 612, 753
Zincblende crystals, 221
  crystal directions in, 224
  piezoelectric coefficients, second-order, 231
  piezoelectric tensor in, 223
  strain-induced polarization in, 224–225
Zinc-blende semiconductors
  with arbitrary crystal orientation, 379–383, 386
  examples of, 387–392
  Luttinger–Kohn Hamiltonian
    block-diagonalized 3 × 3 Hamiltonian, 367–369
    6 × 6 Hamiltonian for valence band, 366–367
  strained-layer QW
    conduction band, 371
    valence band, 371–373
  structure, optical matrix elements, 377–378
ZINDO, *see* Zerner's intermediate neglect of differential overlap



PERCEPTIONS OF A RENEGADE MIND

DAVID ICKE



PERCEPTIONS OF A RENEGADE MIND

DAVID ICKE

PERCEPTIONS OF A RENEGADE MIND

ickonic
publishing

First published in July 2021.

ickonic
publishing

New Enterprise House
St Helens Street
Derby
DE1 3GY
UK

email: gareth.icke@davidicke.com

Copyright © 2021 David Icke

No part of this book may be reproduced in any form without permission from the Publisher, except for the quotation of brief passages in criticism

Cover Design: Gareth Icke
Book Design: Neil Hague

British Library Cataloguing-in Publication Data
A catalogue record for this book is
available from the British Library

eISBN 978-18384153-1-0

PERCEPTIONS
OF A
RENEGADE
MIND

DAVID ICKE

Dedication:

To *Freeeeeedom!*

ICKONIC
THE ALTERNATIVE

NEW. DIFFERENT. REVOLUTIONARY

HUNDREDS OF CUTTING EDGE DOCUMENTARIES,
FEATURE FILMS, SERIES & PODCASTS.

SIGN UP NOW AT ICKONIC.COM

THE LIFE STORY OF DAVID ICKE

RENEGADE

THE FEATURE LENGTH FILM

AVAILABLE NOW AT DAVIDICKE.COM

Renegade:
Adjective
‘Having rejected tradition: Unconventional.’
Merriam-Webster Dictionary

Acquiescence to tyranny is the death of the spirit

You may be 38 years old, as I happen to be. And one day, some great opportunity stands before you and calls you to stand up for some great principle, some great issue, some great cause. And you refuse to do it because you are afraid … You refuse to do it because you want to live longer … You're afraid that you will lose your job, or you are afraid that you will be criticised or that you will lose your popularity, or you're afraid that somebody will stab you, or shoot at you or bomb your house; so you refuse to take the stand.

Well, you may go on and live until you are 90, but you're just as dead at 38 as you would be at 90. And the cessation of breathing in your life is but the belated announcement of an earlier death of the spirit.

Martin Luther King

How the few control the many and always have – the many do whatever they're told

‘Forward, the Light Brigade!’
Was there a man dismayed?
Not though the soldier knew
Someone had blundered.
Theirs not to make reply,
Theirs not to reason why,
Theirs but to do and die.
Into the valley of Death
Rode the six hundred.

Cannon to right of them,
Cannon to left of them,
Cannon in front of them
Volleyed and thundered;
Stormed at with shot and shell,
Boldly they rode and well,
Into the jaws of Death,
Into the mouth of hell
Rode the six hundred

Alfred Lord Tennyson (1809-1892)

The mist is lifting slowly
I can see the way ahead
And I’ve left behind the empty streets
That once inspired my life
And the strength of the emotion
Is like thunder in the air
’Cos the promise that we made each other
Haunts me to the end

The secret of your beauty
And the mystery of your soul
I’ve been searching for in everyone I meet
And the times I’ve been mistaken
It’s impossible to say
And the grass is growing
Underneath our feet

The words that I remember
From my childhood still are true
That there’s none so blind
As those who will not see
And to those who lack the courage
And say it’s dangerous to try
Well they just don’t know
That love eternal will not be denied

I know you’re out there somewhere
Somewhere, somewhere
I know you’re out there somewhere

Somewhere you can hear my voice
I know I'll find you somehow
Somehow, somehow
I know I'll find you somehow
And somehow I'll return again to you

The Moody Blues

Are you a gutless wonder – or a Renegade Mind?

Monuments put from pen to paper,
Turns me into a gutless wonder,
And if you tolerate this,
Then your children will be next.
Gravity keeps my head down,
Or is it maybe shame ...

Manic Street Preachers

Rise like lions after slumber
In unvanquishable number.
Shake your chains to earth like dew
Which in sleep have fallen on you.
Ye are many – they are few.

Percy Shelley

Contents

CHAPTER 1 'I'm thinking' – Oh, but *are* you?
CHAPTER 2 Renegade perception
CHAPTER 3 The Pushbacker sting
CHAPTER 4 'Covid': The calculated catastrophe
CHAPTER 5 There *is no* 'virus'
CHAPTER 6 Sequence of deceit
CHAPTER 7 War on your mind
CHAPTER 8 'Reframing' insanity
CHAPTER 9 We must have it? So what is it?
CHAPTER 10 Human 2.0
CHAPTER 11 Who controls the Cult?
CHAPTER 12 Escaping Wetiko

POSTSCRIPT
APPENDIX Cowan-Kaufman-Morell Statement on Virus Isolation
BIBLIOGRAPHY
INDEX

CHAPTER ONE

I'm thinking' – Oh, but *are* you?

Think for yourself and let others enjoy the privilege of doing so too
Voltaire

French-born philosopher, mathematician and scientist René Descartes became famous for his statement in Latin in the 17th century which translates into English as: 'I think, therefore I am.'

On the face of it that is true. Thought reflects perception and perception leads to both behaviour and self-identity. In that sense 'we' are what we think. But who or what is doing the thinking and is thinking the only route to perception? Clearly, as we shall see, 'we' are not always the source of 'our' perception, indeed with regard to humanity as a whole this is rarely the case; and thinking is far from the only means of perception. Thought is the village idiot compared with other expressions of consciousness that we all have the potential to access and tap into. This has to be true when we *are* those other expressions of consciousness which are infinite in nature. We have forgotten this, or, more to the point, been manipulated to forget.

These are not just the esoteric musings of the navel. The whole foundation of human control and oppression is control of perception. Once perception is hijacked then so is behaviour which is dictated by perception. Collective perception becomes collective behaviour and collective behaviour is what we call human society. Perception is all and those behind human control know that which is why perception is the target 24/7 of the psychopathic manipulators that I call the Global Cult. They know that if they dictate perception they will dictate behaviour and collectively dictate the nature of human society. They are further aware that perception is formed from information received and if they control the circulation of information they will to a vast extent direct human behaviour. Censorship of information and opinion has become globally Nazi-like in recent years and never more blatantly than since the illusory 'virus pandemic' was triggered out of China in 2019 and across the world in 2020. Why have billions submitted to house arrest and accepted fascistic societies in a way they would have never believed possible? Those controlling the information spewing from government, mainstream media and Silicon Valley (all controlled by the same Global Cult networks) told them they were in danger from a 'deadly virus' and only by submitting to house arrest and conceding their most basic of freedoms could they and their families be protected. This monumental and provable lie became the *perception* of the billions and therefore the *behaviour* of the billions. In those few words you have the whole structure and modus operandi of human control. Fear is a perception – **F**alse **E**motion **A**ppearing **R**eal – and fear is the currency of control. In short … get them by the balls (or give them the impression that you have) and their hearts and minds will follow. Nothing grips the dangly bits and freezes the rear-end more comprehensively than fear.

World number 1

There are two 'worlds' in what appears to be one 'world' and the prime difference between them is knowledge. First we have the mass of human society in which the population is maintained in coldly-calculated ignorance through control of information and the 'education' (indoctrination) system. That's all you really need to control to enslave billions in a perceptual delusion in which what are perceived to be *their* thoughts and opinions are ever-repeated mantras that the system has been downloading all their lives through 'education', media, science, medicine, politics and academia in which the personnel and advocates are themselves overwhelmingly the perceptual products of the same repetition. Teachers and academics in general are processed by the same programming machine as everyone else, but unlike the great majority they never leave the ‘education’ program. It gripped them as students and continues to grip them as programmers of subsequent generations of students. The programmed become the programmers – the programmed programmers. The same can largely be said for scientists, doctors and politicians and not least because as the American writer Upton Sinclair said: ‘It is difficult to get a man to understand something when his salary depends upon his not understanding it.’ If your career and income depend on thinking the way the system demands then you will – bar a few free-minded exceptions – concede your mind to the Perceptual Mainframe that I call the Postage Stamp Consensus. This is a tiny band of perceived knowledge and possibility ‘taught’ (downloaded) in the schools and universities, pounded out by the mainstream media and on which all government policy is founded. Try thinking, and especially speaking and acting, outside of the ‘box’ of consensus and see what that does for your career in the Mainstream Everything which bullies, harasses, intimidates and ridicules the population into compliance. Here we have the simple structure which enslaves most of humanity in a perceptual prison cell for an entire lifetime and I’ll go deeper into this process shortly. Most of what humanity is taught as fact is nothing more than programmed belief. American science fiction author Frank Herbert was right when he said: ‘Belief can be manipulated. Only knowledge is dangerous.’ In the ‘Covid’ age belief is promoted and knowledge is censored. It was always so, but never to the extreme of today.

World number 2

A ‘number 2’ is slang for ‘doing a poo’ and how appropriate that is when this other ‘world’ is doing just that on humanity every minute of every day. World number 2 is a global network of secret societies and semi-secret groups dictating the direction of society via governments, corporations and authorities of every kind. I have spent more than 30 years uncovering and exposing this network that I call the Global Cult and knowing its agenda is what has made my books so accurate in predicting current and past events. Secret societies are secret for a reason. They want to keep their hoarded knowledge to themselves and their chosen initiates and to hide it from the population which they seek through ignorance to control and subdue. The whole foundation of the division between World 1 and World 2 is *knowledge*. What number 1 knows number 2 must not. Knowledge they have worked so hard to keep secret includes (a) the agenda to enslave humanity in a centrally-controlled global dictatorship, and (b) the nature of reality and life itself. The latter (b) must be suppressed to allow the former (a) to prevail as I shall be explaining. The way the Cult manipulates and interacts with the population can be likened to a spider's web. The 'spider' sits at the centre in the shadows and imposes its will through the web with each strand represented in World number 2 by a secret society, satanic or semi-secret group, and in World number 1 – the world of the seen – by governments, agencies of government, law enforcement, corporations, the banking system, media conglomerates and Silicon Valley (Fig 1 overleaf). The spider and the web connect and coordinate all these organisations to pursue the same global outcome while the population sees them as individual entities working randomly and independently. At the level of the web governments *are* the banking system *are* the corporations *are* the media *are* Silicon Valley *are* the World Health Organization working from their inner cores as one unit. Apparently unconnected countries, corporations, institutions, organisations and people are on the *same team* pursuing the same global outcome. Strands in the web immediately around the spider are the most secretive and exclusive secret societies and their membership is emphatically restricted to the Cult inner-circle emerging through the generations from particular bloodlines for reasons I will come to. At the core of the core you would get them in a single room. That's how many people are dictating the direction of human society and its transformation through the ‘Covid’ hoax and other means. As the web expands out from the spider we meet the secret societies that many people will be aware of – the Freemasons, Knights Templar, Knights of Malta, Opus Dei, the inner sanctum of the Jesuit Order, and such like. Note how many are connected to the Church of Rome and there is a reason for that. The Roman Church was established as a revamp, a rebranding, of the relocated ‘Church’ of Babylon and the Cult imposing global tyranny today can be tracked back to Babylon and Sumer in what is now Iraq.

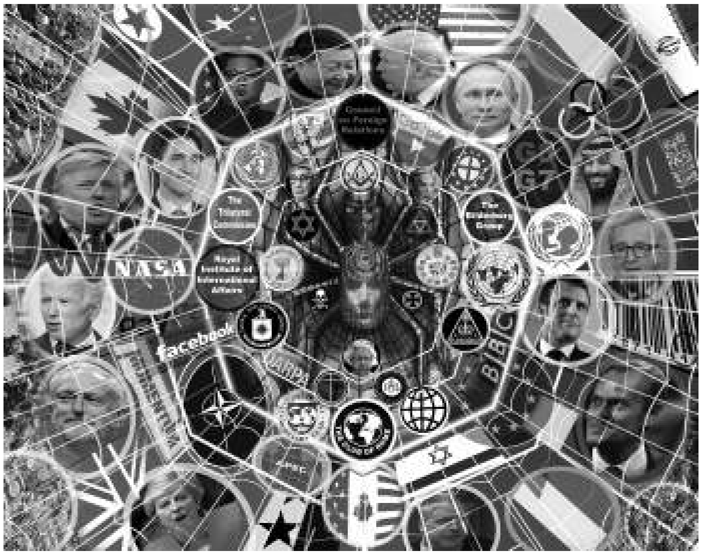

Council on Foreign Relations
The Trilateral Commission
Royal Institute of International Affairs
The Bilderberg Group
NASA
facebook
BBC

Figure 1: The global web through which the few control the many. (Image Neil Hague.)

Inner levels of the web operate in the unseen away from the public eye and then we have what I call the cusp organisations located at the point where the hidden meets the seen. They include a series of satellite organisations answering to a secret society founded in London in the late 19th century called the Round Table and among them are the Royal Institute of International Affairs (UK, founded in 1920); Council on Foreign Relations (US, 1921); Bilderberg Group (worldwide, 1954); Trilateral Commission (US/worldwide, 1972); and the Club of Rome (worldwide, 1968) which was created to exploit environmental concerns to justify the centralisation of global power to ‘save the planet’. The Club of Rome instigated with others the human-caused climate change hoax which has led to all the ‘green new deals' demanding that very centralisation of control. Cusp organisations, which include endless 'think tanks' all over the world, are designed to coordinate a single global policy between political and business leaders, intelligence personnel, media organisations and anyone who can influence the direction of policy in their own sphere of operation. Major players and regular attenders will know what is happening – or some of it – while others come and go and are kept overwhelmingly in the dark about the big picture. I refer to these cusp groupings as semi-secret in that they can be publicly identified, but what goes on at the inner-core is kept very much 'in house' even from most of their members and participants through a fiercely-imposed system of compartmentalisation. Only let them know what they need to know to serve your interests and no more. The structure of secret societies serves as a perfect example of this principle. Most Freemasons never get higher than the bottom three levels of 'degree' (degree of knowledge) when there are 33 official degrees of the Scottish Rite. Initiates only qualify for the next higher 'compartment' or degree if those at that level choose to allow them. Knowledge can be carefully assigned only to those considered 'safe'. I went to my local Freemason's lodge a few years ago when they were having an 'open day' to show how cuddly they were and when I chatted to some of them I was astonished at how little the rank and file knew even about the most ubiquitous symbols they use. The mushroom technique – keep them in the dark and feed them bullshit – applies to most people in the web as well as the population as a whole. Sub-divisions of the web mirror in theme and structure transnational corporations which have a headquarters somewhere in the world dictating to all their subsidiaries in different countries. Subsidiaries operate in their methodology and branding to the same centrally-dictated plan and policy in pursuit of particular ends. The Cult web functions in the same way. Each country has its own web as a subsidiary of the global one. They consist of networks of secret societies, semi-secret groups and bloodline families and their job is to impose the will of the spider and the global web in their particular country. Subsidiary networks control and manipulate the national political system, finance, corporations, media, medicine, etc. to ensure that they follow the globally-dictated Cult agenda. These networks were the means through which the 'Covid' hoax could be played out with almost every country responding in the same way.

The 'Yessir' pyramid

Compartmentalisation is the key to understanding how a tiny few can dictate the lives of billions when combined with a top-down sequence of imposition and acquiescence. The inner core of the Cult sits at the peak of the pyramidal hierarchy of human society (Fig 2 overleaf). It imposes its will – its agenda for the world – on the level immediately below which acquiesces to that imposition. This level then imposes the Cult will on the level below them which acquiesces and imposes on the next level. Very quickly we meet levels in the hierarchy that have no idea there even is a Cult, but the sequence of imposition and acquiescence continues down the pyramid in just the same way. 'I don't know why we are doing this but the order came from "on-high" and so we better just do it.' Alfred Lord Tennyson said of the cannon fodder levels in his poem *The Charge of the Light Brigade*: 'Theirs not to reason why; theirs but to do and die.' The next line says that 'into the valley of death rode the six hundred' and they died because they obeyed without question what their perceived 'superiors' told them to do. In the same way the population capitulated to 'Covid'. The whole hierarchical pyramid functions like this to allow the very few to direct the enormous many. Eventually imposition-acquiescence-imposition-acquiescence comes down to the mass of the population at the foot of the pyramid. If they acquiesce to those levels of the hierarchy imposing on them (governments/law enforcement/doctors/media) a circuit is completed between the population and the handful of super-psychopaths in the Cult inner core at the top of the pyramid. Without a circuit-breaking refusal to obey, the sequence of imposition and acquiescence allows a staggeringly few people to impose their will upon the entirety of humankind. We are looking at the very sequence that has subjugated billions since the start of 2020. Our freedom has not been taken from us. Humanity has given it away. Fascists do not impose fascism because there are not enough of them. Fascism is imposed by the population acquiescing to fascism. Put another way allowing their perceptions to be programmed to the extent that leads to the population giving their freedom away by giving their perceptions – their mind – away. If this circuit is not broken by humanity ceasing to cooperate with their own enslavement then nothing can change. For that to happen people have to critically think and see through the lies and window dressing and then summon the backbone to act upon what they see. The Cult spends its days working to stop either happening and its methodology is systematic and highly detailed, but it can be overcome and that is what this book is all about.

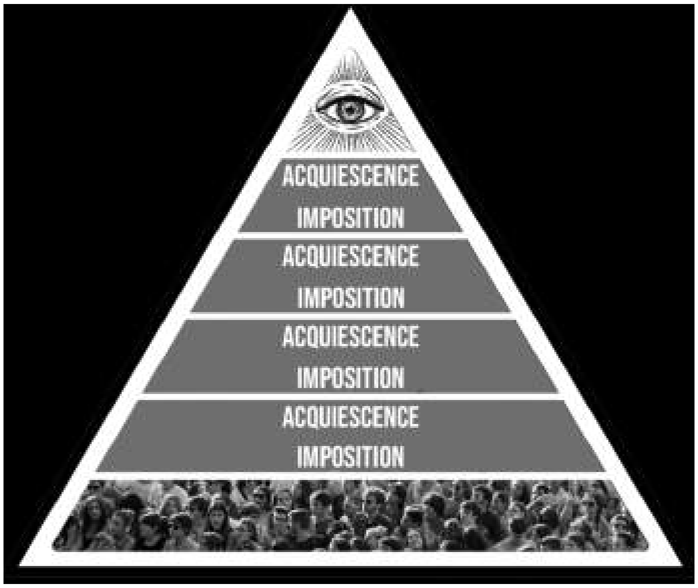


Figure 2: The simple sequence of imposition and compliance that allows a handful of people at the peak of the pyramid to dictate the lives of billions.

The Life Program

Okay, back to world number 1 or the world of the ‘masses’. Observe the process of what we call ‘life’ and it is a perceptual download from cradle to grave. The Cult has created a global structure in which perception can be programmed and the program continually topped-up with what appears to be constant confirmation that the program is indeed true reality. The important word here is ‘appears’.

This is the structure, the fly-trap, the Postage Stamp Consensus or Perceptual Mainframe, which represents that incredibly narrow band of perceived possibility delivered by the 'education' system, mainstream media, science and medicine. From the earliest age the download begins with parents who have themselves succumbed to the very programming their children are about to go through. Most parents don't do this out of malevolence and mostly it is quite the opposite. They do what they believe is best for their children and that is what the program has told them is best. Within three or four years comes the major transition from parental programming to full-blown state (Cult) programming in school, college and university where perceptually-programmed teachers and academics pass on their programming to the next generations. Teachers who resist are soon marginalised and their careers ended while children who resist are called a problem child for whom Ritalin may need to be prescribed. A few years after entering the 'world' children are under the control of authority figures representing the state telling them when they have to be there, when they can leave and when they can speak, eat, even go to the toilet. This is calculated preparation for a lifetime of obeying authority in all its forms. Reflex-action fear of authority is instilled by authority from the start. Children soon learn the carrot and stick consequences of obeying or defying authority which is underpinned daily for the rest of their life. Fortunately I daydreamed through this crap and never obeyed authority simply because it told me to. This approach to my alleged 'betters' continues to this day. There can be consequences of pursuing open-minded freedom in a world of closed-minded conformity. I spent a lot of time in school corridors after being ejected from the classroom for not taking some of it seriously and now I spend a lot of time being ejected from Facebook, YouTube and Twitter. But I can tell you that being true to yourself and not compromising your self-respect is far more exhilarating than bowing to authority for authority's sake. You don't have to be a sheep to the shepherd (authority) and the sheep dog (fear of not obeying authority).

The perceptual download continues throughout the formative years in school, college and university while script-reading 'teachers', 'academics' 'scientists', 'doctors' and 'journalists' insist that ongoing generations must be as programmed as they are. Accept the program or you will not pass your 'exams' which confirm your 'degree' of programming. It is tragic to think that many parents pressure their offspring to work hard at school to download the program and qualify for the next stage at college and university. The late, great, American comedian George Carlin said: 'Here's a bumper sticker I'd like to see: We are proud parents of a child who has resisted his teachers' attempts to break his spirit and bend him to the will of his corporate masters.' Well, the best of luck finding many of those, George. Then comes the moment to leave the formal programming years in academia and enter the 'adult' world of work. There you meet others in your chosen or prescribed arena who went through the same Postage Stamp Consensus program before you did. There is therefore overwhelming agreement between almost everyone on the basic foundations of Postage Stamp reality and the rejection, even contempt, of the few who have a mind of their own and are prepared to use it. This has two major effects. Firstly, the consensus confirms to the programmed that their download is really how things are. I mean, everyone knows that, right? Secondly, the arrogance and ignorance of Postage Stamp adherents ensure that anyone questioning the program will have unpleasant consequences for seeking their own truth and not picking their perceptions from the shelf marked: 'Things you must believe without question and if you don't you're a dangerous lunatic conspiracy theorist and a harebrained nutter'.

Every government, agency and corporation is founded on the same Postage Stamp prison cell and you can see why so many people believe the same thing while calling it their own 'opinion'. Fusion of governments and corporations in pursuit of the same agenda was the definition of fascism described by Italian dictator Benito Mussolini. The pressure to conform to perceptual norms downloaded for a lifetime is incessant and infiltrates society right down to family groups that become censors and condemners of their own ‘black sheep’ for not, ironically, being sheep. We have seen an explosion of that in the ‘Covid’ era. Cult-owned global media unleashes its propaganda all day every day in support of the Postage Stamp and targets with abuse and ridicule anyone in the public eye who won’t bend their mind to the will of the tyranny. Any response to this is denied (certainly in my case). They don’t want to give a platform to expose official lies. Cult-owned-and-created Internet giants like Facebook, Google, YouTube and Twitter delete you for having an unapproved opinion. Facebook boasts that its AI censors delete 97-percent of ‘hate speech’ before anyone even reports it. Much of that ‘hate speech’ will simply be an opinion that Facebook and its masters don’t want people to see. Such perceptual oppression is widely known as fascism. Even Facebook executive Benny Thomas, a ‘CEO Global Planning Lead’, said in comments secretly recorded by investigative journalism operation Project Veritas that Facebook is ‘too powerful’ and should be broken up:

> I mean, no king in history has been the ruler of two billion people, but Mark Zuckerberg is … And he’s 36. That’s too much for a 36-year-old ... You should not have power over two billion people. I just think that’s wrong.

Thomas said Facebook-owned platforms like Instagram, Oculus, and WhatsApp needed to be separate companies. ‘It’s too much power when they’re all one together’. That’s the way the Cult likes it, however. We have an executive of a Cult organisation in Benny Thomas that doesn’t know there is a Cult such is the compartmentalisation. Thomas said that Facebook and Google ‘are no longer companies, they’re countries’. Actually they are more powerful than countries on the basis that if you control information you control perception and control human society.

I love my oppressor

Another expression of this psychological trickery is for those who realise they are being pressured into compliance to eventually convince themselves to believe the official narratives to protect their self-respect from accepting the truth that they have succumbed to meek and subservient compliance. Such people become some of the most vehement defenders of the system. You can see them everywhere screaming abuse at those who prefer to think for themselves and by doing so reminding the compliers of their own capitulation to conformity. ‘You are talking dangerous nonsense you Covidiot!!’ Are you trying to convince me or yourself? It is a potent form of Stockholm syndrome which is defined as: ‘A psychological condition that occurs when a victim of abuse identifies and attaches, or bonds, positively with their abuser.’ An example is hostages bonding and even ‘falling in love’ with their kidnappers. The syndrome has been observed in domestic violence, abused children, concentration camp inmates, prisoners of war and many and various Satanic cults. These are some traits of Stockholm syndrome listed at goodtherapy.org:

- Positive regard towards perpetrators of abuse or captor [see ‘Covid’].
- Failure to cooperate with police and other government authorities when it comes to holding perpetrators of abuse or kidnapping accountable [or in the case of ‘Covid’ cooperating with the police to enforce and defend their captors’ demands].
- Little or no effort to escape [see ‘Covid’].
- Belief in the goodness of the perpetrators or kidnappers [see ‘Covid’].
- Appeasement of captors. This is a manipulative strategy for maintaining one’s safety. As victims get rewarded – perhaps with less abuse or even with life itself – their appeasing behaviours are reinforced [see ‘Covid’].
- Learned helplessness. This can be akin to ‘if you can’t beat ‘em, join ‘em’. As the victims fail to escape the abuse or captivity, they may start giving up and soon realize it’s just easier for everyone if they acquiesce all their power to their captors [see ‘Covid’].

- Feelings of pity toward the abusers, believing they are actually victims themselves. Because of this, victims may go on a crusade or mission to 'save' [protect] their abuser [see the venom unleashed on those challenging the official 'Covid' narrative].
- Unwillingness to learn to detach from their perpetrators and heal. In essence, victims may tend to be less loyal to themselves than to their abuser [*definitely* see 'Covid'].

Ponder on those traits and compare them with the behaviour of great swathes of the global population who have defended governments and authorities which have spent every minute destroying their lives and livelihoods and those of their children and grandchildren since early 2020 with fascistic lockdowns, house arrest and employment deletion to 'protect' them from a 'deadly virus' that their abusers' perceptually created to bring about this very outcome. We are looking at mass Stockholm syndrome. All those that agree to concede their freedom will believe those perceptions are originating in their own independent 'mind' when in fact by conceding their reality to Stockholm syndrome they have by definition conceded any independence of mind. Listen to the 'opinions' of the acquiescing masses in this 'Covid' era and what gushes forth is the repetition of the official version of everything delivered unprocessed, unfiltered and unquestioned. The whole programming dynamic works this way. I must be free because I'm told that I am and so I think that I am.

You can see what I mean with the chapter theme of 'I'm thinking – Oh, but *are* you?' The great majority are not thinking, let alone for themselves. They are repeating what authority has told them to believe which allows them to be controlled. Weaving through this mentality is the fear that the 'conspiracy theorists' are right and this again explains the often hysterical abuse that ensues when you dare to contest the official narrative of anything. Denial is the mechanism of hiding from yourself what you don't want to be true. Telling people what they want to hear is easy, but it's an infinitely greater challenge to tell them what they would rather not be happening.

One is akin to pushing against an open door while the other is met with vehement resistance no matter what the scale of evidence. I don't want it to be true so I'll convince myself that it's not. Examples are everywhere from the denial that a partner is cheating despite all the signs to the reflex-action rejection of any idea that world events in which country after country act in exactly the same way are centrally coordinated. To accept the latter is to accept that a force of unspeakable evil is working to destroy your life and the lives of your children with nothing too horrific to achieve that end. Who the heck wants that to be true? But if we don't face reality the end is duly achieved and the consequences are far worse and ongoing than breaking through the walls of denial today with the courage to make a stand against tyranny.

Connect the dots – but how?

A crucial aspect of perceptual programming is to portray a world in which everything is random and almost nothing is connected to anything else. Randomness cannot be coordinated by its very nature and once you perceive events as random the idea they could be connected is waved away as the rantings of the tinfoil-hat brigade. You can't plan and coordinate random you idiot! No, you can't, but you can hide the coldly-calculated and long-planned behind the *illusion* of randomness. A foundation manifestation of the Renegade Mind is to scan reality for patterns that connect the apparently random and turn pixels and dots into pictures. This is the way I work and have done so for more than 30 years. You look for similarities in people, modus operandi and desired outcomes and slowly, then ever quicker, the picture forms. For instance: There would seem to be no connection between the 'Covid pandemic' hoax and the human-caused global-warming hoax and yet they are masks (appropriately) on the same face seeking the same outcome. Those pushing the global warming myth through the Club of Rome and other Cult agencies are driving the lies about 'Covid' – Bill Gates is an obvious one, but they are endless. Why would the same people be involved in both when they are clearly not connected? Oh, but they *are*. Common themes with personnel are matched by common goals. The ‘solutions’ to both ‘problems’ are centralisation of global power to impose the will of the few on the many to ‘save’ humanity from ‘Covid’ and save the planet from an ‘existential threat’ (we need ‘zero Covid’ and ‘zero carbon emissions’). These, in turn, connect with the ‘dot’ of globalisation which was coined to describe the centralisation of global power in every area of life through incessant political and corporate expansion, trading blocks and superstates like the European Union. If you are the few and you want to control the many you have to centralise power and decision-making. The more you centralise power the more power the few at the centre will have over the many; and the more that power is centralised the more power those at the centre have to centralise even quicker. The momentum of centralisation gets faster and faster which is exactly the process we have witnessed. In this way the hoaxed ‘pandemic’ and the fakery of human-caused global warming serve the interests of globalisation and the seizure of global power in the hands of the Cult inner-circle which is behind ‘Covid’, ‘climate change’ *and* globalisation. At this point random ‘dots’ become a clear and obvious picture or pattern.

Klaus Schwab, the classic Bond villain who founded the Cult’s Gates-funded World Economic Forum, published a book in 2020, *The Great Reset*, in which he used the ‘problem’ of ‘Covid’ to justify a total transformation of human society to ‘save’ humanity from ‘climate change’. Schwab said: ‘The pandemic represents a rare but narrow window of opportunity to reflect, reimagine, and reset our world.’ What he didn’t mention is that the Cult he serves is behind both hoaxes as I show in my book *The Answer*. He and the Cult don’t have to reimagine the world. They know precisely what they want and that’s why they destroyed human society with ‘Covid’ to ‘build back better’ in their grand design. Their job is not to imagine, but to get humanity to imagine and agree with their plans while believing it’s all random. It must be pure coincidence that ‘The Great Reset’ has long been the Cult’s code name for the global imposition of fascism and replaced previous code-names of the ‘New World

Order' used by Cult frontmen like Father George Bush and the 'New Order of the Ages' which emerged from Freemasonry and much older secret societies. New Order of the Ages appears on the reverse of the Great Seal of the United States as 'Novus ordo seclorum' underneath the Cult symbol used since way back of the pyramid and all seeing-eye (Fig 3). The pyramid is the hierarchy of human control headed by the illuminated eye that symbolises the force behind the Cult which I will expose in later chapters. The term 'Annuit Coeptis' translates as 'He favours our undertaking'. We are told the 'He' is the Christian god, but 'He' is not as I will be explaining.

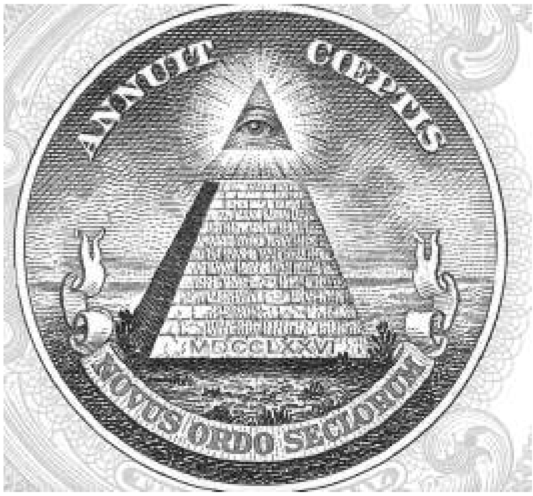


Figure 3: The all-seeing eye of the Cult 'god' on the Freemason-designed Great Seal of the United States and also on the dollar bill.

Having you on

Two major Cult techniques of perceptual manipulation that relate to all this are what I have called since the 1990s Problem-Reaction-Solution (PRS) and the Totalitarian Tiptoe (TT). They can be uncovered by the inquiring mind with a simple question: Who benefits? The answer usually identifies the perpetrators of a given action or happening through the concept of 'he who most benefits from a crime is the one most likely to have committed it'. The Latin 'Cue bono?' – Who benefits? – is widely attributed to the Roman orator and statesman Marcus Tullius Cicero. No wonder it goes back so far when the concept has been relevant to human behaviour since history was recorded. Problem-Reaction-Solution is the technique used to manipulate us every day by covertly creating a problem (or the illusion of one) and offering the solution to the problem (or the illusion of one). In the first phase you create the problem and blame someone or something else for why it has happened. This may relate to a financial collapse, terrorist attack, war, global warming or pandemic, anything in fact that will allow you to impose the 'solution' to change society in the way you desire at that time. The 'problem' doesn't have to be real. PRS is manipulation of perception and all you need is the population to believe the problem is real. Human-caused global warming and the 'Covid pandemic' only have to be *perceived* to be real for the population to accept the 'solutions' of authority. I refer to this technique as NO-Problem-Reaction-Solution. Billions did not meekly accept house arrest from early 2020 because there was a real deadly 'Covid pandemic' but because they perceived – believed – that to be the case. The antidote to Problem-Reaction-Solution is to ask who benefits from the proposed solution. Invariably it will be anyone who wants to justify more control through deletion of freedom and centralisation of power and decision-making.

The two world wars were Problem-Reaction-Solutions that transformed and realigned global society. Both were manipulated into being by the Cult as I have detailed in books since the mid-1990s. They dramatically centralised global power, especially World War Two, which led to the United Nations and other global bodies thanks to the overt and covert manipulations of the Rockefeller family and other Cult bloodlines like the Rothschilds. The UN is a stalking horse for full-blown world government that I will come to shortly. The land on which the UN building stands in New York was donated by the Rockefellers and the same Cult family was behind Big Pharma scalpel and drug 'medicine' and the creation of the World Health Organization as part of the UN. They have been stalwarts of the eugenics movement and funded Hitler's race-purity expert' Ernst Rudin. The human-caused global warming hoax has been orchestrated by the Club of Rome through the UN which is manufacturing both the ‘problem’ through its Intergovernmental Panel on Climate Change and imposing the ‘solution’ through its Agenda 21 and Agenda 2030 which demand the total centralisation of global power to ‘save the world’ from a climate hoax the United Nations is itself perpetrating. What a small world the Cult can be seen to be particularly among the inner circles. The bedfellow of Problem-Reaction-Solution is the Totalitarian Tiptoe which became the Totalitarian Sprint in 2020. The technique is fashioned to hide the carefully-coordinated behind the cover of apparently random events. You start the sequence at ‘A’ and you know you are heading for ‘Z’. You don’t want people to know that and each step on the journey is presented as a random happening while all the steps strung together lead in the same direction. The speed may have quickened dramatically in recent times, but you can still see the incremental approach of the Tiptoe in the case of ‘Covid’ as each new imposition takes us deeper into fascism. Tell people they have to do this or that to get back to ‘normal’, then this and this and this. With each new demand adding to the ones that went before the population’s freedom is deleted until it disappears. The spider wraps its web around the flies more comprehensively with each new diktat. I’ll highlight this in more detail when I get to the ‘Covid’ hoax and how it has been pulled off. Another prime example of the Totalitarian Tiptoe is how the Cult-created European Union went from a ‘free-trade zone’ to a centralised bureaucratic dictatorship through the Tiptoe of incremental centralisation of power until nations became mere administrative units for Cult-owned dark suits in Brussels.

The antidote to ignorance is knowledge which the Cult seeks vehemently to deny us, but despite the systematic censorship to that end the Renegade Mind can overcome this by vociferously seeking out the facts no matter the impediments put in the way. There is also a method of thinking and perceiving – *knowing* – that doesn’t even need names, dates, place-type facts to identify the patterns that reveal the story. I’ll get to that in the final chapter. All you need to know about the manipulation of human society and to what end is still out there – *at the time of writing* – in the form of books, videos and websites for those that really want to breach the walls of programmed perception. To access this knowledge requires the abandonment of the mainstream media as a source of information in the awareness that this is owned and controlled by the Cult and therefore promotes mass perceptions that suit the Cult. Mainstream media lies all day, every day. That is its function and very reason for being. Where it does tell the truth, here and there, is only because the truth and the Cult agenda very occasionally coincide. If you look for fact and insight to the BBC, CNN and virtually all the rest of them you are asking to be conned and perceptually programmed.

Know the outcome and you'll see the journey

Events seem random when you have no idea where the world is being taken. Once you do the random becomes the carefully planned. Know the outcome and you'll see the journey is a phrase I have been using for a long time to give context to daily happenings that appear unconnected. Does a problem, or illusion of a problem, trigger a proposed 'solution' that further drives society in the direction of the outcome? Invariably the answer will be yes and the random – *abracadabra* – becomes the clearly coordinated. So what is this outcome that unlocks the door to a massively expanded understanding of daily events? I will summarise its major aspects – the fine detail is in my other books – and those new to this information will see that the world they thought they were living in is a very different place. The foundation of the Cult agenda is the incessant centralisation of power and all such centralisation is ultimately in pursuit of Cult control on a global level. I have described for a long time the planned world structure of top-down dictatorship as the Hunger Games Society. The term obviously comes from the movie series which portrayed a world in which a few living in military-protected hi-tech luxury were the overlords of a population condemned to abject poverty in isolated 'sectors' that were not allowed to interact. 'Covid' lockdowns and travel bans anyone? The 'Hunger Games' pyramid of structural control has the inner circle of the Cult at the top with pretty much the entire population at the bottom under their control through dependency for survival on the Cult. The whole structure is planned to be protected and enforced by a military-police state (Fig 4).

Here you have the reason for the global lockdowns of the fake pandemic to coldly destroy independent incomes and livelihoods and make everyone dependent on the ‘state’ (the Cult that controls the ‘states’). I have warned in my books for many years about the plan to introduce a ‘guaranteed income’ – a barely survivable pittance – designed to impose dependency when employment was destroyed by AI technology and now even more comprehensively at great speed by the ‘Covid’ scam. Once the pandemic was played and lockdown consequences began to delete independent income the authorities began to talk right on cue about the need for a guaranteed income and a ‘Great Reset’. Guaranteed income will be presented as benevolent governments seeking to help a desperate people – desperate as a direct result of actions of the same governments. The truth is that such payments are a trap. You will only get them if you do exactly what the authorities demand including mass vaccination (genetic manipulation). We have seen this theme already in Australia where those dependent on government benefits have them reduced if parents don’t agree to have their children vaccinated according to an insane health-destroying government-dictated schedule. Calculated economic collapse applies to governments as well as people. The Cult wants rid of countries through the creation of a world state with countries broken up into regions ruled by a world government and super states like the European Union. Countries must be bankrupted, too, to this end and it’s being achieved by the trillions in ‘rescue packages’ and furlough payments, trillions in lost taxation, and money-no-object spending on ‘Covid’ including constant all-medium advertising (programming) which has made the media dependent on government for much of its income. The day of reckoning is coming – as planned – for government spending and given that it has been made possible by printing money and not by production/taxation there is inflation on the way that has the potential to wipe out monetary value. In that case there will be no need for the Cult to steal your money. It just won't be worth anything (see the German Weimar Republic before the Nazis took over). Many have been okay with lockdowns while getting a percentage of their income from so-called furlough payments without having to work. Those payments are dependent, however, on people having at least a theoretical job with a business considered non-essential and ordered to close. As these business go under because they are closed by lockdown after lockdown the furlough stops and it will for everyone eventually. Then what? The 'then what?' is precisely the idea.

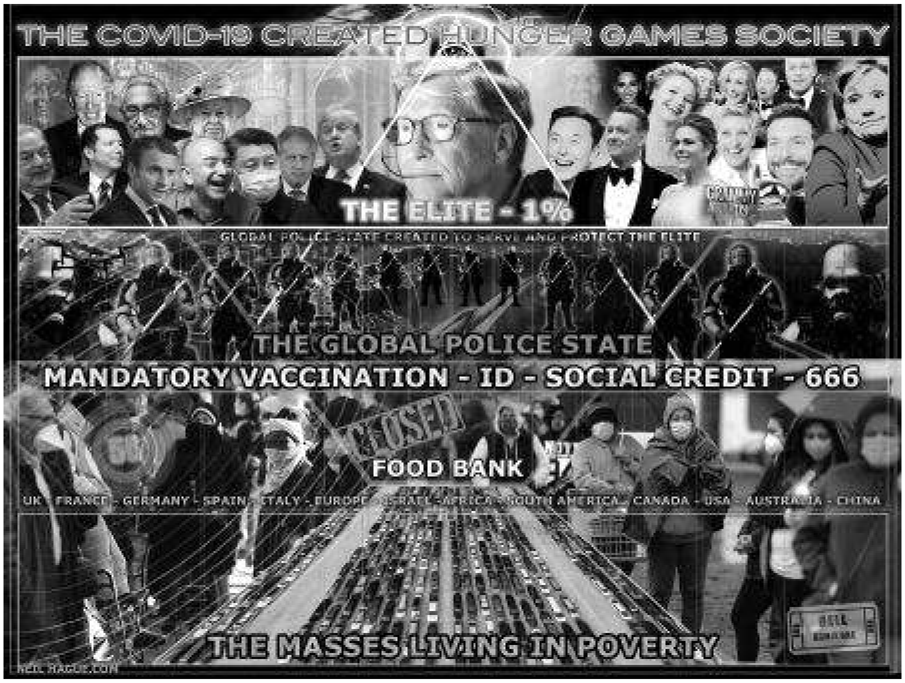

THE COVID-19 CREATED HUNGER GAMES SOCIETY
THE ELITE - 1%
GLOBAL POLICE STATE CREATED TO SERVE AND PROTECT THE ELITE
THE GLOBAL POLICE STATE
MANDATORY VACCINATION - ID - SOCIAL CREDIT - 666
CLOSED
FOOD BANK
UK - FRANCE - GERMANY - SPAIN - ITALY - EUROPE - ISRAEL - AFRICA - SOUTH AMERICA - CANADA - USA - AUSTRALIA - CHINA
THE MASSES LIVING IN POVERTY
NEIL HAGUE.COM

Figure 4: The Hunger Games Society structure I have long warned was planned and now the 'Covid' hoax has made it possible. This is the real reason for lockdowns.

Hired hands

Between the Hunger Games Cult elite and the dependent population is planned to be a vicious military-police state (a fusion of the two into one force). This has been in the making for a long time with police looking ever more like the military and carrying weapons to match. The pandemic scam has seen this process accelerate so fast as lockdown house arrest is brutally enforced by carefully recruited fascist minds and gormless system-servers. The police and military are planned to merge into a centrally-directed world army in a global structure headed by a world government which wouldn't be elected even by the election fixes now in place. The world army is not planned even to be human and instead wars would be fought, primarily against the population, using robot technology controlled by artificial intelligence. I have been warning about this for decades and now militaries around the world are being transformed by this very AI technology. The global regime that I describe is a particular form of fascism known as a technocracy in which decisions are not made by clueless and co-opted politicians but by unelected technocrats – scientists, engineers, technologists and bureaucrats. Cult-owned-and-controlled Silicon Valley giants are examples of technocracy and they already have far more power to direct world events than governments. They are with their censorship *selecting* governments. I know that some are calling the 'Great Reset' a Marxist communist takeover, but fascism and Marxism are different labels for the same tyranny. Tell those who lived in fascist Germany and Stalinist Russia that there was a difference in the way their freedom was deleted and their lives controlled. I could call it a fascist technocracy or a Marxist technocracy and they would be equally accurate. The Hunger Games society with its world government structure would oversee a world army, world central bank and single world cashless currency imposing its will on a microchipped population (Fig 5). Scan its different elements and see how the illusory pandemic is forcing society in this very direction at great speed. Leaders of 23 countries and the World Health Organization (WHO) backed the idea in March, 2021, of a global treaty for 'international cooperation' in 'health emergencies' and nations should 'come together as a global community for peaceful cooperation that extends beyond this crisis'. Cut the Orwellian bullshit and this means another step towards global government. The plan includes a cashless digital money system that I first warned about in 1993. Right at the start of 'Covid' the deeply corrupt Tedros

Adhanom Ghebreyesus, the crooked and merely gofer 'head' of the World Health Organization, said it was possible to catch the 'virus' by touching cash and it was better to use cashless means. The claim was ridiculous nonsense and like the whole 'Covid' mind-trick it was nothing to do with 'health' and everything to do with pushing every aspect of the Cult agenda. As a result of the Tedros lie the use of cash has plummeted. The Cult script involves a single world digital currency that would eventually be technologically embedded in the body. China is a massive global centre for the Cult and if you watch what is happening there you will know what is planned for everywhere. The Chinese government is developing a digital currency which would allow fines to be deducted immediately via AI for anyone caught on camera breaking its fantastic list of laws and the money is going to be programmable with an expiry date to ensure that no one can accrue wealth except the Cult and its operatives.

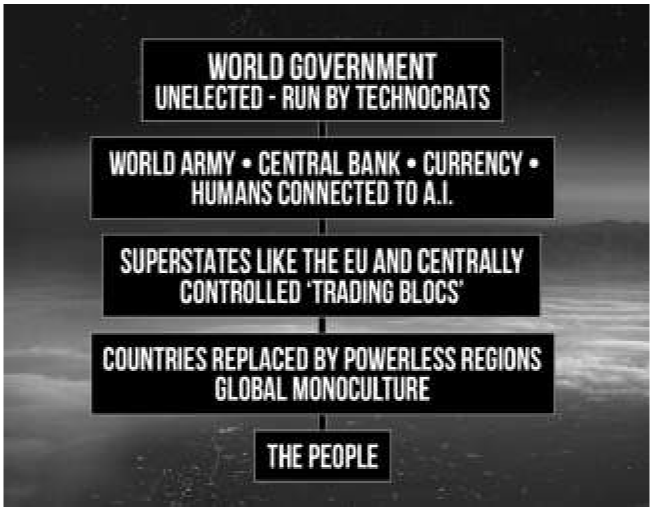

WORLD GOVERNMENT
UNELECTED - RUN BY TECHNOCRATS
WORLD ARMY • CENTRAL BANK • CURRENCY •
HUMANS CONNECTED TO A.I.
SUPERSTATES LIKE THE EU AND CENTRALLY
CONTROLLED 'TRADING BLOCS'
COUNTRIES REPLACED BY POWERLESS REGIONS
GLOBAL MONOCULTURE
THE PEOPLE

Figure 5: The structure of global control the Cult has been working towards for so long and this has been enormously advanced by the 'Covid' illusion.

Serfdom is so smart

The Cult plan is far wider, extreme, and more comprehensive than even most conspiracy researchers appreciate and I will come to the true depths of deceit and control in the chapters 'Who controls the

Cult?' and 'Escaping Wetiko'. Even the world that we know is crazy enough. We are being deluged with ever more sophisticated and controlling technology under the heading of 'smart'. We have smart televisions, smart meters, smart cards, smart cars, smart driving, smart roads, smart pills, smart patches, smart watches, smart skin, smart borders, smart pavements, smart streets, smart cities, smart communities, smart environments, smart growth, smart planet ... smart *everything* around us. Smart technologies and methods of operation are designed to interlock to create a global Smart Grid connecting the entirety of human society including human minds to create a centrally-dictated 'hive' mind. 'Smart cities' is code for densely-occupied megacities of total surveillance and control through AI. Ever more destructive frequency communication systems like 5G have been rolled out without any official testing for health and psychological effects (colossal). 5G/6G/7G systems are needed to run the Smart Grid and each one becomes more destructive of body and mind. Deleting independent income is crucial to forcing people into these AI-policed prisons by ending private property ownership (except for the Cult elite). The Cult's Great Reset now openly foresees a global society in which no one will own any possessions and everything will be rented while the Cult would own literally everything under the guise of government and corporations. The aim has been to use the lockdowns to destroy sources of income on a mass scale and when the people are destitute and in unrepayable amounts of debt (problem) Cult assets come forward with the pledge to write-off debt in return for handing over all property and possessions (solution). Everything – literally everything including people – would be connected to the Internet via AI. I was warning years ago about the coming Internet of Things (IoT) in which all devices and technology from your car to your fridge would be plugged into the Internet and controlled by AI. Now we are already there with much more to come. The next stage is the Internet of Everything (IoE) which is planned to include the connection of AI to the human brain and body to replace the human mind with a centrally-controlled AI mind. Instead of perceptions being manipulated through control of information and censorship those perceptions would come direct from the Cult through AI. What do you think? You think whatever AI decides that you think. In human terms there would be no individual 'think' any longer. Too incredible? The ravings of a lunatic? Not at all. Cult-owned crazies in Silicon Valley have been telling us the plan for years without explaining the real motivation and calculated implications. These include Google executive and 'futurist' Ray Kurzweil who highlights the year 2030 for when this would be underway. He said:

> Our thinking ... will be a hybrid of biological and non-biological thinking ... humans will be able to extend their limitations and 'think in the cloud' ... We're going to put gateways to the cloud in our brains ... We're going to gradually merge and enhance ourselves ... In my view, that's the nature of being human – we transcend our limitations.
>
> As the technology becomes vastly superior to what we are then the small proportion that is still human gets smaller and smaller and smaller until it's just utterly negligible.

The sales-pitch of Kurzweil and Cult-owned Silicon Valley is that this would make us 'super-human' when the real aim is to make us post-human and no longer 'human' in the sense that we have come to know. The entire global population would be connected to AI and become the centrally-controlled 'hive-mind' of externally-delivered perceptions. The Smart Grid being installed to impose the Cult's will on the world is being constructed to allow particular locations – even one location – to control the whole global system. From these prime control centres, which absolutely include China and Israel, anything connected to the Internet would be switched on or off and manipulated at will. Energy systems could be cut, communication via the Internet taken down, computer-controlled driverless autonomous vehicles driven off the road, medical devices switched off, the potential is limitless given how much AI and Internet connections now run human society. We have seen nothing yet if we allow this to continue. Autonomous vehicle makers are working with law enforcement to produce cars designed to automatically pull over if they detect a police or emergency vehicle flashing from up to 100 feet away. At a police stop the car would be unlocked and the window rolled down automatically. Vehicles would only take you where the computer (the state) allowed. The end of petrol vehicles and speed limiters on all new cars in the UK and EU from 2022 are steps leading to electric computerised transport over which ultimately you have no control. The picture is far bigger even than the Cult global network or web and that will become clear when I get to the nature of the 'spider'. There is a connection between all these happenings and the instigation of DNA-manipulating 'vaccines' (which aren't 'vaccines') justified by the 'Covid' hoax. That connection is the unfolding plan to transform the human body from a biological to a synthetic biological state and this is why synthetic biology is such a fast-emerging discipline of mainstream science. 'Covid vaccines' are infusing self-replicating synthetic genetic material into the cells to cumulatively take us on the Totalitarian Tiptoe from Human 1.0 to the synthetic biological Human 2.0 which will be physically and perceptually attached to the Smart Grid to one hundred percent control every thought, perception and deed. Humanity needs to wake up and *fast*.

This is the barest explanation of where the 'outcome' is planned to go but it's enough to see the journey happening all around us. Those new to this information will already see 'Covid' in a whole new context. I will add much more detail as we go along, but for the minutiae evidence see my mega-works, *The Answer, The Trigger* and *Everything You Need to Know But Have Never Been Told*.

Now – how does a Renegade Mind see the 'world'?

CHAPTER TWO

Renegade Perception

It is one thing to be clever and another to be wise
George R.R. Martin

A simple definition of the difference between a programmed mind and a Renegade Mind would be that one sees only dots while the other connects them to see the picture. Reading reality with accuracy requires the observer to (a) know the planned outcome and (b) realise that everything, but *everything*, is connected.

The entirety of infinite reality is connected – that's its very nature – and with human society an expression of infinite reality the same must apply. Simple cause and effect is a connection. The effect is triggered by the cause and the effect then becomes the cause of another effect. Nothing happens in isolation because it *can't*. Life in whatever reality is simple choice and consequence. We make choices and these lead to consequences. If we don't like the consequences we can make different choices and get different consequences which lead to other choices and consequences. The choice and the consequence are not only connected they are indivisible. You can't have one without the other as an old song goes. A few cannot control the world unless those being controlled allow that to happen – cause and effect, choice and consequence. Control – who has it and who doesn't – is a two-way process, a symbiotic relationship, involving the controller and controlled. 'They took my freedom away!!' Well, yes, but you also gave it to them. Humanity is subjected to mass control because humanity has acquiesced to that control. This is all cause and effect and literally a case of give and take. In the same way world events of every kind are connected and the Cult works incessantly to sell the illusion of the random and coincidental to maintain the essential (to them) perception of dots that hide the picture. Renegade Minds know this and constantly scan the world for patterns of connection. This is absolutely pivotal in understanding the happenings in the world and without that perspective clarity is impossible. First you know the planned outcome and then you identify the steps on the journey – the day-by-day apparently random which, when connected in relation to the outcome, no longer appear as individual events, but as the proverbial *chain* of events leading in the same direction. I'll give you some examples:

Political puppet show

We are told to believe that politics is 'adversarial' in that different parties with different beliefs engage in an endless tussle for power. There may have been some truth in that up to a point – and only a point – but today divisions between 'different' parties are rhetorical not ideological. Even the rhetorical is fusing into one-speak as the parties eject any remaining free thinkers while others succumb to the ever-gathering intimidation of anyone with the 'wrong' opinion. The Cult is not a new phenomenon and can be traced back thousands of years as my books have documented. Its intergenerational initiates have been manipulating events with increasing effect the more that global power has been centralised. In ancient times the Cult secured control through the system of monarchy in which 'special' bloodlines (of which more later) demanded the right to rule as kings and queens simply by birthright and by vanquishing others who claimed the same birthright. There came a time, however, when people had matured enough to see the unfairness of such tyranny and demanded a say in who governed them. Note the word – *governed* them. Not served them – *governed* them, hence government defined as 'the political direction and control exercised over the actions of the members, citizens, or inhabitants of communities, societies, and states; direction of the affairs of a state, community, etc.' Governments exercise control over rather than serve just like the monarchies before them. Bizarrely there are still countries like the United Kingdom which are ruled by a monarch *and* a government that officially answers to the monarch. The UK head of state and that of Commonwealth countries such as Canada, Australia and New Zealand is 'selected' by who in a *single family* had unprotected sex with whom and in what order. Pinch me it can't be true. Ouch! Shit, it is. The demise of monarchies in most countries offered a potential vacuum in which some form of free and fair society could arise and the Cult had that base covered. Monarchies had served its interests but they couldn't continue in the face of such widespread opposition and, anyway, replacing a 'royal' dictatorship that people could see with a dictatorship 'of the people' hiding behind the concept of 'democracy' presented far greater manipulative possibilities and ways of hiding coordinated tyranny behind the illusion of 'freedom'.

Democracy is quite wrongly defined as government selected by the population. This is not the case at all. It is government selected by *some* of the population (and then only in theory). This 'some' doesn't even have to be the majority as we have seen so often in first-past-the-post elections in which the so-called majority party wins fewer votes than the 'losing' parties combined. Democracy can give total power to a party in government from a minority of the votes cast. It's a sleight of hand to sell tyranny as freedom. Seventy-four million Trump-supporting Americans didn't vote for the 'Democratic' Party of Joe Biden in the distinctly dodgy election in 2020 and yet far from acknowledging the wishes and feelings of that great percentage of American society the Cult-owned Biden government set out from day one to destroy them and their right to a voice and opinion. Empty shell Biden and his Cult handlers said they were doing this to 'protect democracy'. Such is the level of lunacy and sickness to which politics has descended. Connect the dots and relate them to the desired outcome – a world government run by self-appointed technocrats and no longer even elected politicians. While operating through its political agents in government the Cult is at the same time encouraging public distain for politicians by putting idiots and incompetents in theoretical power on the road to deleting them. The idea is to instil a public reaction that says of the technocrats: 'Well, they couldn't do any worse than the pathetic politicians.' It's all about controlling perception and Renegade Minds can see through that while programmed minds cannot when they are ignorant of both the planned outcome and the manipulation techniques employed to secure that end. This knowledge can be learned, however, and fast if people choose to get informed.

Politics may at first sight appear very difficult to control from a central point. I mean look at the 'different' parties and how would you be able to oversee them all and their constituent parts? In truth, it's very straightforward because of their structure. We are back to the pyramid of imposition and acquiescence. Organisations are structured in the same way as the system as a whole. Political parties are not open forums of free expression. They are hierarchies. I was a national spokesman for the British Green Party which claimed to be a different kind of politics in which influence and power was devolved; but I can tell you from direct experience – and it's far worse now – that Green parties are run as hierarchies like all the others however much they may try to hide that fact or kid themselves that it's not true. A very few at the top of all political parties are directing policy and personnel. They decide if you are elevated in the party or serve as a government minister and to do that you have to be a yes man or woman. Look at all the maverick political thinkers who never ascended the greasy pole. If you want to progress within the party or reach 'high-office' you need to fall into line and conform. Exceptions to this are rare indeed. Should you want to run for parliament or Congress you have to persuade the local or state level of the party to select you and for that you need to play the game as dictated by the hierarchy. If you secure election and wish to progress within the greater structure you need to go on conforming to what is acceptable to those running the hierarchy from the peak of the pyramid. Political parties are perceptual gulags and the very fact that there are party ‘Whips’ appointed to ‘whip’ politicians into voting the way the hierarchy demands exposes the ridiculous idea that politicians are elected to serve the people they are supposed to represent. Cult operatives and manipulation has long seized control of major parties that have any chance of forming a government and at least most of those that haven’t. A new party forms and the Cult goes to work to infiltrate and direct. This has reached such a level today that you see video compilations of ‘leaders’ of all parties whether Democrats, Republicans, Conservative, Labour and Green parroting the same Cult mantra of ‘Build Back Better’ and the ‘Great Reset’ which are straight off the Cult song-sheet to describe the transformation of global society in response to the Cult-instigated hoaxes of the ‘Covid pandemic’ and human-caused ‘climate change’. To see Caroline Lucas, the Green Party MP that I knew when I was in the party in the 1980s, speaking in support of plans proposed by Cult operative Klaus Schwab representing the billionaire global elite is a real head-shaker.

Many parties – one master

The party system is another mind-trick and was instigated to change the nature of the dictatorship by swapping ‘royalty’ for dark suits that people believed – though now ever less so – represented their interests. Understanding this trick is to realise that a single force (the Cult) controls all parties either directly in terms of the major ones or through manipulation of perception and ideology with others. You don’t need to manipulate Green parties to demand your transformation of society in the name of ‘climate change’ when they are obsessed with the lie that this is essential to ‘save the planet’. You just give them a platform and away they go serving your interests while believing they are being environmentally virtuous. America’s political structure is a perfect blueprint for how the two or multi-party system is really a one-party state. The Republican Party is controlled from one step back in the shadows by a group made up of billionaires and their gofers known as neoconservatives or Neocons.

I have exposed them in fine detail in my books and they were the driving force behind the policies of the imbecilic presidency of Boy George Bush which included 9/11 (see *The Trigger* for a comprehensive demolition of the official story), the subsequent 'war on terror' (war *of* terror) and the invasions of Afghanistan and Iraq. The latter was a No-Problem-Reaction-Solution based on claims by Cult operatives, including Bush and British Prime Minister Tony Blair, about Saddam Hussein's 'weapons of mass destruction' which did not exist as war criminals Bush and Blair well knew.

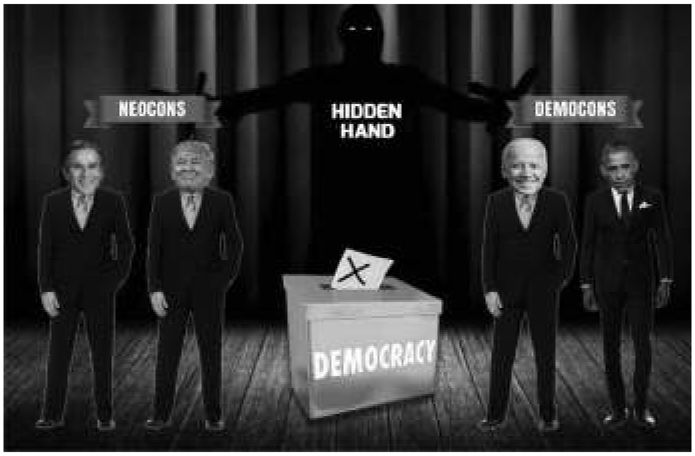


Figure 6: Different front people, different parties – same control system.

The Democratic Party has its own 'Neocon' group controlling from the background which I call the 'Democons' and here's the penny-drop – the Neocons and Democons answer to the same masters one step further back into the shadows (Fig 6). At that level of the Cult the Republican and Democrat parties are controlled by the same people and no matter which is in power the Cult is in power. This is how it works in almost every country and certainly in Britain with Conservative, Labour, Liberal Democrat and Green parties now all on the same page whatever the rhetoric may be in their feeble attempts to appear different. Neocons operated at the time of Bush through a think tank called The Project for the New American Century which in September, 2000, published a document entitled *Rebuilding America's Defenses: Strategies, Forces, and Resources*

For a New Century demanding that America fight 'multiple, simultaneous major theatre wars' as a 'core mission' to force regime-change in countries including Iraq, Libya and Syria. Neocons arranged for Bush ('Republican') and Blair ('Labour Party') to front-up the invasion of Iraq and when they departed the Democons orchestrated the targeting of Libya and Syria through Barack Obama ('Democrat') and British Prime Minister David Cameron ('Conservative Party'). We have 'different' parties and 'different' people, but the same unfolding script. The more the Cult has seized the reigns of parties and personnel the more their policies have transparently pursued the same agenda to the point where the fascist 'Covid' impositions of the Conservative junta of Jackboot Johnson in Britain were opposed by the Labour Party because they were not fascist enough. The Labour Party is likened to the US Democrats while the Conservative Party is akin to a British version of the Republicans and on both sides of the Atlantic they all speak the same language and support the direction demanded by the Cult although some more enthusiastically than others. It's a similar story in country after country because it's all centrally controlled. Oh, but what about Trump? I'll come to him shortly. Political 'choice' in the 'party' system goes like this: You vote for Party A and they get into government. You don't like what they do so next time you vote for Party B and they get into government. You don't like what they do when it's pretty much the same as Party A and why wouldn't that be with both controlled by the same force? Given that only two, sometimes three, parties have any chance of forming a government to get rid of Party B that you don't like you have to vote again for Party A which … you don't like. This, ladies and gentlemen, is what they call 'democracy' which we are told – wrongly – is a term interchangeable with 'freedom'.

The cult of cults

At this point I need to introduce a major expression of the Global Cult known as Sabbatian-Frankism. Sabbatian is also spelt as Sabbatean. I will summarise here. I have published major exposés and detailed background in other works. Sabbatian-Frankism combines the names of two frauds posing as 'Jewish' men, Sabbatai Zevi (1626-1676), a rabbi, black magician and occultist who proclaimed he was the Jewish messiah; and Jacob Frank (1726-1791), the Polish 'Jew', black magician and occultist who said he was the reincarnation of 'messiah' Zevi and biblical patriarch Jacob. They worked across two centuries to establish the Sabbatian-Frankist cult that plays a major, indeed central, role in the manipulation of human society by the Global Cult which has its origins much further back in history than Sabbatai Zevi. I should emphasise two points here in response to the shrill voices that will scream 'anti-Semitism': (1) Sabbatian-Frankists are NOT Jewish and only pose as such to hide their cult behind a Jewish façade; and (2) my information about this cult has come from Jewish sources who have long realised that their society and community has been infiltrated and taken over by interloper Sabbatian-Frankists. Infiltration has been the foundation technique of Sabbatian-Frankism from its official origin in the 17th century. Zevi's Sabbatian sect attracted a massive following described as the biggest messianic movement in Jewish history, spreading as far as Africa and Asia, and he promised a return for the Jews to the 'Promised Land' of Israel. Sabbatianism was not Judaism but an inversion of everything that mainstream Judaism stood for. So much so that this sinister cult would have a feast day when Judaism had a fast day and whatever was forbidden in Judaism the Sabbatians were encouraged and even commanded to do. This included incest and what would be today called Satanism. Members were forbidden to marry outside the sect and there was a system of keeping their children ignorant of what they were part of until they were old enough to be trusted not to unknowingly reveal anything to outsiders. The same system is employed to this day by the Global Cult in general which Sabbatian-Frankism has enormously influenced and now largely controls.

Zevi and his Sabbatians suffered a setback with the intervention by the Sultan of the Islamic Ottoman Empire in the Middle East and what is now the Republic of Turkey where Zevi was located. The

Sultan gave him the choice of proving his ‘divinity’, converting to Islam or facing torture and death. Funnily enough Zevi chose to convert or at least appear to. Some of his supporters were disillusioned and drifted away, but many did not with 300 families also converting – only in theory – to Islam. They continued behind this Islamic smokescreen to follow the goals, rules and rituals of Sabbatianism and became known as ‘crypto-Jews’ or the ‘Dönmeh’ which means ‘to turn’. This is rather ironic because they didn’t ‘turn’ and instead hid behind a fake Islamic persona. The process of appearing to be one thing while being very much another would become the calling card of Sabbatianism especially after Zevi’s death and the arrival of the Satanist Jacob Frank in the 18th century when the cult became Sabbatian-Frankism and plumbed still new depths of depravity and infiltration which included – still includes – human sacrifice and sex with children. Wherever Sabbatians go paedophilia and Satanism follow and is it really a surprise that Hollywood is so infested with child abuse and Satanism when it was established by Sabbatian-Frankists and is still controlled by them? Hollywood has been one of the prime vehicles for global perceptual programming and manipulation. How many believe the version of ‘history’ portrayed in movies when it is a travesty and inversion (again) of the truth? Rabbi Marvin Antelman describes Frankism in his book, *To Eliminate the Opiate*, as ‘a movement of complete evil’ while Jewish professor Gershom Scholem said of Frank in *The Messianic Idea in Judaism*: ‘In all his actions [he was] a truly corrupt and degenerate individual ... one of the most frightening phenomena in the whole of Jewish history.’ Frank was excommunicated by traditional rabbis, as was Zevi, but Frank was undeterred and enjoyed vital support from the House of Rothschild, the infamous banking dynasty whose inner-core are Sabbatian-Frankists and not Jews. Infiltration of the Roman Church and Vatican was instigated by Frank with many Dönmeh ‘turning’ again to convert to Roman Catholicism with a view to hijacking the reins of power. This was the ever-repeating modus operandi and continues to be so. Pose as an advocate of the religion, culture or country that you want to control and then manipulate your people into the positions of authority and influence largely as advisers, administrators and Svengalis for those that appear to be in power. They did this with Judaism, Christianity (Christian Zionism is part of this), Islam and other religions and nations until Sabbatian-Frankism spanned the world as it does today.

Sabbatian Saudis and the terror network

One expression of the Sabbatian-Frankist Dönmeh within Islam is the ruling family of Saudi Arabia, the House of Saud, through which came the vile distortion of Islam known as Wahhabism. This is the violent creed followed by terrorist groups like Al-Qaeda and ISIS or Islamic State. Wahhabism is the hand-chopping, head-chopping 'religion' of Saudi Arabia which is used to keep the people in a constant state of fear so the interloper House of Saud can continue to rule. Al-Qaeda and Islamic State were lavishly funded by the House of Saud while being created and directed by the Sabbatian-Frankist network in the United States that operates through the Pentagon, CIA and the government in general of whichever 'party'. The front man for the establishment of Wahhabism in the middle of the 18th century was a Sabbatian-Frankist 'crypto-Jew' posing as Islamic called Muhammad ibn Abd al-Wahhab. His daughter would marry the son of Muhammad bin Saud who established the first Saudi state before his death in 1765 with support from the British Empire. Bin Saud's successors would establish modern Saudi Arabia in league with the British and Americans in 1932 which allowed them to seize control of Islam's major shrines in Mecca and Medina. They have dictated the direction of Sunni Islam ever since while Iran is the major centre of the Shiite version and here we have the source of at least the public conflict between them. The Sabbatian network has used its Wahhabi extremists to carry out Problem-Reaction-Solution terrorist attacks in the name of 'Al-Qaeda' and 'Islamic State' to justify a devastating 'war on terror', ever-increasing surveillance of the population and to terrify people into compliance. Another insight of the Renegade Mind is the streetwise understanding that just because a country, location or people are attacked doesn’t mean that those apparently representing that country, location or people are not behind the attackers. Often they are *orchestrating* the attacks because of the societal changes that can be then justified in the name of ‘saving the population from terrorists’.

I show in great detail in *The Trigger* how Sabbatian-Frankists were the real perpetrators of 9/11 and not ‘19 Arab hijackers’ who were blamed for what happened. Observe what was justified in the name of 9/11 alone in terms of Middle East invasions, mass surveillance and control that fulfilled the demands of the Project for the New American Century document published by the Sabbatian Neocons. What appear to be enemies are on the deep inside players on the same Sabbatian team. Israel and Arab ‘royal’ dictatorships are all ruled by Sabbatians and the recent peace agreements between Israel and Saudi Arabia, the United Arab Emirates (UAE) and others are only making formal what has always been the case behind the scenes. Palestinians who have been subjected to grotesque tyranny since Israel was bombed and terrorised into existence in 1948 have never stood a chance. Sabbatian-Frankists have controlled Israel (so the constant theme of violence and war which Sabbatians love) and they have controlled the Arab countries that Palestinians have looked to for real support that never comes. ‘Royal families’ of the Arab world in Saudi Arabia, Bahrain, UAE, etc., are all Sabbatians with allegiance to the aims of the cult and not what is best for their Arabic populations. They have stolen the oil and financial resources from their people by false claims to be ‘royal dynasties’ with a genetic right to rule and by employing vicious militaries to impose their will.

Satanic ‘illumination’

The Satanist Jacob Frank formed an alliance in 1773 with two other Sabbatians, Mayer Amschel Rothschild (1744-1812), founder of the Rothschild banking dynasty, and Jesuit-educated fraudulent Jew, Adam Weishaupt, and this led to the formation of the Bavarian Illuminati, firstly under another name, in 1776. The Illuminati would be the manipulating force behind the French Revolution (1789-1799) and was also involved in the American Revolution (1775-1783) before and after the Illuminati's official creation. Weishaupt would later become (in public) a Protestant Christian in archetypal Sabbatian style. I read that his name can be decoded as Adam-Weishaupt or 'the first man to lead those who know'. He wasn't a leader in the sense that he was a subordinate, but he did lead those below him in a crusade of transforming human society that still continues today. The theme was confirmed as early as 1785 when a horseman courier called Lanz was reported to be struck by lighting and extensive Illuminati documents were found in his saddlebags. They made the link to Weishaupt and detailed the plan for world takeover. Current events with 'Covid' fascism have been in the making for a very long time. Jacob Frank was jailed for 13 years by the Catholic Inquisition after his arrest in 1760 and on his release he headed for Frankfurt, Germany, home city and headquarters of the House of Rothschild where the alliance was struck with Mayer Amschel Rothschild and Weishaupt. Rothschild arranged for Frank to be given the title of Baron and he became a wealthy nobleman with a big following of Jews in Germany, the Austro-Hungarian Empire and other European countries. Most of them would have believed he was on their side.

The name 'Illuminati' came from the Zohar which is a body of works in the Jewish mystical 'bible' called the Kabbalah. 'Zohar' is the foundation of Sabbatian-Frankist belief and in Hebrew 'Zohar' means 'splendour', 'radiance', 'illuminated', and so we have 'Illuminati'. They claim to be the 'Illuminated Ones' from their knowledge systematically hidden from the human population and passed on through generations of carefully-chosen initiates in the global secret society network or Cult. Hidden knowledge includes an awareness of the Cult agenda for the world and the nature of our collective reality that I will explore later. Cult 'illumination' is symbolised by the torch held by the Statue of Liberty which was gifted to New York by French Freemasons in Paris who knew exactly what it represents. 'Liberty' symbolises the goddess worshipped in

Babylon as Queen Semiramis or Ishtar. The significance of this will become clear. Notice again the ubiquitous theme of inversion with the Statue of 'Liberty' really symbolising mass control (Fig 7). A mirror-image statute stands on an island in the River Seine in Paris from where New York Liberty originated (Fig 8). A large replica of the Liberty flame stands on top of the Pont de l'Alma tunnel in Paris where Princess Diana died in a Cult ritual described in *The Biggest Secret*. Lucifer 'the light bringer' is related to all this (and much more as we'll see) and 'Lucifer' is a central figure in Sabbatian-Frankism and its associated Satanism. Sabbatians reject the Jewish Torah, or Pentateuch, the 'five books of Moses' in the Old Testament known as Genesis, Exodus, Leviticus, Numbers, and Deuteronomy which are claimed by Judaism and Christianity to have been dictated by 'God' to Moses on Mount Sinai. Sabbatians say these do not apply to them and they seek to replace them with the Zohar to absorb Judaism and its followers into their inversion which is an expression of a much greater global inversion. They want to delete all religions and force humanity to worship a one-world religion – Sabbatian Satanism that also includes worship of the Earth goddess. Satanic themes are being more and more introduced into mainstream society and while Christianity is currently the foremost target for destruction the others are planned to follow.

Figure 7: The Cult goddess of Babylon disguised as the Statue of Liberty holding the flame of Lucifer the 'light bringer'.

Radio France

Figure 8: Liberty's mirror image in Paris where the New York version originated.

Marx brothers

Rabbi Marvin Antelman connects the Illuminati to the Jacobins in *To Eliminate the Opiate* and Jacobins were the force behind the French Revolution. He links both to the Bund der Gerechten, or League of the Just, which was the network that inflicted communism/Marxism on the world. Antelman wrote:

> The original inner circle of the Bund der Gerechten consisted of born Catholics, Protestants and Jews [Sabbatian-Frankist infiltrators], and those representatives of respective subdivisions formulated schemes for the ultimate destruction of their faiths. The heretical Catholics laid plans which they felt would take a century or more for the ultimate destruction of the church; the apostate Jews for the ultimate destruction of the Jewish religion.

Sabbatian-created communism connects into this anti-religion agenda in that communism does not allow for the free practice of religion. The Sabbatian 'Bund' became the International Communist Party and Communist League and in 1848 'Marxism' was born with the Communist Manifesto of Sabbatian assets Karl Marx and Friedrich Engels. It is absolutely no coincidence that Marxism, just a different name for fascist and other centrally-controlled tyrannies, is being imposed worldwide as a result of the 'Covid' hoax and nor that Marxist/fascist China was the place where the hoax originated. The reason for this will become very clear in the chapter 'Covid: The calculated catastrophe'. The so-called 'Woke' mentality has hijacked traditional beliefs of the political left and replaced them with far-right make-believe ‘social justice’ better known as Marxism. Woke will, however, be swallowed by its own perceived ‘revolution’ which is really the work of billionaires and billionaire corporations feigning being ‘Woke’. Marxism is being touted by Wokers as a replacement for ‘capitalism’ when we don’t have ‘capitalism’. We have cartelism in which the market is stitched up by the very Cult billionaires and corporations bankrolling Woke. Billionaires love Marxism which keeps the people in servitude while they control from the top. Terminally naïve Wokers think they are ‘changing the world’ when it’s the Cult that is doing the changing and when they have played their vital part and become surplus to requirements they, too, will be targeted. The Illuminati-Jacobins were behind the period known as ‘The Terror’ in the French Revolution in 1793 and 1794 when Jacobin Maximillian de Robespierre and his Orwellian ‘Committee of Public Safety’ killed 17,000 ‘enemies of the Revolution’ who had once been ‘friends of the Revolution’. Karl Marx (1818-1883), whose Sabbatian creed of Marxism has cost the lives of at least 100 million people, is a hero once again to Wokers who have been systematically kept ignorant of real history by their ‘education’ programming. As a result they now promote a Sabbatian ‘Marxist’ abomination destined at some point to consume them. Rabbi Antelman, who spent decades researching the Sabbatian plot, said of the League of the Just and Karl Marx:

> Contrary to popular opinion Karl Marx did not originate the Communist Manifesto. He was paid for his services by the League of the Just, which was known in its country of origin, Germany, as the Bund der Geaechteten.

Antelman said the text attributed to Marx was the work of other people and Marx ‘was only repeating what others already said’. Marx was ‘a hired hack – lackey of the wealthy Illuminists’. Marx famously said that religion was the ‘opium of the people’ (part of the Sabbatian plan to demonise religion) and Antelman called his books, *To Eliminate the Opiate*. Marx was born Jewish, but his family converted to Christianity (Sabbatian modus operandi) and he attacked Jews, not least in his book, *A World Without Jews*. In doing so he supported the Sabbatian plan to destroy traditional Jewishness and Judaism which we are clearly seeing today with the vindictive targeting of orthodox Jews by the Sabbatian government of Israel over 'Covid' laws. I don't follow any religion and it has done much damage to the world over centuries and acted as a perceptual straightjacket. Renegade Minds, however, are always asking *why* something is being done. It doesn't matter if they agree or disagree with what is happening – *why* is it happening is the question. The 'why?' can be answered with regard to religion in that religions create interacting communities of believers when the Cult wants to dismantle all discourse, unity and interaction (see 'Covid' lockdowns) and the ultimate goal is to delete all religions for a one-world religion of Cult Satanism worshipping their 'god' of which more later. We see the same 'why?' with gun control in America. I don't have guns and don't want them, but why is the Cult seeking to disarm the population at the same time that law enforcement agencies are armed to their molars and why has every tyrant in history sought to disarm people before launching the final takeover? They include Hitler, Stalin, Pol Pot and Mao who followed confiscation with violent seizing of power. You know it's a Cult agenda by the people who immediately race to the microphones to exploit dead people in multiple shootings. Ultra-Zionist Cult lackey Senator Chuck Schumer was straight on the case after ten people were killed in Boulder, Colorado in March, 2121. Simple rule … if Schumer wants it the Cult wants it and the same with his ultra-Zionist mate the wild-eyed Senator Adam Schiff. At the same time they were calling for the disarmament of Americans, many of whom live a long way from a police response, Schumer, Schiff and the rest of these pampered clowns were sitting on Capitol Hill behind a razor-wired security fence protected by thousands of armed troops in addition to their own armed bodyguards. Mom and pop in an isolated home? They're just potential mass shooters.

Zion Mainframe

Sabbatian-Frankists and most importantly the Rothschilds were behind the creation of 'Zionism', a political movement that demanded a Jewish homeland in Israel as promised by Sabbatai Zevi. The very symbol of Israel comes from the German meaning of the name Rothschild. Dynasty founder Mayer Amschel Rothschild changed the family name from Bauer to Rothschild, or 'Red-Shield' in German, in deference to the six-pointed 'Star of David' hexagram displayed on the family's home in Frankfurt. The symbol later appeared on the flag of Israel after the Rothschilds were centrally involved in its creation. Hexagrams are not a uniquely Jewish symbol and are widely used in occult ('hidden') networks often as a symbol for Saturn (see my other books for why). Neither are Zionism and Jewishness interchangeable. Zionism is a political movement and philosophy and not a 'race' or a people. Many Jews oppose Zionism and many non-Jews, including US President Joe Biden, call themselves Zionists as does Israel-centric Donald Trump. America's support for the Israel government is pretty much a gimme with ultra-Zionist billionaires and corporations providing fantastic and dominant funding for both political parties. Former Congresswoman Cynthia McKinney has told how she was approached immediately she ran for office to 'sign the pledge' to Israel and confirm that she would always vote in that country's best interests. All American politicians are approached in this way. Anyone who refuses will get no support or funding from the enormous and all-powerful Zionist lobby that includes organisations like mega-lobby group AIPAC, the American Israel Public Affairs Committee. Trump's biggest funder was ultra-Zionist casino and media billionaire Sheldon Adelson while major funders of the Democratic Party include ultra-Zionist George Soros and ultra-Zionist financial and media mogul, Haim Saban. Some may reel back at the suggestion that Soros is an Israel-firster (Sabbatian-controlled Israel-firster), but Renegade Minds watch the actions not the words and everywhere Soros donates his billions the Sabbatian agenda benefits. In the spirit of Sabbatian inversion Soros pledged $1 billion for a new university network to promote 'liberal values and tackle intolerance'. He made the announcement during his annual speech at the Cult-owned World Economic Forum in Davos, Switzerland, in January, 2020, after his ‘harsh criticism’ of ‘authoritarian rulers’ around the world. You can only laugh at such brazen mendacity. How *he* doesn’t laugh is the mystery. Translated from the Orwellian ‘liberal values and tackle intolerance’ means teaching non-white people to hate white people and for white people to loathe themselves for being born white. The reason for that will become clear.

The ‘Anti-Semitism’ fraud

Zionists support the Jewish homeland in the land of Palestine which has been the Sabbatian-Rothschild goal for so long, but not for the benefit of Jews. Sabbatians and their global Anti-Semitism Industry have skewed public and political opinion to equate opposing the violent extremes of Zionism to be a blanket attack and condemnation of all Jewish people. Sabbatians and their global Anti-Semitism Industry have skewed public and political opinion to equate opposing the violent extremes of Zionism to be a blanket attack and condemnation of all Jewish people. This is nothing more than a Sabbatian protection racket to stop legitimate investigation and exposure of their agendas and activities. The official definition of ‘anti-Semitism’ has more recently been expanded to include criticism of Zionism – a *political movement* – and this was done to further stop exposure of Sabbatian infiltrators who created Zionism as we know it today in the 19th century. Renegade Minds will talk about these subjects when they know the shit that will come their way. People must decide if they want to know the truth or just cower in the corner in fear of what others will say. Sabbatians have been trying to label me as ‘anti-Semitic’ since the 1990s as I have uncovered more and more about their background and agendas. Useless, gutless, fraudulent ‘journalists’ then just repeat the smears without question and on the day I was writing this section a pair of unquestioning repeaters called Ben Quinn and Archie Bland (how appropriate) outright called me an ‘anti-Semite’ in the establishment propaganda sheet, the London *Guardian*, with no supporting evidence. The

Sabbatian Anti-Semitism Industry said so and who are they to question that? They wouldn't dare. Ironically 'Semitic' refers to a group of languages in the Middle East that are almost entirely Arabic. 'Anti-Semitism' becomes 'anti-Arab' which if the consequences of this misunderstanding were not so grave would be hilarious. Don't bother telling Quinn and Bland. I don't want to confuse them, bless 'em. One reason I am dubbed 'anti-Semitic' is that I wrote in the 1990s that Jewish operatives (Sabbatians) were heavily involved in the Russian Revolution when Sabbatians overthrew the Romanov dynasty. This apparently made me 'anti-Semitic'. Oh, really? Here is a section from *The Trigger*:

> British journalist Robert Wilton confirmed these themes in his 1920 book *The Last Days of the Romanovs* when he studied official documents from the Russian government to identify the members of the Bolshevik ruling elite between 1917 and 1919. The Central Committee included 41 Jews among 62 members; the Council of the People's Commissars had 17 Jews out of 22 members; and 458 of the 556 most important Bolshevik positions between 1918 and 1919 were occupied by Jewish people. Only 17 were Russian. Then there were the 23 Jews among the 36 members of the vicious Cheka Soviet secret police established in 1917 who would soon appear all across the country.
>
> Professor Robert Service of Oxford University, an expert on 20th century Russian history, found evidence that ['Jewish'] Leon Trotsky had sought to make sure that Jews were enrolled in the Red Army and were disproportionately represented in the Soviet civil bureaucracy that included the Cheka which performed mass arrests, imprisonment and executions of 'enemies of the people'. A US State Department Decimal File (861.00/5339) dated November 13th, 1918, names [Rothschild banking agent in America] Jacob Schiff and a list of ultra-Zionists as funders of the Russian Revolution leading to claims of a 'Jewish plot', but the key point missed by all is they were not 'Jews' – they were Sabbatian-Frankists.

Britain's Winston Churchill made the same error by mistake or otherwise. He wrote in a 1920 edition of the *Illustrated Sunday Herald* that those behind the Russian revolution were part of a 'worldwide conspiracy for the overthrow of civilisation and for the reconstitution of society on the basis of arrested development, of envious malevolence, and impossible equality' (see 'Woke' today because that has been created by the same network). Churchill said there was no need to exaggerate the part played in the creation of Bolshevism and in the actual bringing about of the Russian

Revolution 'by these international and for the most part atheistical Jews' ['atheistical Jews' = Sabbatians]. Churchill said it is certainly a very great one and probably outweighs all others: 'With the notable exception of Lenin, the majority of the leading figures are Jews.' He went on to describe, knowingly or not, the Sabbatian modus operandi of placing puppet leaders nominally in power while they control from the background:

> Moreover, the principal inspiration and driving power comes from the Jewish leaders. Thus Tchitcherin, a pure Russian, is eclipsed by his nominal subordinate, Litvinoff, and the influence of Russians like Bukharin or Lunacharski cannot be compared with the power of Trotsky, or of Zinovieff, the Dictator of the Red Citadel (Petrograd), or of Krassin or Radek – all Jews. In the Soviet institutions the predominance of Jews is even more astonishing. And the prominent, if not indeed the principal, part in the system of terrorism applied by the Extraordinary Commissions for Combatting Counter-Revolution has been taken by Jews, and in some notable cases by Jewesses.

What I said about seriously disproportionate involvement in the Russian Revolution by Jewish 'revolutionaries' (Sabbatians) is provable fact, but truth is no defence against the Sabbatian Anti-Semitism Industry, its repeater parrots like Quinn and Bland, and the now breathtaking network of so-called 'Woke' 'anti-hate' groups with interlocking leaderships and funding which have the role of discrediting and silencing anyone who gets too close to exposing the Sabbatians. We have seen 'truth is no defence' confirmed in legal judgements with the Saskatchewan Human Rights Commission in Canada decreeing this: 'Truthful statements can be presented in a manner that would meet the definition of hate speech, and not all truthful statements must be free from restriction.' Most 'anti-hate' activists, who are themselves consumed by hatred, are too stupid and ignorant of the world to know how they are being used. They are far too far up their own virtue-signalling arses and it's far too dark for them to see anything.

The 'revolution' game

The background and methods of the 'Russian' Revolution are straight from the Sabbatian playbook seen in the French Revolution and endless others around the world that appear to start as a revolution of the people against tyrannical rule and end up with a regime change to more tyrannical rule overtly or covertly. Wars, terror attacks and regime overthrows follow the Sabbatian cult through history with its agents creating them as Problem-Reaction-Solutions to remove opposition on the road to world domination. Sabbatian dots connect the Rothschilds with the Illuminati, Jacobins of the French Revolution, the 'Bund' or League of the Just, the International Communist Party, Communist League and the Communist Manifesto of Karl Marx and Friedrich Engels that would lead to the Rothschild-funded Russian Revolution. The sequence comes under the heading of 'creative destruction' when you advance to your global goal by continually destroying the status quo to install a new status quo which you then also destroy. The two world wars come to mind. With each new status quo you move closer to your planned outcome. Wars and mass murder are to Sabbatians a collective blood sacrifice ritual. They are obsessed with death for many reasons and one is that death is an inversion of life. Satanists and Sabbatians are obsessed with death and often target churches and churchyards for their rituals. Inversion-obsessed Sabbatians explain the use of inverted symbolism including the *inverted* pentagram and *inverted* cross. The inversion of the cross has been related to targeting Christianity, but the cross was a religious symbol long before Christianity and its inversion is a statement about the Sabbatian mentality and goals more than any single religion.

Sabbatians operating in Germany were behind the rise of the occult-obsessed Nazis and the subsequent Jewish exodus from Germany and Europe to Palestine and the United States after World War Two. The Rothschild dynasty was at the forefront of this both as political manipulators and by funding the operation. Why would Sabbatians help to orchestrate the horrors inflicted on Jews by the Nazis and by Stalin after they organised the Russian Revolution? Sabbatians hate Jews and their religion, that's why. They pose as Jews and secure positions of control within Jewish society and play the 'anti-Semitism' card to protect themselves from exposure through a global network of organisations answering to the Sabbatian-created-and-controlled globe-spanning intelligence network that involves a stunning web of military-intelligence operatives and operations for a tiny country of just nine million. Among them are Jewish assets who are not Sabbatians but have been convinced by them that what they are doing is for the good of Israel and the Jewish community to protect them from what they have been programmed since childhood to believe is a Jew-hating hostile world. The Jewish community is just a highly convenient cover to hide the true nature of Sabbatians. Anyone getting close to exposing their game is accused by Sabbatian place-people and gofers of 'anti-Semitism' and claiming that all Jews are part of a plot to take over the world. I am not saying that. I am saying that Sabbatians – the *real* Jew-haters – have infiltrated the Jewish community to use them both as a cover and an 'anti-Semitic' defence against exposure. Thus we have the Anti-Semitism Industry targeted researchers in this way and most Jewish people think this is justified and genuine. They don't know that their 'Jewish' leaders and institutions of state, intelligence and military are not controlled by Jews at all, but cultists and stooges of Sabbatian-Frankism. I once added my name to a pro-Jewish freedom petition online and the next time I looked my name was gone and text had been added to the petition blurb to attack me as an 'anti-Semite' such is the scale of perceptual programming.

Moving on America

I tell the story in *The Trigger* and a chapter called 'Atlantic Crossing' how particularly after Israel was established the Sabbatians moved in on the United States and eventually grasped control of government administration, the political system via both Democrats and Republicans, the intelligence community like the CIA and National Security Agency (NSA), the Pentagon and mass media. Through this seriously compartmentalised network Sabbatians and their operatives in Mossad, Israeli Defense Forces (IDF) and US agencies pulled off 9/11 and blamed it on 19 'Al-Qaeda hijackers' dominated by men from, or connected to, Sabbatian-ruled Saudi

Arabia. The '19' were not even on the planes let alone flew those big passenger jets into buildings while being largely incompetent at piloting one-engine light aircraft. 'Hijacker' Hani Hanjour who is said to have flown American Airlines Flight 77 into the Pentagon with a turn and manoeuvre most professional pilots said they would have struggled to do was banned from renting a small plane by instructors at the Freeway Airport in Bowie, Maryland, just *six weeks* earlier on the grounds that he was an incompetent pilot. The Jewish population of the world is just 0.2 percent with even that almost entirely concentrated in Israel (75 percent Jewish) and the United States (around two percent). This two percent and globally 0.2 percent refers to *Jewish* people and not Sabbatian interlopers who are a fraction of that fraction. What a sobering thought when you think of the fantastic influence on world affairs of tiny Israel and that the Project for the New America Century (PNAC) which laid out the blueprint in September, 2000, for America's war on terror and regime change wars in Iraq, Libya and Syria was founded and dominated by Sabbatians known as 'Neocons'. The document conceded that this plan would not be supported politically or publicly without a major attack on American soil and a Problem-Reaction-Solution excuse to send troops to war across the Middle East. Sabbatian Neocons said:

> ... [The] process of transformation ... [war and regime change] ... is likely to be a long one, absent some catastrophic and catalysing event – like a new Pearl Harbor.

Four months later many of those who produced that document came to power with their inane puppet George Bush from the long-time Sabbatian Bush family. They included Sabbatian Dick Cheney who was officially vice-president, but really de-facto president for the entirety of the 'Bush' government. Nine months after the 'Bush' inauguration came what Bush called at the time 'the Pearl Harbor of the 21st century' and with typical Sabbatian timing and symbolism 2001 was the 60th anniversary of the attack in 1941 by the Japanese Air Force on Pearl Harbor, Hawaii, which allowed President Franklin Delano Roosevelt to take the United States into a Sabbatian-instigated Second World War that he said in his election campaign that he never would. The evidence is overwhelming that Roosevelt and his military and intelligence networks knew the attack was coming and did nothing to stop it, but they did make sure that America's most essential naval ships were not in Hawaii at the time. Three thousand Americans died in the Pearl Harbor attacks as they did on September 11th. By the 9/11 year of 2001 Sabbatians had widely infiltrated the US government, military and intelligence operations and used their compartmentalised assets to pull off the 'Al-Qaeda' attacks. If you read *The Trigger* it will blow your mind to see the utterly staggering concentration of 'Jewish' operatives (Sabbatian infiltrators) in essential positions of political, security, legal, law enforcement, financial and business power before, during, and after the attacks to make them happen, carry them out, and then cover their tracks – and I do mean *staggering* when you think of that 0.2 percent of the world population and two percent of Americans which are Jewish while Sabbatian infiltrators are a fraction of that. A central foundation of the 9/11 conspiracy was the hijacking of government, military, Air Force and intelligence computer systems in real time through 'back-door' access made possible by Israeli (Sabbatian) 'cyber security' software. Sabbatian-controlled Israel is on the way to rivalling Silicon Valley for domination of cyberspace and is becoming the dominant force in cyber-security which gives them access to entire computer systems and their passcodes across the world. Then add to this that Zionists head (officially) Silicon Valley giants like Google (Larry Page and Sergey Brin), Google-owned YouTube (Susan Wojcicki), Facebook (Mark Zuckerberg and Sheryl Sandberg), and Apple (Chairman Arthur D. Levinson), and that ultra-Zionist hedge fund billionaire Paul Singer has a $1 billion stake in Twitter which is only nominally headed by 'CEO' pothead Jack Dorsey. As cable news host Tucker Carlson said of Dorsey: 'There used to be debate in the medical community whether dropping a ton of acid had permanent effects and I think that debate has now ended.' Carlson made the comment after Dorsey told a hearing on Capitol Hill (if you cut through his bullshit) that he believed in free speech so long as he got to decide what you can hear and see. These 'big names' of Silicon Valley are only front men and women for the Global Cult, not least the Sabbatians, who are the true controllers of these corporations. Does anyone still wonder why these same people and companies have been ferociously censoring and banning people (like me) for exposing any aspect of the Cult agenda and especially the truth about the 'Covid' hoax which Sabbatians have orchestrated?

The Jeffrey Epstein paedophile ring was a Sabbatian operation. He was officially 'Jewish' but he was a Sabbatian and women abused by the ring have told me about the high number of 'Jewish' people involved. The Epstein horror has Sabbatian written all over it and matches perfectly their modus operandi and obsession with sex and ritual. Epstein was running a Sabbatian blackmail ring in which famous people with political and other influence were provided with young girls for sex while everything was being filmed and recorded on hidden cameras and microphones at his New York house, Caribbean island and other properties. Epstein survivors have described this surveillance system to me and some have gone public. Once the famous politician or other figure knew he or she was on video they tended to do whatever they were told. Here we go again …when you've got them by the balls their hearts and minds will follow. Sabbatians use this blackmail technique on a wide scale across the world to entrap politicians and others they need to act as demanded. Epstein's private plane, the infamous 'Lolita Express', had many well-known passengers including Bill Clinton while Bill Gates has flown on an Epstein plane and met with him four years after Epstein had been jailed for paedophilia. They subsequently met many times at Epstein's home in New York according to a witness who was there. Epstein's infamous side-kick was Ghislaine Maxwell, daughter of Mossad agent and ultra-Zionist mega-crooked British businessman, Bob Maxwell, who at one time owned the *Daily Mirror* newspaper. Maxwell was murdered at sea on his boat in 1991 by Sabbatian-controlled Mossad when he became a liability with his business empire collapsing as a former Mossad operative has confirmed (see *The Trigger*).

Money, money, money, funny money ...

Before I come to the Sabbatian connection with the last three US presidents I will lay out the crucial importance to Sabbatians of controlling banking and finance. Sabbatian Mayer Amschel Rothschild set out to dominate this arena in his family's quest for total global control. What is freedom? It is, in effect, choice. The more choices you have the freer you are and the fewer your choices the more you are enslaved. In the global structure created over centuries by Sabbatians the biggest decider and restrictor of choice is … money. Across the world if you ask people what they would like to do with their lives and why they are not doing that they will reply 'I don't have the money'. This is the idea. A global elite of multi-billionaires are described as 'greedy' and that is true on one level; but control of money – who has it and who doesn't – is not primarily about greed. It's about control. Sabbatians have seized ever more control of finance and sucked the wealth of the world out of the hands of the population. We talk now, after all, about the 'One-percent' and even then the wealthiest are a lot fewer even than that. This has been made possible by a money scam so outrageous and so vast it could rightly be called the scam of scams founded on creating 'money' out of nothing and 'loaning' that with interest to the population. Money out of nothing is called 'credit'. Sabbatians have asserted control over governments and banking ever more completely through the centuries and secured financial laws that allow banks to lend hugely more than they have on deposit in a confidence trick known as fractional reserve lending. Imagine if you could lend money that doesn't exist and charge the recipient interest for doing so. You would end up in jail. Bankers by contrast end up in mansions, private jets, Malibu and Monaco.

Banks are only required to keep a fraction of their deposits and wealth in their vaults and they are allowed to lend 'money' they don't have called 'credit. Go into a bank for a loan and if you succeed the banker will not move any real wealth into your account. They will type into your account the amount of the agreed 'loan' – say £100,000. This is not wealth that really exists; it is non-existent, fresh-air, created-out-of-nothing 'credit' which has never, does not, and will never exist except in theory. Credit is backed by nothing except wind and only has buying power because people think that it has buying power and accept it in return for property, goods and services. I have described this situation as like those cartoon characters you see chasing each other and when they run over the edge of a cliff they keep running forward on fresh air until one of them looks down, realises what's happened, and they all crash into the ravine. The whole foundation of the Sabbatian financial system is to stop people looking down except for periodic moments when they want to crash the system (as in 2008 and 2020 ongoing) and reap the rewards from all the property, businesses and wealth their borrowers had signed over as 'collateral' in return for a 'loan' of fresh air. Most people think that money is somehow created by governments when it comes into existence from the start as a debt through banks 'lending' illusory money called credit. Yes, the very currency of exchange is a *debt* from day one issued as an interest-bearing loan. Why don't governments create money interest-free and lend it to their people interest-free? Governments are controlled by Sabbatians and the financial system is controlled by Sabbatians for whom interest-free money would be a nightmare come true. Sabbatians underpin their financial domination through their global network of central banks, including the privately-owned US Federal Reserve and Britain's Bank of England, and this is orchestrated by a privately-owned central bank coordination body called the Bank for International Settlements in Basle, Switzerland, created by the usual suspects including the Rockefellers and Rothschilds. Central bank chiefs don't answer to governments or the people. They answer to the Bank for International Settlements or, in other words, the Global Cult which is dominated today by Sabbatians.

Built-in disaster

There are so many constituent scams within the overall banking scam. When you take out a loan of thin-air credit only the amount of that loan is theoretically brought into circulation to add to the amount in circulation; but you are paying back the principle plus interest. The additional interest is not created and this means that with every 'loan' there is a shortfall in the money in circulation between what is borrowed and what has to be paid back. There is never even close to enough money in circulation to repay all outstanding public and private debt including interest. Coldly weaved in the very fabric of the system is the certainty that some will lose their homes, businesses and possessions to the banking 'lender'. This is less obvious in times of 'boom' when the amount of money in circulation (and the debt) is expanding through more people wanting and getting loans. When a downturn comes and the money supply contracts it becomes painfully obvious that there is not enough money to service all debt and interest. This is less obvious in times of 'boom' when the amount of money in circulation (and the debt) is expanding through more people wanting and getting loans. When a downturn comes and the money supply contracts and it becomes painfully obvious – as in 2008 and currently – that there is not enough money to service all debt and interest. Sabbatian banksters have been leading the human population through a calculated series of booms (more debt incurred) and busts (when the debt can't be repaid and the banks get the debtor's tangible wealth in exchange for non-existent 'credit'). With each 'bust' Sabbatian bankers have absorbed more of the world's tangible wealth and we end up with the One-percent. Governments are in bankruptcy levels of debt to the same system and are therefore owned by a system they do not control. The Federal Reserve, 'America's central bank', is privately-owned and American presidents only nominally appoint its chairman or woman to maintain the illusion that it's an arm of government. It's not. The 'Fed' is a cartel of private banks which handed billions to its associates and friends after the crash of 2008 and has been Sabbatian-controlled since it was manipulated into being in 1913 through the covert trickery of Rothschild banking agents Jacob Schiff and Paul

Warburg, and the Sabbatian Rockefeller family. Somehow from a Jewish population of two-percent and globally 0.2 percent (Sabbatian interlopers remember are far smaller) ultra-Zionists headed the Federal Reserve for 31 years between 1987 and 2018 in the form of Alan Greenspan, Bernard Bernanke and Janet Yellen (now Biden's Treasury Secretary) with Yellen's deputy chairman a Israeli-American duel citizen and ultra-Zionist Stanley Fischer, a former governor of the Bank of Israel. Ultra-Zionist Fed chiefs spanned the presidencies of Ronald Reagan ('Republican'), Father George Bush ('Republican'), Bill Clinton ('Democrat'), Boy George Bush ('Republican') and Barack Obama ('Democrat'). We should really add the pre-Greenspan chairman, Paul Adolph Volcker, 'appointed' by Jimmy Carter ('Democrat') who ran the Fed between 1979 and 1987 during the Carter and Reagan administrations before Greenspan took over. Volcker was a long-time associate and business partner of the Rothschilds. No matter what the 'party' officially in power the United States economy was directed by the same force. Here are members of the Obama, Trump and Biden administrations and see if you can make out a common theme.

Barack Obama ('Democrat')

Ultra-Zionists Robert Rubin, Larry Summers, and Timothy Geithner ran the US Treasury in the Clinton administration and two of them reappeared with Obama. Ultra-Zionist Fed chairman Alan Greenspan had manipulated the crash of 2008 through deregulation and jumped ship just before the disaster to make way for ultra-Zionist Bernard Bernanke to hand out trillions to Sabbatian 'too big to fail' banks and businesses, including the ubiquitous ultra-Zionist Goldman Sachs which has an ongoing staff revolving door operation between itself and major financial positions in government worldwide. Obama inherited the fallout of the crash when he took office in January, 2009, and fortunately he had the support of his ultra-Zionist White House Chief of Staff Rahm Emmanuel, son of a terrorist who helped to bomb Israel into being in 1948, and his ultra-Zionist senior adviser David Axelrod, chief strategist in Obama's two successful presidential campaigns. Emmanuel, later mayor of Chicago and former senior fundraiser and strategist for Bill Clinton, is an example of the Sabbatian policy after Israel was established of migrating insider families to America so their children would be born American citizens. 'Obama' chose this financial team throughout his administration to respond to the Sabbatian-instigated crisis:

Timothy Geithner (ultra-Zionist) Treasury Secretary; Jacob J. Lew, Treasury Secretary; Larry Summers (ultra-Zionist), director of the White House National Economic Council; Paul Adolph Volcker (Rothschild business partner), chairman of the Economic Recovery Advisory Board; Peter Orszag (ultra-Zionist), director of the Office of Management and Budget overseeing all government spending; Penny Pritzker (ultra-Zionist), Commerce Secretary; Jared Bernstein (ultra-Zionist), chief economist and economic policy adviser to Vice President Joe Biden; Mary Schapiro (ultra-Zionist), chair of the Securities and Exchange Commission (SEC); Gary Gensler (ultra-Zionist), chairman of the Commodity Futures Trading Commission (CFTC); Sheila Bair (ultra-Zionist), chair of the Federal Deposit Insurance Corporation (FDIC); Karen Mills (ultra-Zionist), head of the Small Business Administration (SBA); Kenneth Feinberg (ultra-Zionist), Special Master for Executive [bail-out] Compensation. Feinberg would be appointed to oversee compensation (with strings) to 9/11 victims and families in a campaign to stop them having their day in court to question the official story. At the same time ultra-Zionist Bernard Bernanke was chairman of the Federal Reserve and these are only some of the ultra-Zionists with allegiance to Sabbatian-controlled Israel in the Obama government. Obama's biggest corporate donor was ultra-Zionist Goldman Sachs which had employed many in his administration.

Donald Trump ('Republican')

Trump claimed to be an outsider (he wasn't) who had come to 'drain the swamp'. He embarked on this goal by immediately appointing ultra-Zionist Steve Mnuchin, a Goldman Sachs employee for 17 years, as his Treasury Secretary. Others included Gary Cohn (ultra-Zionist), chief operating officer of Goldman Sachs, his first Director of the National Economic Council and chief economic adviser, who was later replaced by Larry Kudlow (ultra-Zionist). Trump's senior adviser throughout his four years in the White House was his sinister son-in-law Jared Kushner, a life-long friend of Israel Prime Minister Benjamin Netanyahu. Kushner is the son of a convicted crook who was pardoned by Trump in his last days in office. Other ultra-Zionists in the Trump administration included: Stephen Miller, Senior Policy Adviser; Avrahm Berkowitz, Deputy Adviser to Trump and his Senior Adviser Jared Kushner; Ivanka Trump, Adviser to the President, who converted to Judaism when she married Jared Kushner; David Friedman, Trump lawyer and Ambassador to Israel; Jason Greenblatt, Trump Organization executive vice president and chief legal officer, who was made Special Representative for International Negotiations and the Israeli-Palestinian Conflict; Rod Rosenstein, Deputy Attorney General; Elliot Abrams, Special Representative for Venezuela, then Iran; John Eisenberg, National Security Council Legal Adviser and Deputy Council to the President for National Security Affairs; Anne Neuberger, Deputy National Manager, National Security Agency; Ezra Cohen-Watnick, Acting Under Secretary of Defense for Intelligence; Elan Carr, Special Envoy to monitor and combat anti-Semitism; Len Khodorkovsky, Deputy Special Envoy to monitor and combat anti-Semitism; Reed Cordish, Assistant to the President, Intragovernmental and Technology Initiatives. Trump Vice President Mike Pence and Secretary of State Mike Pompeo, both Christian Zionists, were also vehement supporters of Israel and its goals and ambitions.

Donald 'free-speech believer' Trump pardoned a number of financial and violent criminals while ignoring calls to pardon Julian Assange and Edward Snowden whose crimes are revealing highly relevant information about government manipulation and corruption and the widespread illegal surveillance of the American people by US 'security' agencies. It's so good to know that Trump is on the side of freedom and justice and not mega-criminals with allegiance to Sabbatian-controlled Israel. These included a pardon for Israeli spy Jonathan Pollard who was jailed for life in 1987 under the Espionage Act. Aviem Sella, the Mossad agent who recruited Pollard, was also pardoned by Trump while Assange sat in jail and Snowden remained in exile in Russia. Sella had 'fled' (was helped to escape) to Israel in 1987 and was never extradited despite being charged under the Espionage Act. A Trump White House statement said that Sella's clemency had been 'supported by Benjamin Netanyahu, Ron Dermer, Israel's US Ambassador, David Friedman, US Ambassador to Israel and Miriam Adelson, wife of leading Trump donor Sheldon Adelson who died shortly before. Other friends of Jared Kushner were pardoned along with Sholom Weiss who was believed to be serving the longest-ever white-collar prison sentence of more than 800 years in 2000. The sentence was commuted of Ponzi-schemer Eliyahu Weinstein who defrauded Jews and others out of $200 million. I did mention that Assange and Snowden were ignored, right? Trump gave Sabbatians almost everything they asked for in military and political support, moving the US Embassy from Tel Aviv to Jerusalem with its critical symbolic and literal implications for Palestinian statehood, and the 'deal of the Century' designed by Jared Kushner and David Friedman which gave the Sabbatian Israeli government the green light to substantially expand its already widespread program of building illegal Jewish-only settlements in the occupied land of the West Bank. This made a two-state 'solution' impossible by seizing all the land of a potential Palestinian homeland and that had been the plan since 1948 and then 1967 when the Arab-controlled Gaza Strip, West Bank, Sinai Peninsula and Syrian Golan Heights were occupied by Israel. All the talks about talks and road maps and delays have been buying time until the West Bank was physically occupied by Israeli real estate. Trump would have to be a monumentally ill-informed idiot not to see that this was the plan he was helping to complete. The Trump administration was in so many ways the Kushner administration which means the Netanyahu administration which means the Sabbatian administration. I understand why many opposing Cult fascism in all its forms gravitated to Trump, but he was a crucial part of the Sabbatian plan and I will deal with this in the next chapter.

Joe Biden ('Democrat')

A barely cognitive Joe Biden took over the presidency in January, 2021, along with his fellow empty shell, Vice-President Kamala Harris, as the latest Sabbatian gofers to enter the White House. Names on the door may have changed and the 'party' – the force behind them remained the same as Zionists were appointed to a stream of pivotal areas relating to Sabbatian plans and policy. They included: Janet Yellen, Treasury Secretary, former head of the Federal Reserve, and still another ultra-Zionist running the US Treasury after Mnuchin (Trump), Lew and Geithner (Obama), and Summers and Rubin (Clinton); Anthony Blinken, Secretary of State; Wendy Sherman, Deputy Secretary of State (so that's 'Biden's' Sabbatian foreign policy sorted); Jeff Zients, White House coronavirus coordinator; Rochelle Walensky, head of the Centers for Disease Control; Rachel Levine, transgender deputy health secretary (that's 'Covid' hoax policy under control); Merrick Garland, Attorney General; Alejandro Mayorkas, Secretary of Homeland Security; Cass Sunstein, Homeland Security with responsibility for new immigration laws; Avril Haines, Director of National Intelligence; Anne Neuberger, National Security Agency cybersecurity director (note, cybersecurity); David Cohen, CIA Deputy Director; Ronald Klain, Biden's Chief of Staff (see Rahm Emanuel); Eric Lander, a 'leading geneticist', Office of Science and Technology Policy director (see Smart Grid, synthetic biology agenda); Jessica Rosenworcel, acting head of the Federal Communications Commission (FCC) which controls Smart Grid technology policy and electromagnetic communication systems including 5G. How can it be that so many pivotal positions are held by two-percent of the American population and 0.2 percent of the world population administration after administration no matter who is the president and what is the party? It's a coincidence? Of course it's not and this is why Sabbatians have built their colossal global web of interlocking 'anti-hate' hate groups to condemn anyone who asks these glaring questions as an 'anti-Semite'. The way that Jewish people horrifically abused in Sabbatian-backed Nazi Germany are exploited to this end is stomach-turning and disgusting beyond words.

Political fusion

Sabbatian manipulation has reversed the roles of Republicans and Democrats and the same has happened in Britain with the Conservative and Labour Parties. Republicans and Conservatives were always labelled the 'right' and Democrats and Labour the 'left', but look at the policy positions now and the Democrat-Labour 'left' has moved further to the 'right' than Republicans and Conservatives under the banner of 'Woke', the Cult-created far-right tyranny. Where once the Democrat-Labour 'left' defended free speech and human rights they now seek to delete them and as I said earlier despite the 'Covid' fascism of the Jackboot Johnson Conservative government in the UK the Labour Party of leader Keir Starmer demanded even more extreme measures. The Labour Party has been very publicly absorbed by Sabbatians after a political and media onslaught against the previous leader, the weak and inept Jeremy Corbyn, over made-up allegations of 'anti-Semitism' both by him and his party. The plan was clear with this 'anti-Semite' propaganda and what was required in response was a swift and decisive 'fuck off' from Corbyn and a statement to expose the Anti-Semitism Industry (Sabbatian) attempt to silence Labour criticism of the Israeli government (Sabbatians) and purge the party of all dissent against the extremes of ultra-Zionism (Sabbatians). Instead Corbyn and his party fell to their knees and appeased the abusers which, by definition, is impossible. Appeasing one demand leads only to a new demand to be appeased until takeover is complete. Like I say – 'fuck off' would have been a much more effective policy and I have used it myself with great effect over the years when Sabbatians are on my case which is most of the time. I consider that fact a great compliment, by the way. The outcome of the Labour Party capitulation is that we now have a Sabbatian-controlled

Conservative Party ‘opposed’ by a Sabbatian-controlled Labour Party in a one-party Sabbatian state that hurtles towards the extremes of tyranny (the Sabbatian cult agenda). In America the situation is the same. Labour’s Keir Starmer spends his days on his knees with his tongue out pointing to Tel Aviv, or I guess now Jerusalem, while Boris Johnson has an ‘anti-Semitism czar’ in the form of former Labour MP John Mann who keeps Starmer company on his prayer mat.

Sabbatian influence can be seen in Jewish members of the Labour Party who have been ejected for criticism of Israel including those from families that suffered in Nazi Germany. Sabbatians despise real Jewish people and target them even more harshly because it is so much more difficult to dub them ‘anti-Semitic’ although in their desperation they do try.

CHAPTER THREE

The Pushbacker sting

Until you realize how easy it is for your mind to be manipulated, you remain the puppet of someone else's game
Evita Ochel

I will use the presidencies of Trump and Biden to show how the manipulation of the one-party state plays out behind the illusion of political choice across the world. No two presidencies could – on the face of it – be more different and apparently at odds in terms of direction and policy.

A Renegade Mind sees beyond the obvious and focuses on outcomes and consequences and not image, words and waffle. The Cult embarked on a campaign to divide America between those who blindly support its agenda (the mentality known as 'Woke') and those who are pushing back on where the Cult and its Sabbatians want to go. This presents infinite possibilities for dividing and ruling the population by setting them at war with each other and allows a perceptual ring fence of demonisation to encircle the Pushbackers in a modern version of the Little Big Horn in 1876 when American cavalry led by Lieutenant Colonel George Custer were drawn into a trap, surrounded and killed by Native American tribes defending their land of thousands of years from being seized by the government. In this modern version the roles are reversed and it's those defending themselves from the Sabbatian government who are surrounded and the government that's seeking to destroy them. This trap was set years ago and to explain how we must return to 2016 and the emergence of Donald Trump as a candidate to be President of the United States. He set out to overcome the best part of 20 other candidates in the Republican Party before and during the primaries and was not considered by many in those early stages to have a prayer of living in the White House. The Republican Party was said to have great reservations about Trump and yet somehow he won the nomination. When you know how American politics works – politics in general – there is no way that Trump could have become the party's candidate unless the Sabbatian-controlled 'Neocons' that run the Republican Party wanted that to happen. We saw the proof in emails and documents made public by WikiLeaks that the Democratic Party hierarchy, or Democons, systematically undermined the campaign of Bernie Sanders to make sure that Sabbatian gofer Hillary Clinton won the nomination to be their presidential candidate. If the Democons could do that then the Neocons in the Republican Party could have derailed Trump in the same way. But they didn't and at that stage I began to conclude that Trump could well be the one chosen to be president. If that was the case the 'why' was pretty clear to see – the goal of dividing America between Cult agenda-supporting Wokers and Pushbackers who gravitated to Trump because he was telling them what they wanted to hear. His constituency of support had been increasingly ignored and voiceless for decades and profoundly through the eight years of Sabbatian puppet Barack Obama. Now here was someone speaking their language of pulling back from the incessant globalisation of political and economic power, the exporting of American jobs to China and elsewhere by 'American' (Sabbatian) corporations, the deletion of free speech, and the mass immigration policies that had further devastated job opportunities for the urban working class of all races and the once American heartlands of the Midwest.

Beware the forked tongue

Those people collectively sighed with relief that at last a political leader was apparently on their side, but another trait of the Renegade Mind is that you look even harder at people telling you what you want to hear than those who are telling you otherwise. Obviously as I said earlier people wish what they want to hear to be true and genuine and they are much more likely to believe that than someone saying what they don't want to here and don't want to be true. Sales people are taught to be skilled in eliciting by calculated questioning what their customers want to hear and repeating that back to them as their own opinion to get their targets to like and trust them. Assets of the Cult are also sales people in the sense of selling perception. To read Cult manipulation you have to play the long and expanded game and not fall for the Vaudeville show of party politics. Both American parties are vehicles for the Cult and they exploit them in different ways depending on what the agenda requires at that moment. Trump and the Republicans were used to be the focus of dividing America and isolating Pushbackers to open the way for a Biden presidency to become the most extreme in American history by advancing the full-blown Woke (Cult) agenda with the aim of destroying and silencing Pushbackers now labelled Nazi Trump supporters and white supremacists.

Sabbatians wanted Trump in office for the reasons described by ultra-Zionist Saul Alinsky (1909-1972) who was promoting the Woke philosophy through 'community organising' long before anyone had heard of it. In those days it still went by its traditional name of Marxism. The reason for the manipulated Trump phenomenon was laid out in Alinsky's 1971 book, *Rules for Radicals*, which was his blueprint for overthrowing democratic and other regimes and replacing them with Sabbatian Marxism. Not surprisingly his to-do list was evident in the Sabbatian French and Russian 'Revolutions' and that in China which will become very relevant in the next chapter about the 'Covid' hoax. Among Alinsky's followers have been the deeply corrupt Barack Obama, House Speaker Nancy Pelosi and Hillary Clinton who described him as a 'hero'. All three are Sabbatian stooges with Pelosi personifying the arrogant corrupt idiocy that so widely fronts up for the Cult inner core. Predictably as a Sabbatian advocate of the 'light-bringer' Alinsky features Lucifer on the dedication page of his book as the original radical who gained his own kingdom ('Earth' as we shall see). One of Alinsky's golden radical rules was to pick an individual and focus all attention, hatred and blame on them and not to target faceless bureaucracies and corporations. *Rules for Radicals* is really a Sabbatian handbook with its contents repeatedly employed all over the world for centuries and why wouldn't Sabbatians bring to power their designer-villain to be used as the individual on which all attention, hatred and blame was bestowed? This is what they did and the only question for me is how much Trump knew that and how much he was manipulated. A bit of both, I suspect. This was Alinsky's Trump technique from a man who died in 1972. The technique has spanned history:

> Pick the target, freeze it, personalize it, polarize it. Don't try to attack abstract corporations or bureaucracies. Identify a responsible individual. Ignore attempts to shift or spread the blame.

From the moment Trump came to illusory power everything was about him. It wasn't about Republican policy or opinion, but all about Trump. Everything he did was presented in negative, derogatory and abusive terms by the Sabbatian-dominated media led by Cult operations such as CNN, MSNBC, *The New York Times* and the Jeff Bezos-owned *Washington Post* – 'Pick the target, freeze it, personalize it, polarize it.' Trump was turned into a demon to be vilified by those who hated him and a demi-god loved by those who worshipped him. This, in turn, had his supporters, too, presented as equally demonic in preparation for the punchline later down the line when Biden was about to take office. It was here's a Trump, there's a Trump, everywhere a Trump, Trump. Virtually every news story or happening was filtered through the lens of 'The Donald'. You loved him or hated him and which one you chose was said to define you as Satan's spawn or a paragon of virtue. Even supporting some Trump policies or statements and not others was enough for an assault on your character. No shades of grey were or are allowed. Everything is black and white (literally and figuratively). A Californian I knew had her head utterly scrambled by her hatred for Trump while telling people they should love each other. She was so totally consumed by

Trump Derangement Syndrome as it became to be known that this glaring contradiction would never have occurred to her. By definition anyone who criticised Trump or praised his opponents was a hero and this lady described Joe Biden as 'a kind, honest gentleman' when he's a provable liar, mega-crook and vicious piece of work to boot. Sabbatians had indeed divided America using Trump as the fall-guy and all along the clock was ticking on the consequences for his supporters.

In hock to his masters

Trump gave Sabbatians via Israel almost everything they wanted in his four years. Ask and you shall receive was the dynamic between himself and Benjamin Netanyahu orchestrated by Trump's ultra-Zionist son-in-law Jared Kushner, his ultra-Zionist Ambassador to Israel, David Friedman, and ultra-Zionist 'Israel adviser', Jason Greenblatt. The last two were central to the running and protecting from collapse of his business empire, the Trump Organisation, and colossal business failures made him forever beholding to Sabbatian networks that bailed him out. By the start of the 1990s Trump owed $4 billion to banks that he couldn't pay and almost $1billion of that was down to him personally and not his companies. This mega-disaster was the result of building two new casinos in Atlantic City and buying the enormous Taj Mahal operation which led to crippling debt payments. He had borrowed fantastic sums from 72 banks with major Sabbatian connections and although the scale of debt should have had him living in a tent alongside the highway they never foreclosed. A plan was devised to lift Trump from the mire by BT Securities Corporation and Rothschild Inc. and the case was handled by Wilber Ross who had worked for the Rothschilds for 27 years. Ross would be named US Commerce Secretary after Trump's election. Another crucial figure in saving Trump was ultra-Zionist 'investor' Carl Icahn who bought the Taj Mahal casino. Icahn was made special economic adviser on financial regulation in the Trump administration. He didn't stay long but still managed to find time to make a tidy sum of a reported $31.3 million when he sold his holdings affected by the price of steel three days before Trump imposed a 235 percent tariff on steel imports. What amazing bits of luck these people have. Trump and Sabbatian operatives have long had a close association and his mentor and legal adviser from the early 1970s until 1986 was the dark and genetically corrupt ultra-Zionist Roy Cohn who was chief counsel to Senator Joseph McCarthy's 'communist' witch-hunt in the 1950s. *Esquire* magazine published an article about Cohn with the headline 'Don't mess with Roy Cohn'. He was described as the most feared lawyer in New York and 'a ruthless master of dirty tricks ... [with] ... more than one Mafia Don on speed dial'. Cohn's influence, contacts, support and protection made Trump a front man for Sabbatians in New York with their connections to one of Cohn's many criminal employers, the 'Russian' Sabbatian Mafia. Israel-centric media mogul Rupert Murdoch was introduced to Trump by Cohn and they started a long friendship. Cohn died in 1986 weeks after being disbarred for unethical conduct by the Appellate Division of the New York State Supreme Court. The wheels of justice do indeed run slow given the length of Cohn's crooked career.

QAnon-sense

We are asked to believe that Donald Trump with his fundamental connections to Sabbatian networks and operatives has been leading the fight to stop the Sabbatian agenda for the fascistic control of America and the world. Sure he has. A man entrapped during his years in the White House by Sabbatian operatives and whose biggest financial donor was casino billionaire Sheldon Adelson who was Sabbatian to his DNA?? Oh, do come on. Trump has been used to divide America and isolate Pushbackers on the Cult agenda under the heading of 'Trump supporters', 'insurrectionists' and 'white supremacists'. The US Intelligence/Mossad Psyop or psychological operation known as QAnon emerged during the Trump years as a central pillar in the Sabbatian campaign to lead Pushbackers into the trap set by those that wished to destroy them. I knew from the start that QAnon was a scam because I had seen the same scenario many times before over 30 years under different names and I had written about one in particular in the books. ‘Not again’ was my reaction when QAnon came to the fore. The same script is pulled out every few years and a new name added to the letterhead. The story always takes the same form: ‘Insiders’ or ‘the good guys’ in the government-intelligence-military ‘Deep State’ apparatus were going to instigate mass arrests of the ‘bad guys’ which would include the Rockefellers, Rothschilds, Barack Obama, Hillary Clinton, George Soros, etc., etc. Dates are given for when the ‘good guys’ are going to move in, but the dates pass without incident and new dates are given which pass without incident. The central message to Pushbackers in each case is that they don’t have to do anything because there is ‘a plan’ and it is all going to be sorted by the ‘good guys’ on the inside. ‘Trust the plan’ was a QAnon mantra when the only plan was to misdirect Pushbackers into putting their trust in a Psyop they believed to be real. Beware, beware, those who tell you what you want to hear and always check it out. Right up to Biden’s inauguration QAnon was still claiming that ‘the Storm’ was coming and Trump would stay on as president when Biden and his cronies were arrested and jailed. It was never going to happen and of course it didn’t, but what did happen as a result provided that punchline to the Sabbatian Trump/QAnon Psyop.

On January 6th, 2021, a very big crowd of Trump supporters gathered in the National Mall in Washington DC down from the Capitol Building to protest at what they believed to be widespread corruption and vote fraud that stopped Trump being re-elected for a second term as president in November, 2020. I say as someone that does not support Trump or Biden that the evidence is clear that major vote-fixing went on to favour Biden, a man with cognitive problems so advanced he can often hardly string a sentence together without reading the words written for him on the Teleprompter. Glaring ballot discrepancies included serious questions about electronic voting machines that make vote rigging a comparative cinch and hundreds of thousands of paper votes that suddenly appeared during already advanced vote counts and virtually all of them for Biden. Early Trump leads in crucial swing states suddenly began to close and disappear. The pandemic hoax was used as the excuse to issue almost limitless numbers of mail-in ballots with no checks to establish that the recipients were still alive or lived at that address. They were sent to streams of people who had not even asked for them. Private organisations were employed to gather these ballots and who knows what they did with them before they turned up at the counts. The American election system has been manipulated over decades to become a sick joke with more holes than a Swiss cheese for the express purpose of dictating the results. Then there was the criminal manipulation of information by Sabbatian tech giants like Facebook, Twitter and Google-owned YouTube which deleted pro-Trump, anti-Biden accounts and posts while everything in support of Biden was left alone. Sabbatians wanted Biden to win because after the dividing of America it was time for full-on Woke and every aspect of the Cult agenda to be unleashed.

Hunter gatherer

Extreme Silicon Valley bias included blocking information by the *New York Post* exposing a Biden scandal that should have ended his bid for president in the final weeks of the campaign. Hunter Biden, his monumentally corrupt son, is reported to have sent a laptop to be repaired at a local store and failed to return for it. Time passed until the laptop became the property of the store for non-payment of the bill. When the owner saw what was on the hard drive he gave a copy to the FBI who did nothing even though it confirmed widespread corruption in which the Joe Biden family were using his political position, especially when he was vice president to Obama, to make multiple millions in countries around the world and most notably Ukraine and China. Hunter Biden's one-time business partner Tony Bobulinski went public when the story broke in the *New York Post* to confirm the corruption he saw and that Joe Biden not only knew what was going on he also profited from the spoils. Millions were handed over by a Chinese company with close connections – like all major businesses in China – to the Chinese communist party of President Xi Jinping. Joe Biden even boasted at a meeting of the Cult's World Economic Forum that as vice president he had ordered the government of Ukraine to fire a prosecutor. What he didn't mention was that the same man just happened to be investigating an energy company which was part of Hunter Biden's corrupt portfolio. The company was paying him big bucks for no other reason than the influence his father had. Overnight Biden's presidential campaign should have been over given that he had lied publicly about not knowing what his son was doing. Instead almost the entire Sabbatian-owned mainstream media and Sabbatian-owned Silicon Valley suppressed circulation of the story. This alone went a mighty way to rigging the election of 2020. Cult assets like Mark Zuckerberg at Facebook also spent hundreds of millions to be used in support of Biden and vote 'administration'.

The Cult had used Trump as the focus to divide America and was now desperate to bring in moronic, pliable, corrupt Biden to complete the double-whammy. No way were they going to let little things like the will of the people thwart their plan. Silicon Valley widely censored claims that the election was rigged because it *was* rigged. For the same reason anyone claiming it was rigged was denounced as a 'white supremacist' including the pathetically few Republican politicians willing to say so. Right across the media where the claim was mentioned it was described as a 'false claim' even though these excuses for 'journalists' would have done no research into the subject whatsoever. Trump won seven million more votes than any sitting president had ever achieved while somehow a cognitively-challenged soon to be 78-year-old who was hidden away from the public for most of the campaign managed to win more votes than any presidential candidate in history. It makes no sense. You only had to see election rallies for both candidates to witness the enthusiasm for Trump and the apathy for Biden. Tens of thousands would attend Trump events while Biden was speaking in empty car parks with often only television crews attending and framing their shots to hide the fact that no one was there. It was pathetic to see footage come to light of Biden standing at a podium making speeches only to TV crews and party fixers while reading the words written for him on massive Teleprompter screens. So, yes, those protestors on January 6th had a point about election rigging, but some were about to walk into a trap laid for them in Washington by the Cult Deep State and its QAnon Psyop. This was the Capitol Hill riot ludicrously dubbed an 'insurrection'.

The spider and the fly

Renegade Minds know there are not two 'sides' in politics, only one side, the Cult, working through all 'sides'. It's a stage show, a puppet show, to direct the perceptions of the population into focusing on diversions like parties and candidates while missing the puppeteers with their hands holding all the strings. The Capitol Hill 'insurrection' brings us back to the Little Big Horn. Having created two distinct opposing groupings – Woke and Pushbackers – the trap was about to be sprung. Pushbackers were to be encircled and isolated by associating them all in the public mind with Trump and then labelling Trump as some sort of Confederate leader. I knew immediately that the Capitol riot was a set-up because of two things. One was how easy the rioters got into the building with virtually no credible resistance and secondly I could see – as with the 'Covid' hoax in the West at the start of 2020 – how the Cult could exploit the situation to move its agenda forward with great speed. My experience of Cult techniques and activities over more than 30 years has showed me that while they do exploit situations they haven't themselves created this never happens with events of fundamental agenda significance. Every time major events giving cultists the excuse to rapidly advance their plan you find they are manipulated into being for the specific reason of providing that excuse – Problem-Reaction-Solution. Only a tiny minority of the huge crowd of Washington protestors sought to gain entry to the Capitol by smashing windows and breaching doors. That didn't matter. The whole crowd and all Pushbackers, even if they did not support Trump, were going to be lumped together as dangerous insurrectionists and conspiracy theorists. The latter term came into widespread use through a CIA memo in the 1960s aimed at discrediting those questioning the nonsensical official story of the Kennedy assassination and it subsequently became widely employed by the media. It's still being used by inept 'journalists' with no idea of its origin to discredit anyone questioning anything that authority claims to be true. When you are perpetrating a conspiracy you need to discredit the very word itself even though the dictionary definition of conspiracy is merely 'the activity of secretly planning with other people to do something bad or illegal' and 'a general agreement to keep silent about a subject for the purpose of keeping it secret'. On that basis there are conspiracies almost wherever you look. For obvious reasons the Cult and its lapdog media have to claim there are no conspiracies even though the word appears in state laws as with conspiracy to defraud, to murder, and to corrupt public morals.

Agent provocateurs are widely used by the Cult Deep State to manipulate genuine people into acting in ways that suit the desired outcome. By genuine in this case I mean protestors genuinely supporting Trump and claims that the election was stolen. In among them, however, were agents of the state wearing the garb of Trump supporters and QAnon to pump-prime the Capital riot which some genuine Trump supporters naively fell for. I described the situation as 'Come into my parlour said the spider to the fly'. Leaflets appeared through the Woke paramilitary arm Antifa, the anti-fascist fascists, calling on supporters to turn up in Washington looking like Trump supporters even though they hated him. Some of those arrested for breaching the Capitol Building were sourced to Antifa and its stable mate Black Lives Matter. Both organisations are funded by Cult billionaires and corporations. One man charged for the riot was according to his lawyer a former FBI agent who had held top secret security clearance for 40 years. Attorney Thomas Plofchan said of his client, 66-year-old Thomas Edward Caldwell:

> He has held a Top Secret Security Clearance since 1979 and has undergone multiple Special Background Investigations in support of his clearances. After retiring from the Navy, he

> worked as a section chief for the Federal Bureau of Investigation from 2009-2010 as a GS-12 [mid-level employee].
>
> He also formed and operated a consulting firm performing work, often classified, for U.S government customers including the US. Drug Enforcement Agency, Department of Housing and Urban Development, the US Coast Guard, and the US Army Personnel Command.

A judge later released Caldwell pending trial in the absence of evidence about a conspiracy or that he tried to force his way into the building. *The New York Post* reported a 'law enforcement source' as saying that 'at least two known Antifa members were spotted' on camera among Trump supporters during the riot while one of the rioters arrested was John Earle Sullivan, a seriously extreme Black Lives Matter Trump-hater from Utah who was previously arrested and charged in July, 2020, over a BLM-Antifa riot in which drivers were threatened and one was shot. Sullivan is the founder of Utah-based Insurgence USA which is an affiliate of the Cult-created-and-funded Black Lives Matter movement. Footage appeared and was then deleted by Twitter of Trump supporters calling out Antifa infiltrators and a group was filmed changing into pro-Trump clothing before the riot. Security at the building was *pathetic* – as planned. Colonel Leroy Fletcher Prouty, a man with long experience in covert operations working with the US security apparatus, once described the tell-tale sign to identify who is involved in an assassination. He said:

> No one has to direct an assassination – it happens. The active role is played secretly by permitting it to happen. This is the greatest single clue. Who has the power to call off or reduce the usual security precautions?

This principle applies to many other situations and certainly to the Capitol riot of January 6th, 2021.

The sting

With such a big and potentially angry crowd known to be gathering near the Capitol the security apparatus would have had a major police detail to defend the building with National Guard troops on standby given the strength of feeling among people arriving from all over America encouraged by the QAnon Psyop and statements by Donald Trump. Instead Capitol Police 'security' was flimsy, weak, and easily breached. The same number of officers was deployed as on a regular day and that is a blatant red flag. They were not staffed or equipped for a possible riot that had been an obvious possibility in the circumstances. No protective and effective fencing worth the name was put in place and there were no contingency plans. The whole thing was basically a case of standing aside and waving people in. Once inside police mostly backed off apart from one Capitol police officer who ridiculously shot dead unarmed Air Force veteran protestor Ashli Babbitt without a warning as she climbed through a broken window. The 'investigation' refused to name or charge the officer after what must surely be considered a murder in the circumstances. They just lifted a carpet and swept. The story was endlessly repeated about five people dying in the 'armed insurrection' when there was no report of rioters using weapons. Apart from Babbitt the other four died from a heart attack, strokes and apparently a drug overdose. Capitol police officer Brian Sicknick was reported to have died after being bludgeoned with a fire extinguisher when he was alive after the riot was over and died later of what the Washington Medical Examiner's Office said was a stroke. Sicknick had no external injuries. The lies were delivered like rapid fire. There was a narrative to build with incessant repetition of the lie until the lie became the accepted 'everybody knows that' truth. The 'Big Lie' technique of Nazi Propaganda Minister Joseph Goebbels is constantly used by the Cult which was behind the Nazis and is today behind the 'Covid' and 'climate change' hoaxes. Goebbels said:

> If you tell a lie big enough and keep repeating it, people will eventually come to believe it. The lie can be maintained only for such time as the State can shield the people from the political, economic and/or military consequences of the lie. It thus becomes vitally important for the State to use all of its powers to repress dissent, for the truth is the mortal enemy of the lie, and thus by extension, the truth is the greatest enemy of the State.

Most protestors had a free run of the Capitol Building. This allowed pictures to be taken of rioters in iconic parts of the building including the Senate chamber which could be used as propaganda images against all Pushbackers. One Congresswoman described the scene as ‘the worst kind of non-security anybody could ever imagine’. Well, the first part was true, but someone obviously did imagine it and made sure it happened. Some photographs most widely circulated featured people wearing QAnon symbols and now the Psyop would be used to dub all QAnon followers with the ubiquitous fit-all label of ‘white supremacist’ and ‘insurrectionists’. When a Muslim extremist called Noah Green drove his car at two police officers at the Capitol Building killing one in April, 2021, there was no such political and media hysteria. They were just disappointed he wasn’t white.

The witch-hunt

Government prosecutor Michael Sherwin, an aggressive, dark-eyed, professional Rottweiler led the ‘investigation’ and to call it over the top would be to understate reality a thousand fold. Hundreds were tracked down and arrested for the crime of having the wrong political views and people were jailed who had done nothing more than walk in the building, committed no violence or damage to property, took a few pictures and left. They were labelled a ‘threat to the Republic’ while Biden sat in the White House signing executive orders written for him that were dismantling ‘the Republic’. Even when judges ruled that a mother and son should not be in jail the government kept them there. Some of those arrested have been badly beaten by prison guards in Washington and lawyers for one man said he suffered a fractured skull and was made blind in one eye. Meanwhile a woman is shot dead for no reason by a Capitol Police officer and we are not allowed to know who he is never mind what has happened to him although that will be *nothing*. The Cult’s QAnon/Trump sting to identify and isolate Pushbackers and then target them on the road to crushing and deleting them was a resounding success. You would have thought the Russians had invaded the building at gunpoint and lined up senators for a firing squad to see the political and media reaction. Congresswoman Alexandria Ocasio-Cortez is a child in a woman's body, a terrible-twos, me, me, me, Woker narcissist of such proportions that words have no meaning. She said she thought she was going to die when 'insurrectionists' banged on her office door. It turned out she wasn't even in the Capitol Building when the riot was happening and the 'banging' was a Capitol Police officer. She referred to herself as a 'survivor' which is an insult to all those true survivors of violent and sexual abuse while she lives her pampered and privileged life talking drivel for a living. Her Woke colleague and fellow mega-narcissist Rashida Tlaib broke down describing the devastating effect on her, too, of *not being* in the building when the rioters were there. Ocasio-Cortez and Tlaib are members of a fully-Woke group of Congresswomen known as 'The Squad' along with Ilhan Omar and Ayanna Pressley. The Squad from what I can see can be identified by its vehement anti-white racism, anti-white men agenda, and, as always in these cases, the absence of brain cells on active duty.

The usual suspects were on the riot case immediately in the form of Democrat ultra-Zionist senators and operatives Chuck Schumer and Adam Schiff demanding that Trump be impeached for 'his part in the insurrection'. The same pair of prats had led the failed impeachment of Trump over the invented 'Russia collusion' nonsense which claimed Russia had helped Trump win the 2016 election. I didn't realise that Tel Aviv had been relocated just outside Moscow. I must find an up-to-date map. The Russia hoax was a Sabbatian operation to keep Trump occupied and impotent and to stop any rapport with Russia which the Cult wants to retain as a perceptual enemy to be pulled out at will. Puppet Biden began attacking Russia when he came to office as the Cult seeks more upheaval, division and war across the world. A two-year stage show 'Russia collusion inquiry' headed by the not-very-bright former 9/11 FBI chief Robert Mueller, with support from 19 lawyers, 40 FBI agents plus intelligence analysts, forensic accountants and other staff, devoured tens of millions of dollars and found no evidence of Russia collusion which a ten-year-old could have told them on day one. Now the same moronic Schumer and Schiff wanted a second impeachment of Trump over the Capitol 'insurrection' (riot) which the arrested development of Schumer called another 'Pearl Harbor' while others compared it with 9/11 in which 3,000 died and, in the case of CNN, with the Rwandan genocide in the 1990s in which an estimated 500,000 to 600,000 were murdered, between 250, 000 and 500,000 women were raped, and populations of whole towns were hacked to death with machetes. To make those comparisons purely for Cult political reasons is beyond insulting to those that suffered and lost their lives and confirms yet again the callous inhumanity that we are dealing with. Schumer is a monumental idiot and so is Schiff, but they serve the Cult agenda and do whatever they're told so they get looked after. Talking of idiots – another inane man who spanned the Russia and Capitol impeachment attempts was Senator Eric Swalwell who had the nerve to accuse Trump of collusion with the Russians while sleeping with a Chinese spy called Christine Fang or 'Fang Fang' which is straight out of a Bond film no doubt starring Klaus Schwab as the bloke living on a secret island and controlling laser weapons positioned in space and pointing at world capitals. Fang Fang plays the part of Bond's infiltrator girlfriend which I'm sure she would enjoy rather more than sharing a bed with the brainless Swalwell, lying back and thinking of China. The FBI eventually warned Swalwell about Fang Fang which gave her time to escape back to the Chinese dictatorship. How very thoughtful of them. The second Trump impeachment also failed and hardly surprising when an impeachment is supposed to remove a sitting president and by the time it happened Trump was no longer president. These people are running your country America, well, officially anyway. Terrifying isn't it?

Outcomes tell the story - always

The outcome of all this – and it's the *outcome* on which Renegade Minds focus, not the words – was that a vicious, hysterical and obviously pre-planned assault was launched on Pushbackers to censor, silence and discredit them and even targeted their right to earn a living. They have since been condemned as ‘domestic terrorists’ that need to be treated like Al-Qaeda and Islamic State. ‘Domestic terrorists’ is a label the Cult has been trying to make stick since the period of the Oklahoma bombing in 1995 which was blamed on ‘far-right domestic terrorists’. If you read *The Trigger* you will see that the bombing was clearly a Problem-Reaction-Solution carried out by the Deep State during a Bill Clinton administration so corrupt that no dictionary definition of the term would even nearly suffice. Nearly 30, 000 troops were deployed from all over America to the empty streets of Washington for Biden’s inauguration. Ten thousand of them stayed on with the pretext of protecting the capital from insurrectionists when it was more psychological programming to normalise the use of the military in domestic law enforcement in support of the Cult plan for a police-military state. Biden’s fascist administration began a purge of ‘wrong-thinkers’ in the military which means anyone that is not on board with Woke. The Capitol Building was surrounded by a fence with razor wire and the Land of the Free was further symbolically and literally dismantled. The circle was completed with the installation of Biden and the exploitation of the QAnon Psyop.

America had never been so divided since the civil war of the 19th century, Pushbackers were isolated and dubbed terrorists and now, as was always going to happen, the Cult immediately set about deleting what little was left of freedom and transforming American society through a swish of the hand of the most controlled ‘president’ in American history leading (officially at least) the most extreme regime since the country was declared an independent state on July 4th, 1776. Biden issued undebated, dictatorial executive orders almost by the hour in his opening days in office across the whole spectrum of the Cult wish-list including diluting controls on the border with Mexico allowing thousands of migrants to illegally enter the United States to transform the demographics of America and import an election-changing number of perceived Democrat voters. Then there were Biden deportation amnesties for the already illegally resident (estimated to be as high as 20 or even 30 million). A bill before Congress awarded American citizenship to anyone who could prove they had worked in agriculture for just 180 days in the previous two years as ‘Big Ag’ secured its slave labour long-term. There were the plans to add new states to the union such as Puerto Rico and making Washington DC a state. They are all parts of a plan to ensure that the Cult-owned Woke Democrats would be permanently in power.

Border – what border?

I have exposed in detail in other books how mass immigration into the United States and Europe is the work of Cult networks fuelled by the tens of billions spent to this and other ends by George Soros and his global Open Society (open borders) Foundations. The impact can be seen in America alone where the population has increased by *100 million* in little more than 30 years mostly through immigration. I wrote in *The Answer* that the plan was to have so many people crossing the southern border that the numbers become unstoppable and we are now there under Cult-owned Biden. El Salvador in Central America puts the scale of what is happening into context. A third of the population now lives in the United States, much of it illegally, and many more are on the way. The methodology is to crush Central and South American countries economically and spread violence through machete-wielding psychopathic gangs like MS-13 based in El Salvador and now operating in many American cities. Biden-imposed lax security at the southern border means that it is all but open. He said before his ‘election’ that he wanted to see a surge towards the border if he became president and that was the green light for people to do just that after election day to create the human disaster that followed for both America and the migrants. When that surge came the imbecilic Alexandria Ocasio-Cortez said it wasn’t a ‘surge’ because they are ‘children, not insurgents’ and the term ‘surge’ (used by Biden) was a claim of ‘white supremacists’.

This disingenuous lady may one day enter the realm of the most basic intelligence, but it won't be any time soon.

Sabbatians and the Cult are in the process of destroying America by importing violent people and gangs in among the genuine to terrorise American cities and by overwhelming services that cannot cope with the sheer volume of new arrivals. Something similar is happening in Europe as Western society in general is targeted for demographic and cultural transformation and upheaval. The plan demands violence and crime to create an environment of intimidation, fear and division and Soros has been funding the election of district attorneys across America who then stop prosecuting many crimes, reduce sentences for violent crimes and free as many violent criminals as they can. Sabbatians are creating the chaos from which order – their order – can respond in a classic Problem-Reaction-Solution. A Freemasonic moto says 'Ordo Ab Chao' (Order out of Chaos) and this is why the Cult is constantly creating chaos to impose a new 'order'. Here you have the reason the Cult is constantly creating chaos. The 'Covid' hoax can be seen with those entering the United States by plane being forced to take a 'Covid' test while migrants flooding through southern border processing facilities do not. Nothing is put in the way of mass migration and if that means ignoring the government's own 'Covid' rules then so be it. They know it's all bullshit anyway. Any pushback on this is denounced as 'racist' by Wokers and Sabbatian fronts like the ultra-Zionist Anti-Defamation League headed by the appalling Jonathan Greenblatt which at the same time argues that Israel should not give citizenship and voting rights to more Palestinian Arabs or the 'Jewish population' (in truth the Sabbatian network) will lose control of the country.

Society-changing numbers

Biden's masters have declared that countries like El Salvador are so dangerous that their people must be allowed into the United States for humanitarian reasons when there are fewer murders in large parts of many Central American countries than in US cities like

Baltimore. That is not to say Central America cannot be a dangerous place and Cult-controlled American governments have been making it so since way back, along with the dismantling of economies, in a long-term plan to drive people north into the United States. Parts of Central America are very dangerous, but in other areas the story is being greatly exaggerated to justify relaxing immigration criteria. Migrants are being offered free healthcare and education in the United States as another incentive to head for the border and there is no requirement to be financially independent before you can enter to prevent the resources of America being drained. You can't blame migrants for seeking what they believe will be a better life, but they are being played by the Cult for dark and nefarious ends. The numbers since Biden took office are huge. In February, 2021, more than 100,000 people were known to have tried to enter the US illegally through the southern border (it was 34,000 in the same month in 2020) and in March it was 170,000 – a 418 percent increase on March, 2020. These numbers are only known people, not the ones who get in unseen. The true figure for migrants illegally crossing the border in a single month was estimated by one congressman at 250,000 and that number will only rise under Biden's current policy. Gangs of murdering drug-running thugs that control the Mexican side of the border demand money – thousands of dollars – to let migrants cross the Rio Grande into America. At the same time gun battles are breaking out on the border several times a week between rival Mexican drug gangs (which now operate globally) who are equipped with sophisticated military-grade weapons, grenades and armoured vehicles. While the Capitol Building was being 'protected' from a non-existent 'threat' by thousands of troops, and others were still deployed at the time in the Cult Neocon war in Afghanistan, the southern border of America was left to its fate. This is not incompetence, it is cold calculation.

By March, 2021, there were 17,000 unaccompanied children held at border facilities and many of them are ensnared by people traffickers for paedophile rings and raped on their journey north to America. This is not conjecture – this is fact. Many of those designated children are in reality teenage boys or older. Meanwhile Wokers posture their self-purity for encouraging poor and tragic people to come to America and face this nightmare both on the journey and at the border with the disgusting figure of House Speaker Nancy Pelosi giving disingenuous speeches about caring for migrants. The woman's evil. Wokers condemned Trump for having children in cages at the border (so did Obama, *Shhhh*), but now they are sleeping on the floor without access to a shower with one border facility 729 percent over capacity. The Biden insanity even proposed flying migrants from the southern border to the northern border with Canada for 'processing'. The whole shambles is being overseen by ultra-Zionist Secretary of Homeland Security, the moronic liar Alejandro Mayorkas, who banned news cameras at border facilities to stop Americans seeing what was happening. Mayorkas said there was not a ban on news crews; it was just that they were not allowed to film. Alongside him at Homeland Security is another ultra-Zionist Cass Sunstein appointed by Biden to oversee new immigration laws. Sunstein despises conspiracy researchers to the point where he suggests they should be banned or *taxed* for having such views. The man is not bonkers or anything. He's perfectly well-adjusted, but adjusted to what is the question. Criticise what is happening and you are a 'white supremacist' when earlier non-white immigrants also oppose the numbers which effect their lives and opportunities. Black people in poor areas are particularly damaged by uncontrolled immigration and the increased competition for work opportunities with those who will work for less. They are also losing voting power as Hispanics become more dominant in former black areas. It's a downward spiral for them while the billionaires behind the policy drone on about how much they care about black people and 'racism'. None of this is about compassion for migrants or black people – that's just wind and air. Migrants are instead being mercilessly exploited to transform America while the countries they leave are losing their future and the same is true in Europe. Mass immigration may now be the work of Woke Democrats, but it can be traced back to the 1986 Immigration Reform and Control Act (it wasn't) signed into law by Republican hero President Ronald Reagan which gave amnesty to millions living in the United States illegally and other incentives for people to head for the southern border. Here we have the one-party state at work again.

Save me syndrome

Almost every aspect of what I have been exposing as the Cult agenda was on display in even the first days of 'Biden' with silencing of Pushbackers at the forefront of everything. A Renegade Mind will view the Trump years and QAnon in a very different light to their supporters and advocates as the dots are connected. The QAnon/Trump Psyop has given the Cult all it was looking for. We may not know how much, or little, that Trump realised he was being used, but that's a side issue. This pincer movement produced the desired outcome of dividing America and having Pushbackers isolated. To turn this around we have to look at new routes to empowerment which do not include handing our power to other people and groups through what I will call the 'Save Me Syndrome' – 'I want someone else to do it so that I don't have to'. We have seen this at work throughout human history and the QAnon/Trump Psyop is only the latest incarnation alongside all the others. Religion is an obvious expression of this when people look to a 'god' or priest to save them or tell them how to be saved and then there are 'save me' politicians like Trump. Politics is a diversion and not a 'saviour'. It is a means to block positive change, not make it possible.

Save Me Syndrome always comes with the same repeating theme of handing your power to whom or what you believe will save you while your real 'saviour' stares back from the mirror every morning. Renegade Minds are constantly vigilant in this regard and always asking the question 'What can I do?' rather than 'What can someone else do for me?' Gandhi was right when he said: 'You must be the change you want to see in the world.' We are indeed the people we have been waiting for. We are presented with a constant raft of reasons to concede that power to others and forget where the real power is. Humanity has the numbers and the Cult does not. It has to use diversion and division to target the unstoppable power that comes from unity. Religions, governments, politicians, corporations, media, QAnon, are all different manifestations of this power-diversion and dilution. Refusing to give your power to governments and instead handing it to Trump and QAnon is not to take a new direction, but merely to recycle the old one with new names on the posters. I will explore this phenomenon as we proceed and how to break the cycles and recycles that got us here through the mists of repeating perception and so repeating history.

For now we shall turn to the most potent example in the entire human story of the consequences that follow when you give your power away. I am talking, of course, of the ‘Covid’ hoax.

CHAPTER FOUR

‘Covid’: Calculated catastrophe

Facts are threatening to those invested in fraud
DaShanne Stokes

We can easily unravel the real reason for the ‘Covid pandemic’ hoax by employing the Renegade Mind methodology that I have outlined this far. We’ll start by comparing the long-planned Cult outcome with the ‘Covid pandemic’ outcome. Know the outcome and you’ll see the journey.

I have highlighted the plan for the Hunger Games Society which has been in my books for so many years with the very few controlling the very many through ongoing dependency. To create this dependency it is essential to destroy independent livelihoods, businesses and employment to make the population reliant on the state (the Cult) for even the basics of life through a guaranteed pittance income. While independence of income remained these Cult ambitions would be thwarted. With this knowledge it was easy to see where the ‘pandemic’ hoax was going once talk of ‘lockdowns’ began and the closing of all but perceived ‘essential’ businesses to ‘save’ us from an alleged ‘deadly virus’. Cult corporations like Amazon and Walmart were naturally considered ‘essential’ while mom and pop shops and stores had their doors closed by fascist decree. As a result with every new lockdown and new regulation more small and medium, even large businesses not owned by the Cult, went to the wall while Cult giants and their frontmen and women grew financially fatter by the second. Mom and pop were denied an income and the right to earn a living and the wealth of people like Jeff Bezos (Amazon), Mark Zuckerberg (Facebook) and Sergei Brin and Larry Page (Google/Alphabet) have reached record levels. The Cult was increasing its own power through further dramatic concentrations of wealth while the competition was being destroyed and brought into a state of dependency. Lockdowns have been instigated to secure that very end and were never anything to do with health. My brother Paul spent 45 years building up a bus repair business, but lockdowns meant buses were running at a fraction of normal levels for months on end. Similar stories can told in their hundreds of millions worldwide. Efforts of a lifetime coldly destroyed by Cult multi-billionaires and their lackeys in government and law enforcement who continued to earn their living from the taxation of the people while denying the right of the same people to earn theirs. How different it would have been if those making and enforcing these decisions had to face the same financial hardships of those they affected, but they never do.

Gates of Hell

Behind it all in the full knowledge of what he is doing and why is the psychopathic figure of Cult operative Bill Gates. His puppet Tedros at the World Health Organization declared 'Covid' a pandemic in March, 2020. The WHO had changed the definition of a 'pandemic' in 2009 just a month before declaring the 'swine flu pandemic' which would not have been so under the previous definition. The same applies to 'Covid'. The definition had included… 'an infection by an infectious agent, occurring simultaneously in different countries, with a significant mortality rate relative to the proportion of the population infected'. The new definition removed the need for 'significant mortality'. The 'pandemic' has been fraudulent even down to the definition, but Gates demanded economy-destroying lockdowns, school closures, social distancing, mandatory masks, a 'vaccination' for every man, woman and child on the planet and severe consequences and restrictions for those that refused. Who gave him this power? The

Cult did which he serves like a little boy in short trousers doing what his daddy tells him. He and his psychopathic missus even smiled when they said that much worse was to come (what they knew was planned to come). Gates responded in the matter-of-fact way of all psychopaths to a question about the effect on the world economy of what he was doing:

> Well, it won't go to zero but it will shrink. Global GDP is probably going to take the biggest hit ever [Gates was smiling as he said this] … in my lifetime this will be the greatest economic hit. But you don't have a choice. People act as if you have a choice. People don't feel like going to the stadium when they might get infected … People are deeply affected by seeing these stats, by knowing they could be part of the transmission chain, old people, their parents and grandparents, could be affected by this, and so you don't get to say ignore what is going on here.
>
> There will be the ability to open up, particularly in rich countries, if things are done well over the next few months, but for the world at large normalcy only returns when we have largely vaccinated the entire population.

The man has no compassion or empathy. How could he when he's a psychopath like all Cult players? My own view is that even beyond that he is very seriously mentally ill. Look in his eyes and you can see this along with his crazy flailing arms. You don't do what he has done to the world population since the start of 2020 unless you are mentally ill and at the most extreme end of psychopathic. You especially don't do it when to you know, as we shall see, that cases and deaths from 'Covid' are fakery and a product of monumental figure massaging. 'These stats' that Gates referred to are based on a 'test' that's not testing for the 'virus' as he has known all along. He made his fortune with big Cult support as an infamously ruthless software salesman and now buys global control of 'health' (death) policy without the population he affects having any say. It's a breathtaking outrage. Gates talked about people being deeply affected by fear of 'Covid' when that was because of *him* and his global network lying to them minute-by-minute supported by a lying media that he seriously influences and funds to the tune of hundreds of millions. He's handed big sums to media operations including the BBC, NBC, Al Jazeera, Univision, *PBS NewsHour,* *ProPublica, National Journal, The Guardian, The Financial Times, The Atlantic, Texas Tribune, USA Today* publisher Gannett, *Washington Monthly, Le Monde,* Center for Investigative Reporting, Pulitzer Center on Crisis Reporting, National Press Foundation, International Center for Journalists, Solutions Journalism Network, the Poynter Institute for Media Studies, and many more. Gates is everywhere in the 'Covid' hoax and the man must go to prison – or a mental facility – for the rest of his life and his money distributed to those he has taken such enormous psychopathic pleasure in crushing.

The Muscle

The Hunger Games global structure demands a police-military state – a fusion of the two into one force – which viciously imposes the will of the Cult on the population and protects the Cult from public rebellion. In that regard, too, the 'Covid' hoax just keeps on giving. Often unlawful, ridiculous and contradictory 'Covid' rules and regulations have been policed across the world by moronic automatons and psychopaths made faceless by face-nappy masks and acting like the Nazi SS and fascist blackshirts and brownshirts of Hitler and Mussolini. The smallest departure from the rules decreed by the psychos in government and their clueless gofers were jumped upon by the face-nappy fascists. Brutality against public protestors soon became commonplace even on girls, women and old people as the brave men with the batons – the Face-Nappies as I call them – broke up peaceful protests and handed out fines like confetti to people who couldn't earn a living let alone pay hundreds of pounds for what was once an accepted human right. Robot Face-Nappies of Nottingham police in the English East Midlands fined one group £11,000 for attending a child's birthday party. For decades I charted the transformation of law enforcement as genuine, decent officers were replaced with psychopaths and the brain dead who would happily and brutally do whatever their masters told them. Now they were let loose on the public and I would emphasise the point that none of this just happened. The step-by-step change in the dynamic between police and public was orchestrated from the shadows by those who knew where this was all going and the same with the perceptual reframing of those in all levels of authority and official administration through 'training courses' by organisations such as Common Purpose which was created in the late 1980s and given a massive boost in Blair era Britain until it became a global phenomenon. Supposed public 'servants' began to view the population as the enemy and the same was true of the police. This was the start of the explosion of behaviour manipulation organisations and networks preparing for the all-war on the human psyche unleashed with the dawn of 2020. I will go into more detail about this later in the book because it is a core part of what is happening.

Police desecrated beauty spots to deter people gathering and arrested women for walking in the countryside alone 'too far' from their homes. We had arrogant, clueless sergeants in the Isle of Wight police where I live posting on Facebook what they insisted the population must do or else. A schoolmaster sergeant called Radford looked young enough for me to ask if his mother knew he was out, but he was posting what he *expected* people to do while a Sergeant Wilkinson boasted about fining lads for meeting in a McDonald's car park where they went to get a lockdown takeaway. Wilkinson added that he had even cancelled their order. What a pair of prats these people are and yet they have increasingly become the norm among Jackboot Johnson's Yellowshirts once known as the British police. This was the theme all over the world with police savagery common during lockdown protests in the United States, the Netherlands, and the fascist state of Victoria in Australia under its tyrannical and again moronic premier Daniel Andrews. Amazing how tyrannical and moronic tend to work as a team and the same combination could be seen across America as arrogant, narcissistic Woke governors and mayors such as Gavin Newsom (California), Andrew Cuomo (New York), Gretchen Whitmer (Michigan), Lori Lightfoot (Chicago) and Eric Garcetti (Los Angeles) did their Nazi and Stalin impressions with the full support of the compliant brutality of their enforcers in uniform as they arrested small business owners defying fascist shutdown orders and took them to jail in ankle shackles and handcuffs. This happened to bistro owner Marlena Pavlos-Hackney in Gretchen Whitmer's fascist state of Michigan when police arrived to enforce an order by a state-owned judge for 'putting the community at risk' at a time when other states like Texas were dropping restrictions and migrants were pouring across the southern border without any 'Covid' questions at all. I'm sure there are many officers appalled by what they are ordered to do, but not nearly enough of them. If they were truly appalled they would not do it. As the months passed every opportunity was taken to have the military involved to make their presence on the streets ever more familiar and 'normal' for the longer-term goal of police-military fusion.

Another crucial element to the Hunger Games enforcement network has been encouraging the public to report neighbours and others for 'breaking the lockdown rules'. The group faced with £11,000 in fines at the child's birthday party would have been dobbed-in by a neighbour with a brain the size of a pea. The technique was most famously employed by the Stasi secret police in communist East Germany who had public informants placed throughout the population. A police chief in the UK says his force doesn't need to carry out 'Covid' patrols when they are flooded with so many calls from the public reporting other people for visiting the beach. Dorset police chief James Vaughan said people were so enthusiastic about snitching on their fellow humans they were now operating as an auxiliary arm of the police: 'We are still getting around 400 reports a week from the public, so we will respond to reports …We won't need to be doing hotspot patrols because people are very quick to pick the phone up and tell us.' Vaughan didn't say that this is a pillar of all tyrannies of whatever complexion and the means to hugely extend the reach of enforcement while spreading distrust among the people and making them wary of doing anything that might get them reported. Those narcissistic Isle of Wight sergeants Radford and Wilkinson never fail to add a link to their Facebook posts where the public can inform on their fellow slaves.

Neither would be self-aware enough to realise they were imitating the Stasi which they might well never have heard of. Government psychologists that I will expose later laid out a policy to turn communities against each other in the same way.

A coincidence? Yep, and I can knit fog

I knew from the start of the alleged pandemic that this was a Cult operation. It presented limitless potential to rapidly advance the Cult agenda and exploit manipulated fear to demand that every man, woman and child on the planet was 'vaccinated' in a process never used on humans before which infuses self-replicating *synthetic* material into human cells. Remember the plan to transform the human body from a biological to a synthetic biological state. I'll deal with the 'vaccine' (that's not actually a vaccine) when I focus on the genetic agenda. Enough to say here that mass global 'vaccination' justified by this 'new virus' set alarms ringing after 30 years of tracking these people and their methods. The 'Covid' hoax officially beginning in China was also a big red flag for reasons I will be explaining. The agenda potential was so enormous that I could dismiss any idea that the 'virus' appeared naturally. Major happenings with major agenda implications never occur without Cult involvement in making them happen. My questions were twofold in early 2020 as the media began its campaign to induce global fear and hysteria: Was this alleged infectious agent released on purpose by the Cult or did it even exist at all? I then did what I always do in these situations. I sat, observed and waited to see where the evidence and information would take me. By March and early April synchronicity was strongly – and ever more so since then – pointing me in the direction of *there is no 'virus'*. I went public on that with derision even from swathes of the alternative media that voiced a scenario that the Chinese government released the 'virus' in league with Deep State elements in the United States from a top-level bio-lab in Wuhan where the 'virus' is said to have first appeared. I looked at that possibility, but I didn't buy it for several reasons. Deaths from the 'virus' did not in any way match what they would have been with a ‘deadly bioweapon’ and it is much more effective if you sell the *illusion* of an infectious agent rather than having a real one unless you can control through injection who has it and who doesn’t. Otherwise you lose control of events. A made-up ‘virus’ gives you a blank sheet of paper on which you can make it do whatever you like and have any symptoms or mutant ‘variants’ you choose to add while a real infectious agent would limit you to what it actually does. A phantom disease allows you to have endless ludicrous ‘studies’ on the ‘Covid’ dollar to widen the perceived impact by inventing ever more ‘at risk’ groups including one study which said those who walk slowly may be almost four times more likely to die from the ‘virus’. People are in psychiatric wards for less.

A real ‘deadly bioweapon’ can take out people in the hierarchy that are not part of the Cult, but essential to its operation. Obviously they don’t want that. Releasing a real disease means you immediately lose control of it. Releasing an illusory one means you don’t. Again it’s vital that people are extra careful when dealing with what they want to hear. A bioweapon unleashed from a Chinese laboratory in collusion with the American Deep State may fit a conspiracy narrative, but is it true? Would it not be far more effective to use the excuse of a ‘virus’ to justify the real bioweapon – the ‘vaccine’? That way your disease agent does not have to be transmitted and arrives directly through a syringe. I saw a French virologist Luc Montagnier quoted in the alternative media as saying he had discovered that the alleged ‘new’ severe acute respiratory syndrome coronavirus , or SARS-CoV-2, was made artificially and included elements of the human immunodeficiency ‘virus’ (HIV) and a parasite that causes malaria. SARS-CoV-2 is alleged to trigger an alleged illness called Covid-19. I remembered Montagnier’s name from my research years before into claims that an HIV ‘retrovirus’ causes AIDs – claims that were demolished by Berkeley virologist Peter Duesberg who showed that no one had ever proved that HIV causes acquired immunodeficiency syndrome or AIDS. Claims that become accepted as fact, publicly and medically, with no proof whatsoever are an ever-recurring story that profoundly applies to

'Covid'. Nevertheless, despite the lack of proof, Montagnier's team at the Pasteur Institute in Paris had a long dispute with American researcher Robert Gallo over which of them discovered and isolated the HIV 'virus' and with *no evidence* found it to cause AIDS. You will see later that there is also no evidence that any 'virus' causes any disease or that there is even such a thing as a 'virus' in the way it is said to exist. The claim to have 'isolated' the HIV 'virus' will be presented in its real context as we come to the shocking story – and it is a story – of SARS-CoV-2 and so will Montagnier's assertion that he identified the full SARS-CoV-2 genome.

Hoax in the making

We can pick up the 'Covid' story in 2010 and the publication by the Rockefeller Foundation of a document called 'Scenarios for the Future of Technology and International Development'. The inner circle of the Rockefeller family has been serving the Cult since John D. Rockefeller (1839-1937) made his fortune with Standard Oil. It is less well known that the same Rockefeller – the Bill Gates of his day – was responsible for establishing what is now referred to as 'Big Pharma', the global network of pharmaceutical companies that make outrageous profits dispensing scalpel and drug 'medicine' and are obsessed with pumping vaccines in ever-increasing number into as many human arms and backsides as possible. John D. Rockefeller was the driving force behind the creation of the 'education' system in the United States and elsewhere specifically designed to program the perceptions of generations thereafter. The Rockefeller family donated exceptionally valuable land in New York for the United Nations building and were central in establishing the World Health Organization in 1948 as an agency of the UN which was created from the start as a Trojan horse and stalking horse for world government. Now enter Bill Gates. His family and the Rockefellers have long been extremely close and I have seen genealogy which claims that if you go back far enough the two families fuse into the same bloodline. Gates has said that the Bill and Melinda Gates Foundation was inspired by the Rockefeller Foundation and why not when both are serving the same Cult? Major tax-exempt foundations are overwhelmingly criminal enterprises in which Cult assets fund the Cult agenda in the guise of 'philanthropy' while avoiding tax in the process. Cult operatives can become mega-rich in their role of front men and women for the psychopaths at the inner core and they, too, have to be psychopaths to knowingly serve such evil. Part of the deal is that a big percentage of the wealth gleaned from representing the Cult has to be spent advancing the ambitions of the Cult and hence you have the Rockefeller Foundation, Bill and Melinda Gates Foundation (and *so* many more) and people like George Soros with his global Open Society Foundations spending their billions in pursuit of global Cult control. Gates is a global public face of the Cult with his interventions in world affairs including Big Tech influence; a central role in the 'Covid' and 'vaccine' scam; promotion of the climate change shakedown; manipulation of education; geoengineering of the skies; and his food-control agenda as the biggest owner of farmland in America, his GMO promotion and through other means. As one writer said: 'Gates monopolizes or wields disproportionate influence over the tech industry, global health and vaccines, agriculture and food policy (including biopiracy and fake food), weather modification and other climate technologies, surveillance, education and media.' The almost limitless wealth secured through Microsoft and other not-allowed-to-fail ventures (including vaccines) has been ploughed into a long, long list of Cult projects designed to enslave the entire human race. Gates and the Rockefellers have been working as one unit with the Rockefeller-established World Health Organization leading global 'Covid' policy controlled by Gates through his mouth-piece Tedros. Gates became the WHO's biggest funder when Trump announced that the American government would cease its donations, but Biden immediately said he would restore the money when he took office in January, 2021. The Gates Foundation (the Cult) owns through limitless funding the world health system and the major players across the globe in the 'Covid' hoax.

Okay, with that background we return to that Rockefeller Foundation document of 2010 headed 'Scenarios for the Future of Technology and International Development' and its 'imaginary' epidemic of a virulent and deadly influenza strain which infected 20 percent of the global population and killed eight million in seven months. The Rockefeller scenario was that the epidemic destroyed economies, closed shops, offices and other businesses and led to governments imposing fierce rules and restrictions that included mandatory wearing of face masks and body-temperature checks to enter communal spaces like railway stations and supermarkets. The document predicted that even after the height of the Rockefeller-envisaged epidemic the authoritarian rule would continue to deal with further pandemics, transnational terrorism, environmental crises and rising poverty. Now you may think that the Rockefellers are our modern-day seers or alternatively, and rather more likely, that they well knew what was planned a few years further on. Fascism had to be imposed, you see, to 'protect citizens from risk and exposure'. The Rockefeller scenario document said:

> During the pandemic, national leaders around the world flexed their authority and imposed airtight rules and restrictions, from the mandatory wearing of face masks to body-temperature checks at the entries to communal spaces like train stations and supermarkets. Even after the pandemic faded, this more authoritarian control and oversight of citizens and their activities stuck and even intensified. In order to protect themselves from the spread of increasingly global problems – from pandemics and transnational terrorism to environmental crises and rising poverty – leaders around the world took a firmer grip on power.
>
> At first, the notion of a more controlled world gained wide acceptance and approval. Citizens willingly gave up some of their sovereignty – and their privacy – to more paternalistic states in exchange for greater safety and stability. Citizens were more tolerant, and even eager, for top-down direction and oversight, and national leaders had more latitude to impose order in the ways they saw fit.
>
> In developed countries, this heightened oversight took many forms: biometric IDs for all citizens, for example, and tighter regulation of key industries whose stability was deemed vital to national interests. In many developed countries, enforced cooperation with a suite of new regulations and agreements slowly but steadily restored both order and, importantly, economic growth.

There we have the prophetic Rockefellers in 2010 and three years later came their paper for the Global Health Summit in Beijing, China, when government representatives, the private sector, international organisations and groups met to discuss the next 100 years of 'global health'. The Rockefeller Foundation-funded paper was called 'Dreaming the Future of Health for the Next 100 Years and more prophecy ensued as it described a dystopian future: 'The abundance of data, digitally tracking and linking people may mean the 'death of privacy' and may replace physical interaction with transient, virtual connection, generating isolation and raising questions of how values are shaped in virtual networks.' Next in the 'Covid' hoax preparation sequence came a 'table top' simulation in 2018 for another 'imaginary' pandemic of a disease called Clade X which was said to kill 900 million people. The exercise was organised by the Gates-funded Johns Hopkins University's Center for Health Security in the United States and this is the very same university that has been compiling the disgustingly and systematically erroneous global figures for 'Covid' cases and deaths. Similar Johns Hopkins health crisis scenarios have included the Dark Winter exercise in 2001 and Atlantic Storm in 2005.

Nostradamus 201

For sheer predictive genius look no further prophecy-watchers than the Bill Gates-funded Event 201 held only six weeks before the 'coronavirus pandemic' is supposed to have broken out in China and Event 201 was based on a scenario of a global 'coronavirus pandemic'. Melinda Gates, the great man's missus, told the BBC that he had 'prepared for years' for a coronavirus pandemic which told us what we already knew. Nostradamugates had predicted in a TED talk in 2015 that a pandemic was coming that would kill a lot of people and demolish the world economy. My god, the man is a machine – possibly even literally. Now here he was only weeks before the real thing funding just such a simulated scenario and involving his friends and associates at Johns Hopkins, the World Economic Forum Cult-front of Klaus Schwab, the United Nations, Johnson & Johnson, major banks, and officials from China and the Centers for Disease Control in the United States. What synchronicity – Johns Hopkins would go on to compile the fraudulent 'Covid' figures, the World Economic Forum and Schwab would push the 'Great Reset' in response to 'Covid', the Centers for Disease Control would be at the forefront of 'Covid' policy in the United States, Johnson & Johnson would produce a 'Covid vaccine', and everything would officially start just weeks later in China. Spooky, eh? They were even accurate in creating a simulation of a 'virus' pandemic because the 'real thing' would also be a simulation. Event 201 was not an exercise preparing for something that might happen; it was a rehearsal for what those in control knew was *going* to happen and very shortly. Hours of this simulation were posted on the Internet and the various themes and responses mirrored what would soon be imposed to transform human society. News stories were inserted and what they said would be commonplace a few weeks later with still more prophecy perfection. Much discussion focused on the need to deal with misinformation and the 'anti-vax movement' which is exactly what happened when the 'virus' arrived – was said to have arrived – in the West.

Cult-owned social media banned criticism and exposure of the official 'virus' narrative and when I said there *was* no 'virus' in early April, 2020, I was banned by one platform after another including YouTube, Facebook and later Twitter. The mainstream broadcast media in Britain was in effect banned from interviewing me by the Tony-Blair-created government broadcasting censor Ofcom headed by career government bureaucrat Melanie Dawes who was appointed just as the 'virus' hoax was about to play out in January, 2020. At the same time the Ickonic media platform was using Vimeo, another ultra-Zionist-owned operation, while our own player was being created and they deleted in an instant hundreds of videos, documentaries, series and shows to confirm their unbelievable vindictiveness. We had copies, of course, and they had to be restored one by one when our player was ready. These people have no class. Sabbatian Facebook promised free advertisements for the Gates-controlled World Health Organization narrative while deleting ‘false claims and conspiracy theories’ to stop ‘misinformation’ about the alleged coronavirus. All these responses could be seen just a short while earlier in the scenarios of Event 201. Extreme censorship was absolutely crucial for the Cult because the official story was so ridiculous and unsupportable by the evidence that it could never survive open debate and the free-flow of information and opinion. If you can’t win a debate then don’t have one is the Cult’s approach throughout history. Facebook’s little boy front man – front boy – Mark Zuckerberg equated ‘credible and accurate information’ with official sources and exposing their lies with ‘misinformation’.

Silencing those that can see

The censorship dynamic of Event 201 is now the norm with an army of narrative-supporting ‘fact-checker’ organisations whose entire reason for being is to tell the public that official narratives are true and those exposing them are lying. One of the most appalling of these ‘fact-checkers’ is called NewsGuard founded by ultra-Zionist Americans Gordon Crovitz and Steven Brill. Crovitz is a former publisher of *The Wall Street Journal*, former Executive Vice President of Dow Jones, a member of the Council on Foreign Relations (CFR), and on the board of the American Association of Rhodes Scholars. The CFR and Rhodes Scholarships, named after Rothschild agent Cecil Rhodes who plundered the gold and diamonds of South Africa for his masters and the Cult, have featured widely in my books. NewsGuard don’t seem to like me for some reason – I really can’t think why – and they have done all they can to have me censored and discredited which is, to quote an old British politician, like being savaged by a dead sheep. They are, however, like all in the censorship network, very well connected and funded by organisations themselves funded by, or connected to, Bill Gates. As you would expect with anything associated with Gates NewsGuard has an offshoot called HealthGuard which ‘fights online health care hoaxes’. How very kind. Somehow the NewsGuard European Managing Director Anna-Sophie Harling, a remarkably young-looking woman with no broadcasting experience and little hands-on work in journalism, has somehow secured a position on the ‘Content Board’ of UK government broadcast censor Ofcom. An executive of an organisation seeking to discredit dissidents of the official narratives is making decisions for the government broadcast ‘regulator’ about content?? Another appalling ‘fact-checker’ is Full Fact funded by George Soros and global censors Google and Facebook.

It’s amazing how many activists in the ‘fact-checking’, ‘anti-hate’, arena turn up in government-related positions – people like UK Labour Party activist Imran Ahmed who heads the Center for Countering Digital Hate founded by people like Morgan McSweeney, now chief of staff to the Labour Party’s hapless and useless ‘leader’ Keir Starmer. Digital Hate – which is what it really is – uses the American spelling of Center to betray its connection to a transatlantic network of similar organisations which in 2020 shapeshifted from attacking people for ‘hate’ to attacking them for questioning the ‘Covid’ hoax and the dangers of the ‘Covid vaccine’. It’s just a coincidence, you understand. This is one of Imran Ahmed’s hysterical statements: ‘I would go beyond calling anti-vaxxers conspiracy theorists to say they are an extremist group that pose a national security risk.’ No one could ever accuse this prat of understatement and he’s including in that those parents who are now against vaccines after their children were damaged for life or killed by them. He’s such a nice man. Ahmed does the rounds of the Woke media getting soft-ball questions from spineless ‘journalists’ who never ask what right he has to campaign to destroy the freedom of speech of others while he demands it for himself. There also seems to be an overrepresentation in Ofcom of people connected to the narrative-worshipping BBC. This incredible global network of narrative-support was super-vital when the ‘Covid’ hoax was played in the light of the mega-whopper lies that have to be defended from the spotlight cast by the most basic intelligence.

Setting the scene

The Cult plays the long game and proceeds step-by-step ensuring that everything is in place before major cards are played and they don't come any bigger than the 'Covid' hoax. The psychopaths can't handle events where the outcome isn't certain and as little as possible – preferably nothing – is left to chance. Politicians, government and medical officials who would follow direction were brought to illusory power in advance by the Cult web whether on the national stage or others like state governors and mayors of America. For decades the dynamic between officialdom, law enforcement and the public was changed from one of service to one of control and dictatorship. Behaviour manipulation networks established within government were waiting to impose the coming 'Covid' rules and regulations specifically designed to subdue and rewire the psyche of the people in the guise of protecting health. These included in the UK the Behavioural Insights Team part-owned by the British government Cabinet Office; the Scientific Pandemic Insights Group on Behaviours (SPI-B); and a whole web of intelligence and military groups seeking to direct the conversation on social media and control the narrative. Among them are the cyberwarfare (on the people) 77th Brigade of the British military which is also coordinated through the Cabinet Office as civilian and military leadership continues to combine in what they call the Fusion Doctrine. The 77th Brigade is a British equivalent of the infamous Israeli (Sabbatian) military cyberwarfare and Internet manipulation operation Unit 8200 which I expose at length in *The Trigger*. Also carefully in place were the medical and science advisers to government – many on the payroll past or present of Bill Gates – and a whole alternative structure of unelected government stood by to take control when elected parliaments were effectively closed down once the 'Covid' card was slammed on the table. The structure I have described here and so much more was installed in every major country through the Cult networks. The top-down control hierarchy looks like this: The Cult – Cult-owned Gates – the World Health Organization and Tedros – Gates-funded or controlled chief medical officers and science 'advisers' (dictators) in each country –

political 'leaders'– law enforcement – The People. Through this simple global communication and enforcement structure the policy of the Cult could be imposed on virtually the entire human population so long as they acquiesced to the fascism. With everything in place it was time for the button to be pressed in late 2019/early 2020.

These were the prime goals the Cult had to secure for its will to prevail:

1) Locking down economies, closing all but designated 'essential' businesses (Cult-owned corporations were 'essential'), and putting the population under house arrest was an imperative to destroy independent income and employment and ensure dependency on the Cult-controlled state in the Hunger Games Society. Lockdowns had to be established as the global blueprint from the start to respond to the 'virus' and followed by pretty much the entire world.

2) The global population had to be terrified into believing in a deadly 'virus' that didn't actually exist so they would unquestioningly obey authority in the belief that authority must know how best to protect them and their families. Software salesman Gates would suddenly morph into the world's health expert and be promoted as such by the Cult-owned media.

3) A method of testing that wasn't testing for the 'virus', but was only claimed to be, had to be in place to provide the illusion of 'cases' and subsequent 'deaths' that had a very different cause to the 'Covid-19' that would be scribbled on the death certificate.

4) Because there was no 'virus' and the great majority testing positive with a test not testing for the 'virus' would have no symptoms of anything the lie had to be sold that people without symptoms (without the 'virus') could still pass it on to others. This was crucial to justify for the first time quarantining – house arresting – healthy people. Without this the economy-destroying lockdown of *everybody* could not have been credibly sold.

5) The 'saviour' had to be seen as a vaccine which beyond evil drug companies were working like angels of mercy to develop as quickly as possible, with all corners cut, to save the day. The public must absolutely not know that the 'vaccine' had nothing to do with a 'virus' or that the contents were ready and waiting with a very different motive long before the 'Covid' card was even lifted from the pack.

I said in March, 2020, that the 'vaccine' would have been created way ahead of the 'Covid' hoax which justified its use and the following December an article in the New York *Intelligencer* magazine said the Moderna 'vaccine' had been 'designed' by

January, 2020. This was 'before China had even acknowledged that the disease could be transmitted from human to human, more than a week before the first confirmed coronavirus case in the United States'. The article said that by the time the first American death was announced a month later 'the vaccine had already been manufactured and shipped to the National Institutes of Health for the beginning of its Phase I clinical trial'. The 'vaccine' was actually 'designed' long before that although even with this timescale you would expect the article to ask how on earth it could have been done that quickly. Instead it asked why the 'vaccine' had not been rolled out then and not months later. Journalism in the mainstream is truly dead. I am going to detail in the next chapter why the 'virus' has never existed and how a hoax on that scale was possible, but first the foundation on which the Big Lie of 'Covid' was built.

The test that doesn't test

Fraudulent 'testing' is the bottom line of the whole 'Covid' hoax and was the means by which a 'virus' that did not exist *appeared* to exist. They could only achieve this magic trick by using a test not testing for the 'virus'. To use a test that *was* testing for the 'virus' would mean that every test would come back negative given there was no 'virus'. They chose to exploit something called the RT-PCR test invented by American biochemist Kary Mullis in the 1980s who said publicly that his PCR test … *cannot detect infectious disease*. Yes, the 'test' used worldwide to detect infectious 'Covid' to produce all the illusory 'cases' and 'deaths' compiled by Johns Hopkins and others *cannot detect infectious disease*. This fact came from the mouth of the man who invented PCR and was awarded the Nobel Prize in Chemistry in 1993 for doing so. Sadly, and incredibly conveniently for the Cult, Mullis died in August, 2019, at the age of 74 just before his test would be fraudulently used to unleash fascism on the world. He was said to have died from pneumonia which was an irony in itself. A few months later he would have had 'Covid-19' on his death certificate. I say the timing of his death was convenient because had he lived Mullis, a brilliant, honest and decent man, would have been vociferously speaking out against the use of his test to detect ‘Covid’ when it was never designed, or able, to do that. I know that to be true given that Mullis made the same point when his test was used to ‘detect’ – not detect – HIV. He had been seriously critical of the Gallo/Montagnier claim to have isolated the HIV ‘virus’ and shown it to cause AIDS for which Mullis said there was no evidence. AIDS is actually not a disease but a series of diseases from which people die all the time. When they die from those *same diseases* after a positive ‘test’ for HIV then AIDS goes on their death certificate. I think I’ve heard that before somewhere. Countries instigated a policy with ‘Covid’ that anyone who tested positive with a test not testing for the ‘virus’ and died of any other cause within 28 days and even longer ‘Covid-19’ had to go on the death certificate. Cases have come from the test that can’t test for infectious disease and the deaths are those who have died of *anything* after testing positive with a test not testing for the ‘virus’. I’ll have much more later about the death certificate scandal.

Mullis was deeply dismissive of the now US ‘Covid’ star Anthony Fauci who he said was a liar who didn’t know anything about anything – ‘and I would say that to his face – nothing.’ He said of Fauci: ‘The man thinks he can take a blood sample, put it in an electron microscope and if it’s got a virus in there you’ll know it – he doesn’t understand electron microscopy and he doesn’t understand medicine and shouldn’t be in a position like he’s in.’ That position, terrifyingly, has made him the decider of ‘Covid’ fascism policy on behalf of the Cult in his role as director since 1984 of the National Institute of Allergy and Infectious Diseases (NIAID) while his record of being wrong is laughable; but being wrong, so long as it’s the *right kind* of wrong, is why the Cult loves him. He’ll say anything the Cult tells him to say. Fauci was made Chief Medical Adviser to the President immediately Biden took office. Biden was installed in the White House by Cult manipulation and one of his first decisions was to elevate Fauci to a position of even more control. This is a coincidence? Yes, and I identify as a flamenco dancer called Lola. How does such an incompetent criminal like Fauci remain in that pivotal position in American health since *the 1980s*? When you serve the Cult it looks after you until you are surplus to requirements. Kary Mullis said prophetically of Fauci and his like: 'Those guys have an agenda and it's not an agenda we would like them to have … they make their own rules, they change them when they want to, and Tony Fauci does not mind going on television in front of the people who pay his salary and lie directly into the camera.' Fauci has done that almost daily since the 'Covid' hoax began. Lying is in Fauci's DNA. To make the situation crystal clear about the PCR test this is a direct quote from its inventor Kary Mullis:

> It [the PCR test] doesn't tell you that you're sick and doesn't tell you that the thing you ended up with was really going to hurt you ...'

Ask yourself why governments and medical systems the world over have been using this very test to decide who is 'infected' with the SARS-CoV-2 'virus' and the alleged disease it allegedly causes, 'Covid-19'. The answer to that question will tell you what has been going on. By the way, here's a little show-stopper – the 'new' SARS-CoV-2 'virus' was 'identified' as such right from the start using … *the PCR test not testing for the 'virus'*. If you are new to this and find that shocking then stick around. I have hardly started yet. Even worse, other 'tests', like the 'Lateral Flow Device' (LFD), are considered so useless that they have to be *confirmed* by the PCR test! Leaked emails written by Ben Dyson, adviser to UK 'Health' Secretary Matt Hancock, said they were 'dangerously unreliable'. Dyson, executive director of strategy at the Department of Health, wrote: 'As of today, someone who gets a positive LFD result in (say) London has at best a 25 per cent chance of it being a true positive, but if it is a self-reported test potentially as low as 10 per cent (on an optimistic assumption about specificity) or as low as 2 per cent (on a more pessimistic assumption).' These are the 'tests' that schoolchildren and the public are being urged to have twice a week or more and have to isolate if they get a positive. Each fake positive goes in the statistics as a 'case' no matter how ludicrously inaccurate and the ‘cases’ drive lockdown, masks and the pressure to ‘vaccinate’. The government said in response to the email leak that the ‘tests’ were accurate which confirmed yet again what shocking bloody liars they are. The real false positive rate is *100 percent* as we’ll see. In another ‘you couldn’t make it up’ the UK government agreed to pay £2.8 billion to California’s Innova Medical Group to supply the irrelevant lateral flow tests. The company’s primary test-making centre is in China. Innova Medical Group, established in March, 2020, is owned by Pasaca Capital Inc, chaired by Chinese-American millionaire Charles Huang who was born in Wuhan.

How it works – and how it doesn’t

The RT-PCR test, known by its full title of Polymerase chain reaction, is used across the world to make millions, even billions, of copies of a DNA/RNA genetic information sample. The process is called ‘amplification’ and means that a tiny sample of genetic material is amplified to bring out the detailed content. I stress that it is not testing for an infectious disease. It is simply amplifying a sample of genetic material. In the words of Kary Mullis: ‘PCR is … just a process that’s used to make a whole lot of something out of something.’ To emphasise the point companies that make the PCR tests circulated around the world to ‘test’ for ‘Covid’ warn on the box that it can’t be used to detect ‘Covid’ or infectious disease and is for research purposes only. It’s okay, rest for a minute and you’ll be fine. This is the test that produces the ‘cases’ and ‘deaths’ that have been used to destroy human society. All those global and national medical and scientific ‘experts’ demanding this destruction to ‘save us’ *KNOW* that the test is not testing for the ‘virus’ and the cases and deaths they claim to be real are an almost unimaginable fraud. Every one of them and so many others including politicians and psychopaths like Gates and Tedros must be brought before Nuremburg-type trials and jailed for the rest of their lives. The more the genetic sample is amplified by PCR the more elements of that material become sensitive to the test and by that I don’t mean sensitive for a ‘virus’ but for elements of the genetic material which is *naturally* in the body or relates to remnants of old conditions of various kinds lying dormant and causing no disease. Once the amplification of the PCR reaches a certain level *everyone* will test positive. So much of the material has been made sensitive to the test that everyone will have some part of it in their body. Even lying criminals like Fauci have said that once PCR amplifications pass 35 cycles everything will be a false positive that cannot be trusted for the reasons I have described. I say, like many proper doctors and scientists, that 100 percent of the 'positives' are false, but let's just go with Fauci for a moment.

He says that any amplification over 35 cycles will produce false positives and yet the US Centers for Disease Control (CDC) and Food and Drug Administration (FDA) have recommended up to *40 cycles* and the National Health Service (NHS) in Britain admitted in an internal document for staff that it was using *45 cycles* of amplification. A long list of other countries has been doing the same and at least one 'testing' laboratory has been using *50 cycles*. Have you ever heard a doctor, medical 'expert' or the media ask what level of amplification has been used to claim a 'positive'. The 'test' comes back 'positive' and so you have the 'virus', end of story. Now we can see how the government in Tanzania could send off samples from a goat and a pawpaw fruit under human names and both came back positive for 'Covid-19'. Tanzania president John Magufuli mocked the 'Covid' hysteria, the PCR test and masks and refused to import the DNA-manipulating 'vaccine'. The Cult hated him and an article sponsored by the Bill Gates Foundation appeared in the London *Guardian* in February, 2021, headed 'It's time for Africa to rein in Tanzania's anti-vaxxer president'. Well, 'reined in' he shortly was. Magufuli appeared in good health, but then, in March, 2021, he was dead at 61 from 'heart failure'. He was replaced by Samia Hassan Suhulu who is connected to Klaus Schwab's World Economic Forum and she immediately reversed Magufuli's 'Covid' policy. A sample of cola tested positive for 'Covid' with the PCR test in Germany while American actress and singer-songwriter Erykah Badu tested positive in one nostril and negative in the other. Footballer Ronaldo called the PCR test ‘bullshit’ after testing positive three times and being forced to quarantine and miss matches when there was nothing wrong with him. The mantra from Tedros at the World Health Organization and national governments (same thing) has been test, test, test. They know that the more tests they can generate the more fake ‘cases’ they have which go on to become ‘deaths’ in ways I am coming to. The UK government has its Operation Moonshot planned to test multiple millions every day in workplaces and schools with free tests for everyone to use twice a week at home in line with the Cult plan from the start to make testing part of life. A government advertisement for an ‘Interim Head of Asymptomatic Testing Communication’ said the job included responsibility for delivering a ‘communications strategy’ (propaganda) ‘to support the expansion of asymptomatic testing that *‘normalises testing as part of everyday life*’. More tests means more fake ‘cases’, ‘deaths’ and fascism. I have heard of, and from, many people who booked a test, couldn’t turn up, and yet got a positive result through the post for a test they’d never even had. The whole thing is crazy, but for the Cult there’s method in the madness. Controlling and manipulating the level of amplification of the test means the authorities can control whenever they want the number of apparent ‘cases’ and ‘deaths’. If they want to justify more fascist lockdown and destruction of livelihoods they keep the amplification high. If they want to give the illusion that lockdowns and the ‘vaccine’ are working then they lower the amplification and ‘cases’ and ‘deaths’ will appear to fall. In January, 2021, the Cult-owned World Health Organization suddenly warned laboratories about over-amplification of the test and to lower the threshold. Suddenly headlines began appearing such as: ‘Why ARE “Covid” cases plummeting?’ This was just when the vaccine rollout was underway and I had predicted months before they would make cases appear to fall through amplification tampering when the ‘vaccine’ came. These people are so predictable.

Cow vaccines?

The question must be asked of what is on the test swabs being poked far up the nose of the population to the base of the brain? A nasal swab punctured one woman's brain and caused it to leak fluid. Most of these procedures are being done by people with little training or medical knowledge. Dr Lorraine Day, former orthopaedic trauma surgeon and Chief of Orthopaedic Surgery at San Francisco General Hospital, says the tests are really a *'vaccine'*. Cows have long been vaccinated this way. She points out that masks have to cover the nose and the mouth where it is claimed the 'virus' exists in saliva. Why then don't they take saliva from the mouth as they do with a DNA test instead of pushing a long swab up the nose towards the brain? The ethmoid bone separates the nasal cavity from the brain and within that bone is the cribriform plate. Dr Day says that when the swab is pushed up against this plate and twisted the procedure is 'depositing things back there'. She claims that among these 'things' are nanoparticles that can enter the brain. Researchers have noted that a team at the Gates-funded Johns Hopkins have designed tiny, star-shaped micro-devices that can latch onto intestinal mucosa and release drugs into the body. Mucosa is the thin skin that covers the inside surface of parts of the body such as *the nose* and mouth and produces mucus to protect them. The Johns Hopkins micro-devices are called 'theragrippers' and were 'inspired' by a parasitic worm that digs its sharp teeth into a host's intestines. Nasal swabs are also coated in the sterilisation agent ethylene oxide. The US National Cancer Institute posts this explanation on its website:

> At room temperature, ethylene oxide is a flammable colorless gas with a sweet odor. It is used primarily to produce other chemicals, including antifreeze. In smaller amounts, ethylene oxide is used as a pesticide and a sterilizing agent. The ability of ethylene oxide to damage DNA makes it an effective sterilizing agent but also accounts for its cancer-causing activity.

The Institute mentions lymphoma and leukaemia as cancers most frequently reported to be associated with occupational exposure to ethylene oxide along with stomach and breast cancers. How does anyone think this is going to work out with the constant testing regime being inflicted on adults and children at home and at school that will accumulate in the body anything that's on the swab?

Doctors know best

It is vital for people to realise that 'hero' doctors 'know' only what the Big Pharma-dominated medical authorities tell them to 'know' and if they refuse to 'know' what they are told to 'know' they are out the door. They are mostly not physicians or healers, but repeaters of the official narrative – or else. I have seen alleged professional doctors on British television make shocking statements that we are supposed to take seriously. One called 'Dr' Amir Khan, who is actually telling patients how to respond to illness, said that men could take the birth pill to 'help slow down the effects of Covid-19'. In March, 2021, another ridiculous 'Covid study' by an American doctor proposed injecting men with the female sex hormone progesterone as a 'Covid' treatment. British doctor Nighat Arif told the BBC that face coverings were now going to be part of ongoing normal. Yes, the vaccine protects you, she said (evidence?) … but the way to deal with viruses in the community was always going to come down to hand washing, face covering and keeping a physical distance. That's not what we were told before the 'vaccine' was circulating. Arif said she couldn't imagine ever again going on the underground or in a lift without a mask. I was just thanking my good luck that she was not my doctor when she said – in March, 2021 – that if 'we are *behaving* and we are doing all the right things' she thought we could 'have our nearest and dearest around us at home … around *Christmas* and N*ew Year!* Her patronising delivery was the usual school teacher talking to six-year-olds as she repeated every government talking point and probably believed them all. If we have learned anything from the 'Covid' experience surely it must be that humanity's perception of doctors needs a fundamental rethink. NHS 'doctor' Sara Kayat told her television audience that the 'Covid vaccine' would '100 percent prevent hospitalisation and death'. Not even Big Pharma claimed that. We have to stop taking 'experts' at their word without question when so many of them are clueless and only repeating the party line on which their careers depend. That is not to say there are not brilliants doctors – there are and I have spoken to many of them since all this began – but you won't see them in the mainstream media or quoted by the psychopaths and yes-people in government.

Remember the name – Christian Drosten

German virologist Christian Drosten, Director of Charité Institute of Virology in Berlin, became a national star after the pandemic hoax began. He was feted on television and advised the German government on 'Covid' policy. Most importantly to the wider world Drosten led a group that produced the 'Covid' testing protocol for the PCR test. What a remarkable feat given the PCR cannot test for infectious disease and even more so when you think that Drosten said that his method of testing for SARS-CoV-2 was developed 'without having virus material available'. *He developed a test for a 'virus' that he didn't have and had never seen*. Let that sink in as you survey the global devastation that came from what he did. The whole catastrophe of Drosten's 'test' was based on the alleged genetic sequence published by Chinese scientists on the Internet. We will see in the next chapter that this alleged 'genetic sequence' has never been produced by China or anyone and cannot be when there *is no* SARS-CoV-2. Drosten, however, doesn't seem to let little details like that get in the way. He was the lead author with Victor Corman from the same Charité Hospital of the paper 'Detection of 2019 novel coronavirus (2019-nCoV) by real-time PCR' published in a magazine called *Eurosurveillance*. This became known as the Corman-Drosten paper. In November, 2020, with human society devastated by the effects of the Corman-Drosten test baloney, the protocol was publicly challenged by 22 international scientists and independent researchers from Europe, the United States, and Japan. Among them were senior molecular geneticists, biochemists, immunologists, and microbiologists. They produced a document headed 'External peer review of the RTPCR test to detect SARS-Cov-2 Reveals 10 Major Flaws At The Molecular and Methodological Level: Consequences

For False-Positive Results’. The flaws in the Corman-Drosten test included the following:

- The test is non-specific because of erroneous design
- Results are enormously variable
- The test is unable to discriminate between the whole ‘virus’ and viral fragments
- It doesn’t have positive or negative controls
- The test lacks a standard operating procedure
- It is unsupported by proper peer view

The scientists said the PCR ‘Covid’ testing protocol was not founded on science and they demanded the Corman-Drosten paper be retracted by *Eurosurveillance*. They said all present and previous Covid deaths, cases, and ‘infection rates’ should be subject to a massive retroactive inquiry. Lockdowns and travel restrictions should be reviewed and relaxed and those diagnosed through PCR to have ‘Covid-19’ should not be forced to isolate. Dr Kevin Corbett, a health researcher and nurse educator with a long academic career producing a stream of peer-reviewed publications at many UK universities, made the same point about the PCR test debacle. He said of the scientists’ conclusions: ‘Every scientific rationale for the development of that test has been totally destroyed by this paper. It’s like Hiroshima/Nagasaki to the Covid test.’ He said that China hadn’t given them an isolated ‘virus’ when Drosten developed the test. Instead they had developed the test from *a sequence in a gene bank*.’ Put another way … *they made it up*! The scientists were supported in this contention by a Portuguese appeals court which ruled in November, 2020, that PCR tests are unreliable and it is unlawful to quarantine people based solely on a PCR test. The point about China not providing an isolated virus must be true when the ‘virus’ has never been isolated to this day and the consequences of that will become clear. Drosten and company produced this useless ‘protocol’ right on cue in January, 2020, just as the ‘virus’ was said to be moving westward and it somehow managed to successfully pass a peer-review in 24 hours. In other words there was no peer-review for a test that would be used to decide who had ‘Covid’ and who didn’t across the world. The Cult-created, Gates-controlled World Health Organization immediately recommended all its nearly 200 member countries to use the Drosten PCR protocol to detect ‘cases’ and ‘deaths’. The sting was underway and it continues to this day.

So who is this Christian Drosten that produced the means through which death, destruction and economic catastrophe would be justified? His education background, including his doctoral thesis, would appear to be somewhat shrouded in mystery and his track record is dire as with another essential player in the ‘Covid’ hoax, the Gates-funded Professor Neil Ferguson at the Gates-funded Imperial College in London of whom more shortly. Drosten predicted in 2003 that the alleged original SARS ‘virus’ (SARS-1’) was an epidemic that could have serious effects on economies and an effective vaccine would take at least two years to produce. Drosten’s answer to every alleged ‘outbreak’ is a vaccine which you won’t be shocked to know. What followed were just 774 official deaths worldwide and none in Germany where there were only nine cases. That is even if you believe there ever was a SARS ‘virus’ when the evidence is zilch and I will expand on this in the next chapter. Drosten claims to be co-discoverer of ‘SARS-1’ and developed a test for it in 2003. He was screaming warnings about ‘swine flu’ in 2009 and how it was a widespread infection far more severe than any dangers from a vaccine could be and people should get vaccinated. It would be helpful for Drosten’s vocal chords if he simply recorded the words ‘the virus is deadly and you need to get vaccinated’ and copies could be handed out whenever the latest made-up threat comes along. Drosten’s swine flu epidemic never happened, but Big Pharma didn’t mind with governments spending hundreds of millions on vaccines that hardly anyone bothered to use and many who did wished they hadn’t. A study in 2010 revealed that the risk of dying from swine flu, or H1N1, was no higher than that of the annual seasonal flu which is what at least most of ‘it’ really was as in the case of ‘Covid-19’. A media investigation into Drosten asked how with such a record of inaccuracy he could be *the* government adviser on these issues. The answer to that question is the same with Drosten, Ferguson and Fauci – they keep on giving the authorities the ‘conclusions’ and ‘advice’ they want to hear. Drosten certainly produced the goods for them in January, 2020, with his PCR protocol garbage and provided the foundation of what German internal medicine specialist Dr Claus Köhnlein, co-author of *Virus Mania*, called the ‘test pandemic’. The 22 scientists in the *Eurosurveillance* challenge called out conflicts of interest within the Drosten ‘protocol’ group and with good reason. Olfert Landt, a regular co-author of Drosten ‘studies’, owns the biotech company TIB Molbiol Syntheselabor GmbH in Berlin which manufactures and sells the tests that Drosten and his mates come up with. They have done this with SARS, Enterotoxigenic E. coli (ETEC), MERS, Zika ‘virus’, yellow fever, and now ‘Covid’. Landt told the *Berliner Zeitung* newspaper:

> The testing, design and development came from the Charité [Drosten and Corman]. We simply implemented it immediately in the form of a kit. And if we don’t have the virus, which originally only existed in Wuhan, we can make a synthetic gene to simulate the genome of the virus. That’s what we did very quickly.

This is more confirmation that the Drosten test was designed without access to the ‘virus’ and only a synthetic simulation which is what SARS-CoV-2 really is – a computer-generated synthetic fiction. It’s quite an enterprise they have going here. A Drosten team decides what the test for something should be and Landt’s biotech company flogs it to governments and medical systems across the world. His company must have made an absolute fortune since the ‘Covid’ hoax began. Dr Reiner Fuellmich, a prominent German consumer protection trial lawyer in Germany and California, is on Drosten’s case and that of Tedros at the World Health Organization for crimes against humanity with a class-action lawsuit being prepared in the United States and other legal action in Germany.

Why China?

Scamming the world with a 'virus' that doesn't exist would seem impossible on the face of it, but not if you have control of the relatively few people that make policy decisions and the great majority of the global media. Remember it's not about changing 'real' reality it's about controlling *perception* of reality. You don't have to make something happen you only have make people *believe* that it's happening. Renegade Minds understand this and are therefore much harder to swindle. 'Covid-19' is not a 'real' 'virus'. It's a mind virus, like a computer virus, which has infected the minds, not the bodies, of billions. It all started, publically at least, in China and that alone is of central significance. The Cult was behind the revolution led by its asset Mao Zedong, or Chairman Mao, which established the People's Republic of China on October 1st, 1949. It should have been called The Cult's Republic of China, but the name had to reflect the recurring illusion that vicious dictatorships are run by and for the people (see all the 'Democratic Republics' controlled by tyrants). In the same way we have the 'Biden' Democratic Republic of America officially ruled by a puppet tyrant (at least temporarily) on behalf of Cult tyrants. The creation of Mao's merciless communist/fascist dictatorship was part of a frenzy of activity by the Cult at the conclusion of World War Two which, like the First World War, it had instigated through its assets in Germany, Britain, France, the United States and elsewhere. Israel was formed in 1948; the Soviet Union expanded its 'Iron Curtain' control, influence and military power with the Warsaw Pact communist alliance in 1955; the United Nations was formed in 1945 as a Cult precursor to world government; and a long list of world bodies would be established including the World Health Organization (1948), World Trade Organization (1948 under another name until 1995), International Monetary Fund (1945) and World Bank (1944). Human society was redrawn and hugely centralised in the global Problem-Reaction-Solution that was World War Two. All these changes were significant. Israel would become the headquarters of the Sabbatians and the revolution in China would prepare the ground and control system for the events of 2019/2020.

Renegade Minds know there are no borders except for public consumption. The Cult is a seamless, borderless global entity and to understand the game we need to put aside labels like borders, nations, countries, communism, fascism and democracy. These delude the population into believing that countries are ruled within their borders by a government of whatever shade when these are mere agencies of a global power. America's illusion of democracy and China's communism/fascism are subsidiaries – vehicles – for the same agenda. We may hear about conflict and competition between America and China and on the lower levels that will be true; but at the Cult level they are branches of the same company in the way of the McDonald's example I gave earlier. I have tracked in the books over the years support by US governments of both parties for Chinese Communist Party infiltration of American society through allowing the sale of land, even military facilities, and the acquisition of American business and university influence. All this is underpinned by the infamous stealing of intellectual property and technological know-how. Cult-owned Silicon Valley corporations waive their fraudulent 'morality' to do business with human-rights-free China; Cult-controlled Disney has become China's PR department; and China in effect owns 'American' sports such as basketball which depends for much of its income on Chinese audiences. As a result any sports player, coach or official speaking out against China's horrific human rights record is immediately condemned or fired by the China-worshipping National Basketball Association. One of the first acts of China-controlled Biden was to issue an executive order telling federal agencies to stop making references to the 'virus' by the 'geographic location of its origin'. Long-time Congressman Jerry Nadler warned that criticising China, America's biggest rival, leads to hate crimes against Asian people in the United States. So shut up you bigot. China is fast closing in on Israel as a country that must not be criticised which is apt, really, given that Sabbatians control them both. The two countries have developed close economic, military, technological and strategic ties which include involvement in China's 'Silk Road' transport and economic initiative to connect China with Europe. Israel was the first country in the Middle East to recognise the establishment of Mao's tyranny in 1950 months after it was established.

Project Wuhan – the 'Covid' Psyop

I emphasise again that the Cult plays the long game and what is happening to the world today is the result of centuries of calculated manipulation following a script to take control step-by-step of every aspect of human society. I will discuss later the common force behind all this that has spanned those centuries and thousands of years if the truth be told. Instigating the Mao revolution in China in 1949 with a 2020 'pandemic' in mind is not only how they work – the 71 years between them is really quite short by the Cult's standards of manipulation preparation. The reason for the Cult's Chinese revolution was to create a fiercely-controlled environment within which an extreme structure for human control could be incubated to eventually be unleashed across the world. We have seen this happen since the 'pandemic' emerged from China with the Chinese control-structure founded on AI technology and tyrannical enforcement sweep across the West. Until the moment when the Cult went for broke in the West and put its fascism on public display Western governments had to pay some lip-service to freedom and democracy to not alert too many people to the tyranny-in-the-making. Freedoms were more subtly eroded and power centralised with covert government structures put in place waiting for the arrival of 2020 when that smokescreen of 'freedom' could be dispensed with. The West was not able to move towards tyranny before 2020 anything like as fast as China which was created as a tyranny and had no limits on how fast it could construct the Cult's blueprint for global control. When the time came to impose that structure on the world it was the same Cult-owned Chinese communist/fascist government that provided the excuse – the 'Covid pandemic'. It was absolutely crucial to the Cult plan for the Chinese response to the 'pandemic' –

draconian lockdowns of the entire population – to become the blueprint that Western countries would follow to destroy the livelihoods and freedom of their people. This is why the Cult-owned, Gates-owned, WHO Director-General Tedros said early on:

> The Chinese government is to be congratulated for the extraordinary measures it has taken to contain the outbreak. China is actually setting a new standard for outbreak response and it is not an exaggeration.

Forbes magazine said of China: '… those measures protected untold millions from getting the disease'. The Rockefeller Foundation 'epidemic scenario' document in 2010 said 'prophetically':

> However, a few countries did fare better – China in particular. The Chinese government's quick imposition and enforcement of mandatory quarantine for all citizens, as well as its instant and near-hermetic sealing off of all borders, saved millions of lives, stopping the spread of the virus far earlier than in other countries and enabling a swifter post-pandemic recovery.

Once again – *spooky*.

The first official story was the 'bat theory' or rather the bat diversion. The source of the 'virus outbreak' we were told was a ''wet market' in Wuhan where bats and other animals are bought and eaten in horrifically unhygienic conditions. Then another story emerged through the alternative media that the 'virus' had been released on purpose or by accident from a BSL-4 (biosafety level 4) laboratory in Wuhan not far from the wet market. The lab was reported to create and work with lethal concoctions and bioweapons. Biosafety level 4 is the highest in the World Health Organization system of safety and containment. Renegade Minds are aware of what I call designer manipulation. The ideal for the Cult is for people to buy its prime narrative which in the opening salvoes of the 'pandemic' was the wet market story. It knows, however, that there is now a considerable worldwide alternative media of researchers sceptical of anything governments say and they are often given a version of events in a form they can perceive as credible while misdirecting them from the real truth. In this case let them think that the conspiracy involved is a ‘bioweapon virus’ released from the Wuhan lab to keep them from the real conspiracy – *there is no ‘virus’*. The WHO’s current position on the source of the outbreak at the time of writing appears to be: ‘We haven’t got a clue, mate.’ This is a good position to maintain mystery and bewilderment. The inner circle will know where the ‘virus’ came from – *nowhere*. The bottom line was to ensure the public believed there *was* a ‘virus’ and it didn’t much matter if they thought it was natural or had been released from a lab. The belief that there was a ‘deadly virus’ was all that was needed to trigger global panic and fear. The population was terrified into handing their power to authority and doing what they were told. They had to or they were ‘all gonna die’.

In March, 2020, information began to come my way from real doctors and scientists and my own additional research which had my intuition screaming: ‘Yes, that’s it! *There is no virus*.’ The ‘bioweapon’ was not the ‘virus’; it was the *‘vaccine’* already being talked about that would be the bioweapon. My conclusion was further enhanced by happenings in Wuhan. The ‘virus’ was said to be sweeping the city and news footage circulated of people collapsing in the street (which they’ve never done in the West with the same ‘virus’). The Chinese government was building ‘new hospitals’ in a matter of ten days to ‘cope with demand’ such was the virulent nature of the ‘virus’. Yet in what seemed like no time the ‘new hospitals’ closed – even if they even opened – and China declared itself ‘virus-free’. It was back to business as usual. This was more propaganda to promote the Chinese draconian lockdowns in the West as the way to ‘beat the virus’. Trouble was that we subsequently had lockdown after lockdown, but never business as usual. As the people of the West and most of the rest of the world were caught in an ever-worsening spiral of lockdown, social distancing, masks, isolated old people, families forced apart, and livelihood destruction, it was party-time in Wuhan. Pictures emerged of thousands of people enjoying pool parties and concerts. It made no sense until you realised there never was a ‘virus’ and the whole thing was a Cult set-up to transform human society out of one its major global strongholds – China.

How is it possible to deceive virtually the entire world population into believing there is a deadly virus when there is not even a ‘virus’ let alone a deadly one? It’s nothing like as difficult as you would think and that’s clearly true because it happened.

Postscript: See end of book Postscript for more on the ‘Wuhan lab virus release’ story which the authorities and media were pushing heavily in the summer of 2021 to divert attention from the truth that the ‘Covid virus’ is pure invention.

CHAPTER FIVE

There *is no* 'virus'

You can fool some of the people all of the time, and all of the people some of the time, but you cannot fool all of the people all of the time
Abraham Lincoln

The greatest form of mind control is repetition. The more you repeat the same mantra of alleged 'facts' the more will accept them to be true. It becomes an 'everyone knows that, mate'. If you can also censor any other version or alternative to your alleged 'facts' you are pretty much home and cooking.

By the start of 2020 the Cult owned the global mainstream media almost in its entirety to spew out its 'Covid' propaganda and ignore or discredit any other information and view. Cult-owned social media platforms in Cult-owned Silicon Valley were poised and ready to unleash a campaign of ferocious censorship to obliterate all but the official narrative. To complete the circle many demands for censorship by Silicon Valley were led by the mainstream media as 'journalists' became full-out enforcers for the Cult both as propagandists and censors. Part of this has been the influx of young people straight out of university who have become 'journalists' in significant positions. They have no experience and a headful of programmed perceptions from their years at school and university at a time when today's young are the most perceptually-targeted generations in known human history given the insidious impact of technology. They enter the media perceptually prepared and ready to repeat the narratives of the system that programmed them to repeat its narratives. The BBC has a truly pathetic ‘specialist disinformation reporter’ called Marianna Spring who fits this bill perfectly. She is clueless about the world, how it works and what is really going on. Her role is to discredit anyone doing the job that a proper journalist would do and system-serving hacks like Spring wouldn’t dare to do or even see the need to do. They are too busy licking the arse of authority which can never be wrong and, in the case of the BBC propaganda programme, *Panorama*, contacting payments systems such as PayPal to have a donations page taken down for a film company making documentaries questioning vaccines. Even the BBC soap opera *EastEnders* included a disgracefully biased scene in which an inarticulate white working class woman was made to look foolish for questioning the ‘vaccine’ while a well-spoken black man and Asian woman promoted the government narrative. It ticked every BBC box and the fact that the black and minority community was resisting the ‘vaccine’ had nothing to do with the way the scene was written. The BBC has become a disgusting tyrannical propaganda and censorship operation that should be defunded and disbanded and a free media take its place with a brief to stop censorship instead of demanding it. A BBC ‘interview’ with Gates goes something like: ‘Mr Gates, sir, if I can call you sir, would you like to tell our audience why you are such a great man, a wonderful humanitarian philanthropist, and why you should absolutely be allowed as a software salesman to decide health policy for approaching eight billion people? Thank you, sir, please sir.’ Propaganda programming has been incessant and merciless and when all you hear is the same story from the media, repeated by those around you who have only heard the same story, is it any wonder that people on a grand scale believe absolute mendacious garbage to be true? You are about to see, too, why this level of information control is necessary when the official ‘Covid’ narrative is so nonsensical and unsupportable by the evidence.

Structure of Deceit

The pyramid structure through which the ‘Covid’ hoax has been manifested is very simple and has to be to work. As few people as possible have to be involved with full knowledge of what they are doing – and why – or the real story would get out. At the top of the pyramid are the inner core of the Cult which controls Bill Gates who, in turn, controls the World Health Organization through his pivotal funding and his puppet Director-General mouthpiece, Tedros. Before he was appointed Tedros was chair of the Gates-founded Global Fund to ‘fight against AIDS, tuberculosis and malaria’, a board member of the Gates-funded ‘vaccine alliance’ GAVI, and on the board of another Gates-funded organisation. Gates owns him and picked him for a specific reason – Tedros is a crook and worse. ‘Dr’ Tedros (he’s not a medical doctor, the first WHO chief not to be) was a member of the tyrannical Marxist government of Ethiopia for decades with all its human rights abuses. He has faced allegations of corruption and misappropriation of funds and was exposed three times for covering up cholera epidemics while Ethiopia’s health minister. Tedros appointed the mass-murdering genocidal Zimbabwe dictator Robert Mugabe as a WHO goodwill ambassador for public health which, as with Tedros, is like appointing a psychopath to run a peace and love campaign. The move was so ridiculous that he had to drop Mugabe in the face of widespread condemnation. American economist David Steinman, a Nobel peace prize nominee, lodged a complaint with the International Criminal Court in The Hague over alleged genocide by Tedros when he was Ethiopia’s foreign minister. Steinman says Tedros was a ‘crucial decision maker’ who directed the actions of Ethiopia’s security forces from 2013 to 2015 and one of three officials in charge when those security services embarked on the ‘killing’ and ‘torturing’ of Ethiopians. You can see where Tedros is coming from and it’s sobering to think that he has been the vehicle for Gates and the Cult to direct the global response to ‘Covid’. Think about that. A psychopathic Cult dictates to psychopath Gates who dictates to psychopath Tedros who dictates how countries of the world must respond to a ‘Covid virus’ never scientifically shown to exist. At the same time psychopathic Cult-owned Silicon Valley information giants like Google, YouTube, Facebook and Twitter announced very early on that they would give the Cult/Gates/Tedros/WHO version of the narrative free advertising and censor those who challenged their intelligence-insulting, mendacious story.

The next layer in the global 'medical' structure below the Cult, Gates and Tedros are the chief medical officers and science 'advisers' in each of the WHO member countries which means virtually all of them. Medical officers and arbiters of science (they're not) then take the WHO policy and recommended responses and impose them on their country's population while the political 'leaders' say they are deciding policy (they're clearly not) by 'following the science' on the advice of the 'experts' – the same medical officers and science 'advisers' (dictators). In this way with the rarest of exceptions the entire world followed the same policy of lockdown, people distancing, masks and 'vaccines' dictated by the psychopathic Cult, psychopathic Gates and psychopathic Tedros who we are supposed to believe give a damn about the health of the world population they are seeking to enslave. That, amazingly, is all there is to it in terms of crucial decision-making. Medical staff in each country then follow like sheep the dictates of the shepherds at the top of the national medical hierarchies – chief medical officers and science 'advisers' who themselves follow like sheep the shepherds of the World Health Organization and the Cult. Shepherds at the national level often have major funding and other connections to Gates and his Bill and Melinda Gates Foundation which carefully hands out money like confetti at a wedding to control the entire global medical system from the WHO down.

Follow the money

Christopher Whitty, Chief Medical Adviser to the UK Government at the centre of 'virus' policy, a senior adviser to the government's Scientific Advisory Group for Emergencies (SAGE), and Executive Board member of the World Health Organization, was gifted a grant of $40 million by the Bill and Melinda Gates Foundation for malaria research in Africa. The BBC described the unelected Whitty as 'the official who will probably have the greatest impact on our everyday lives of any individual policymaker in modern times' and so it turned out. What Gates and Tedros have said Whitty has done like his equivalents around the world. Patrick Vallance, co-chair of SAGE and the government's Chief Scientific Adviser, is a former executive of Big Pharma giant GlaxoSmithKline with its fundamental financial and business connections to Bill Gates. In September, 2020, it was revealed that Vallance owned a deferred bonus of shares in GlaxoSmithKline worth £600,000 while the company was 'developing' a 'Covid vaccine'. Move along now – nothing to see here – what could possibly be wrong with that? Imperial College in London, a major player in 'Covid' policy in Britain and elsewhere with its 'Covid-19' Response Team, is funded by Gates and has big connections to China while the now infamous Professor Neil Ferguson, the useless 'computer modeller' at Imperial College is also funded by Gates. Ferguson delivered the dramatically inaccurate excuse for the first lockdowns (much more in the next chapter). The Institute for Health Metrics and Evaluation (IHME) in the United States, another source of outrageously false 'Covid' computer models to justify lockdowns, is bankrolled by Gates who is a vehement promotor of lockdowns. America's version of Whitty and Vallance, the again now infamous Anthony Fauci, has connections to 'Covid vaccine' maker Moderna as does Bill Gates through funding from the Bill and Melinda Gates Foundation. Fauci is director of the National Institute of Allergy and Infectious Diseases (NIAID), a major recipient of Gates money, and they are very close. Deborah Birx who was appointed White House Coronavirus Response Coordinator in February, 2020, is yet another with ties to Gates. Everywhere you look at the different elements around the world behind the coordination and decision making of the 'Covid' hoax there is Bill Gates and his money. They include the World Health Organization; Centers for Disease Control (CDC) in the United States; National Institutes of Health (NIH) of Anthony Fauci; Imperial College and Neil Ferguson; the London School of Hygiene where Chris Whitty worked; Regulatory agencies like the UK Medicines & Healthcare products Regulatory Agency (MHRA) which gave emergency approval for ‘Covid vaccines’; Wellcome Trust; GAVI, the Vaccine Alliance; the Coalition for Epidemic Preparedness Innovations (CEPI); Johns Hopkins University which has compiled the false ‘Covid’ figures; and the World Economic Forum. A Nationalfile.com article said:

> Gates has a lot of pull in the medical world, he has a multi-million dollar relationship with Dr. Fauci, and Fauci originally took the Gates line supporting vaccines and casting doubt on [the drug hydroxychloroquine]. Coronavirus response team member Dr. Deborah Birx, appointed by former president Obama to serve as United States Global AIDS Coordinator, also sits on the board of a group that has received billions from Gates’ foundation, and Birx reportedly used a disputed Bill Gates-funded model for the White House’s Coronavirus effort. Gates is a big proponent for a population lockdown scenario for the Coronavirus outbreak.

Another funder of Moderna is the Defense Advanced Research Projects Agency (DARPA), the technology-development arm of the Pentagon and one of the most sinister organisations on earth. DARPA had a major role with the CIA covert technology-funding operation In-Q-Tel in the development of Google and social media which is now at the centre of global censorship. Fauci and Gates are extremely close and openly admit to talking regularly about ‘Covid’ policy, but then why wouldn’t Gates have a seat at every national ‘Covid’ table after his Foundation committed $1.75 billion to the ‘fight against Covid-19’. When passed through our Orwellian Translation Unit this means that he has bought and paid for the Cult-driven ‘Covid’ response worldwide. Research the major ‘Covid’ response personnel in your own country and you will find the same Gates funding and other connections again and again. Medical and science chiefs following World Health Organization ‘policy’ sit atop a medical hierarchy in their country of administrators, doctors and nursing staff. These ‘subordinates’ are told they must work and behave in accordance with the policy delivered from the ‘top’ of the national ‘health’ pyramid which is largely the policy delivered by the WHO which is the policy delivered by Gates and the Cult. The whole ‘Covid’ narrative has been imposed on medical staff by a climate of fear although great numbers don’t even need that to comply. They do so through breathtaking levels of ignorance and include doctors who go through life simply repeating what Big Pharma and their hierarchical masters tell them to say and believe. No wonder Big Pharma 'medicine' is one of the biggest killers on Planet Earth.

The same top-down system of intimidation operates with regard to the Cult Big Pharma cartel which also dictates policy through national and global medical systems in this way. The Cult and Big Pharma agendas are the same because the former controls and owns the latter. 'Health' administrators, doctors, and nursing staff are told to support and parrot the dictated policy or they will face consequences which can include being fired. How sad it's been to see medical staff meekly repeating and imposing Cult policy without question and most of those who can see through the deceit are only willing to speak anonymously off the record. They know what will happen if their identity is known. This has left the courageous few to expose the lies about the 'virus', face masks, overwhelmed hospitals that aren't, and the dangers of the 'vaccine' that isn't a vaccine. When these medical professionals and scientists, some renowned in their field, have taken to the Internet to expose the truth their articles, comments and videos have been deleted by Cult-owned Facebook, Twitter and YouTube. What a real head-shaker to see YouTube videos with leading world scientists and highly qualified medical specialists with an added link underneath to the notorious Cult propaganda website *Wikipedia* to find the 'facts' about the same subject.

HIV – the 'Covid' trial-run

I'll give you an example of the consequences for health and truth that come from censorship and unquestioning belief in official narratives. The story was told by PCR inventor Kary Mullis in his book *Dancing Naked in the Mind Field*. He said that in 1984 he accepted as just another scientific fact that Luc Montagnier of France's Pasteur Institute and Robert Gallo of America's National Institutes of Health had independently discovered that a 'retrovirus' dubbed HIV (human immunodeficiency virus) caused AIDS. They were, after all, Mullis writes, specialists in retroviruses. This is how the medical and science pyramids work. Something is announced or *assumed* and then becomes an everybody-knows-that purely through repetition of the assumption as if it is fact. Complete crap becomes accepted truth with no supporting evidence and only repetition of the crap. This is how a 'virus' that doesn't exist became the 'virus' that changed the world. The HIV-AIDS fairy story became a multi-billion pound industry and the media poured out propaganda terrifying the world about the deadly HIV 'virus' that caused the lethal AIDS. By then Mullis was working at a lab in Santa Monica, California, to detect retroviruses with his PCR test in blood donations received by the Red Cross. In doing so he asked a virologist where he could find a reference for HIV being the cause of AIDS. 'You don't need a reference,' the virologist said … *'Everybody knows it.'* Mullis said he wanted to quote a reference in the report he was doing and he said he felt a little funny about not knowing the source of such an important discovery when everyone else seemed to. The virologist suggested he cite a report by the Centers for Disease Control and Prevention (CDC) on morbidity and mortality. Mullis read the report, but it only said that an organism had been identified and did not say how. The report did not identify the original scientific work. Physicians, however, *assumed* (key recurring theme) that if the CDC was convinced that HIV caused AIDS then proof must exist. Mullis continues:

> I did computer searches. Neither Montagnier, Gallo, nor anyone else had published papers describing experiments which led to the conclusion that HIV probably caused AIDS. I read the papers in Science for which they had become well known as AIDS doctors, but all they had said there was that they had found evidence of a past infection by something which was probably HIV in some AIDS patients.
>
> They found antibodies. Antibodies to viruses had always been considered evidence of past disease, not present disease. Antibodies signaled that the virus had been defeated. The patient had saved himself. There was no indication in these papers that this virus caused a disease. They didn't show that everybody with the antibodies had the disease. In fact they found some healthy people with antibodies.

Mullis asked why their work had been published if Montagnier and Gallo hadn't really found this evidence, and why had they been fighting so hard to get credit for the discovery? He says he was hesitant to write 'HIV is the probable cause of AIDS' until he found published evidence to support that. 'Tens of thousands of scientists and researchers were spending billions of dollars a year doing research based on this idea,' Mullis writes. 'The reason had to be there somewhere; otherwise these people would not have allowed their research to settle into one narrow channel of investigation.' He said he lectured about PCR at numerous meetings where people were always talking about HIV and he asked them how they knew that HIV was the cause of AIDS:

> Everyone said something. Everyone had the answer at home, in the office, in some drawer. They all knew, and they would send me the papers as soon as they got back. But I never got any papers. Nobody ever sent me the news about how AIDS was caused by HIV.

Eventually Mullis was able to ask Montagnier himself about the reference proof when he lectured in San Diego at the grand opening of the University of California AIDS Research Center. Mullis says this was the last time he would ask his question without showing anger. Montagnier said he should reference the CDC report. 'I read it', Mullis said, and it didn't answer the question. 'If Montagnier didn't know the answer who the hell did?' Then one night Mullis was driving when an interview came on National Public Radio with Peter Duesberg, a prominent virologist at Berkeley and a California Scientist of the Year. Mullis says he finally understood why he could not find references that connected HIV to AIDS – *there weren't any!* No one had ever proved that HIV causes AIDS even though it had spawned a multi-billion pound global industry and the media was repeating this as fact every day in their articles and broadcasts terrifying the shit out of people about AIDS and giving the impression that a positive test for HIV (see 'Covid') was a death sentence. Duesberg was a threat to the AIDS gravy train and the agenda that underpinned it. He was therefore abused and castigated after he told the Proceedings of the National Academy of Sciences there was no good evidence implicating the new ‘virus’. Editors rejected his manuscripts and his research funds were deleted. Mullis points out that the CDC has defined AIDS as one of more than 30 diseases *if accompanied* by a positive result on a test that detects antibodies to HIV; but those same diseases are not defined as AIDS cases when antibodies are not detected:

> If an HIV-positive woman develops uterine cancer, for example, she is considered to have AIDS. If she is not HIV positive, she simply has uterine cancer. An HIV-positive man with tuberculosis has AIDS; if he tests negative he simply has tuberculosis. If he lives in Kenya or Colombia, where the test for HIV antibodies is too expensive, he is simply presumed to have the antibodies and therefore AIDS, and therefore he can be treated in the World Health Organization’s clinic. It’s the only medical help available in some places. And it’s free, because the countries that support WHO are worried about AIDS.

Mullis accuses the CDC of continually adding new diseases (see ever more ‘Covid symptoms’) to the grand AIDS definition and of virtually doctoring the books to make it appear as if the disease continued to spread. He cites how in 1993 the CDC enormously broadened its AIDS definition and county health authorities were delighted because they received $2,500 per year from the Federal government for every reported AIDS case. Ladies and gentlemen, I have just described, via Kary Mullis, the ‘Covid pandemic’ of 2020 and beyond. Every element is the same and it’s been pulled off in the same way by the same networks.

The ‘Covid virus’ exists? Okay – prove it. Er ... still waiting

What Kary Mullis described with regard to ‘HIV’ has been repeated with ‘Covid’. A claim is made that a new, or ‘novel’, infection has been found and the entire medical system of the world repeats that as fact exactly as they did with HIV and AIDS. No one in the mainstream asks rather relevant questions such as ‘How do you know?’ and ‘Where is your proof?’ The SARS-Cov-2 ‘virus’ and the ‘Covid-19 disease’ became an overnight ‘everybody-knows-that’. The origin could be debated and mulled over, but what you could not suggest was that ‘SARS-Cov-2’ didn’t exist. That would be ridiculous. ‘Everybody knows’ the ‘virus’ exists. Well, I didn’t for one along with American proper doctors like Andrew Kaufman and Tom Cowan and long-time American proper journalist Jon Rappaport. We dared to pursue the obvious and simple question: ‘Where’s the evidence?’ The overwhelming majority in medicine, journalism and the general public did not think to ask that. After all, *everyone knew* there was a new ‘virus’. Everyone was saying so and I heard it on the BBC. Some would eventually argue that the ‘deadly virus’ was nothing like as deadly as claimed, but few would venture into the realms of its very existence. Had they done so they would have found that the evidence for that claim had gone AWOL as with HIV causes AIDS. In fact, not even that. For something to go AWOL it has to exist in the first place and scientific proof for a ‘SARS-Cov-2’ can be filed under nothing, nowhere and zilch.

Dr Andrew Kaufman is a board-certified forensic psychiatrist in New York State, a Doctor of Medicine and former Assistant Professor and Medical Director of Psychiatry at SUNY Upstate Medical University, and Medical Instructor of Hematology and Oncology at the Medical School of South Carolina. He also studied biology at the Massachusetts Institute of Technology (MIT) and trained in Psychiatry at Duke University. Kaufman is retired from allopathic medicine, but remains a consultant and educator on natural healing, I saw a video of his very early on in the ‘Covid’ hoax in which he questioned claims about the ‘virus’ in the absence of any supporting evidence and with plenty pointing the other way. I did everything I could to circulate his work which I felt was asking the pivotal questions that needed an answer. I can recommend an excellent pull-together interview he did with the website The Last Vagabond entitled *Dr Andrew Kaufman: Virus Isolation, Terrain Theory and Covid-19* and his website is andrewkaufmanmd.com. Kaufman is not only a forensic psychiatrist; he is forensic in all that he does. He always reads original scientific papers, experiments and studies instead of second-third-fourth-hand reports about the ‘virus’ in the media which are repeating the repeated repetition of the narrative. When he did so with the original Chinese ‘virus’ papers Kaufman realised that there was no evidence of a ‘SARS-Cov-2’. They had never – from the start – shown it to exist and every repeat of this claim worldwide was based on the accepted existence of proof that was nowhere to be found – see Kary Mullis and HIV. Here we go again.

Let’s postulate

Kaufman discovered that the Chinese authorities immediately concluded that the cause of an illness that broke out among about 200 initial patients in Wuhan was a ‘new virus’ when there were no grounds to make that conclusion. The alleged ‘virus’ was not isolated from other genetic material in their samples and then shown through a system known as Koch’s postulates to be the causative agent of the illness. The world was told that the SARS-Cov-2 ‘virus’ caused a disease they called ‘Covid-19’ which had ‘flu-like’ symptoms and could lead to respiratory problems and pneumonia. If it wasn’t so tragic it would almost be funny. *‘Flu-like’ symptoms’? Pneumonia? Respiratory disease?* What in *CHINA* and particularly in *Wuhan*, one of the most polluted cities in the world with a resulting epidemic of respiratory disease?? Three hundred thousand people get pneumonia in China every year and there are nearly a billion cases worldwide of ‘flu-like symptoms’. These have a whole range of causes – including pollution in Wuhan – but no other possibility was credibly considered in late 2019 when the world was told there was a new and deadly ‘virus’. The global prevalence of pneumonia and ‘flu-like systems’ gave the Cult networks unlimited potential to re-diagnose these other causes as the mythical ‘Covid-19’ and that is what they did from the very start. Kaufman revealed how Chinese medical and science authorities (all subordinates to the Cult-owned communist government) took genetic material from the lungs of only a few of the first patients. The material contained their own cells, bacteria, fungi and other microorganisms living in their bodies. The only way you could prove the existence of the ‘virus’ and its responsibility for the alleged ‘Covid-19’ was to isolate the virus from all the other material – a process also known as ‘purification’ – and then follow the postulates sequence developed in the late 19th century by German physician and bacteriologist Robert Koch which became the 'gold standard' for connecting an alleged causation agent to a disease:

1. The microorganism (bacteria, fungus, virus, etc.) must be present in every case of the disease and all patients must have the same symptoms. It must also *not be present in healthy individuals*.

2. The microorganism must be isolated from the host with the disease. If the microorganism is a bacteria or fungus it must be grown in a pure culture. If it is a virus, it must be purified (i.e. containing no other material except the virus particles) from a clinical sample.

3. The specific disease, with all of its characteristics, must be reproduced when the infectious agent (the purified virus or a pure culture of bacteria or fungi) is inoculated into a healthy, susceptible host.

4. The microorganism must be recoverable from the experimentally infected host as in step 2.

Not one of these criteria has been met in the case of 'SARS-Cov-2' and 'Covid-19'. Not ONE. *EVER*. Robert Koch refers to bacteria and not viruses. What are called 'viral particles' are so minute (hence masks are useless by any definition) that they could only be seen after the invention of the electron microscope in the 1930s and can still only be observed through that means. American bacteriologist and virologist Thomas Milton Rivers, the so-called 'Father of Modern Virology' who was very significantly director of the Rockefeller Institute for Medical Research in the 1930s, developed a less stringent version of Koch's postulates to identify 'virus' causation known as 'Rivers criteria'. 'Covid' did not pass that process either. Some even doubt whether any 'virus' can be isolated from other particles containing genetic material in the Koch method. Freedom of Information requests in many countries asking for scientific proof that the 'Covid virus' has been purified and isolated and shown to exist have all come back with a 'we don't have that' and when this happened with a request to the UK Department of Health they added this comment:

> However, outside of the scope of the [Freedom of Information Act] and on a discretionary basis, the following information has been advised to us, which may be of interest. Most infectious diseases are caused by viruses, bacteria or fungi. Some bacteria or fungi have the capacity to grow on their own in isolation, for example in colonies on a petri dish. Viruses are different in that they are what we call 'obligate pathogens' – that is, they cannot survive or reproduce without infecting a host ...
>
> … For some diseases, it is possible to establish causation between a microorganism and a disease by isolating the pathogen from a patient, growing it in pure culture and reintroducing it to a healthy organism. These are known as 'Koch's postulates' and were developed in 1882. However, as our understanding of disease and different disease-causing agents has advanced, these are no longer the method for determining causation [Andrew Kaufman asks why in that case are there two published articles falsely claiming to satisfy Koch's postulates].
>
> It has long been known that viral diseases cannot be identified in this way as viruses cannot be grown in 'pure culture'. When a patient is tested for a viral illness, this is normally done by looking for the presence of antigens, or viral genetic code in a host with molecular biology techniques [Kaufman asks how you could know the origin of these chemicals without having a pure culture for comparison].

> For the record 'antigens' are defined so:
>
> Invading microorganisms have antigens on their surface that the human body can recognise as being foreign – meaning not belonging to it. When the body recognises a foreign antigen, lymphocytes (white blood cells) produce antibodies, which are complementary in shape to the antigen.

Notwithstanding that this is open to question in relation to 'SARS-Cov-2' the presence of 'antibodies' can have many causes and they are found in people that are perfectly well. Kary Mullis said: 'Antibodies … had always been considered evidence of past disease, not present disease.'

'Covid' really is a *computer* 'virus'

Where the UK Department of Health statement says 'viruses' are now 'diagnosed' through a 'viral genetic code in a host with molecular biology techniques', they mean … *the PCR test* which its inventor said cannot test for infectious disease. They have no credible method of connecting a 'virus' to a disease and we will see that there is no scientific proof that any 'virus' causes any disease or there is any such thing as a 'virus' in the way that it is described. Tenacious Canadian researcher Christine Massey and her team made some 40 Freedom of Information requests to national public health agencies in different countries asking for proof that SARS-CoV-2 has been isolated and not one of them could supply that information. Massey said of her request in Canada: ‘Freedom of Information reveals Public Health Agency of Canada has no record of ‘SARS-COV-2’ isolation performed by anyone, anywhere, ever.’ If you accept the comment from the UK Department of Health it’s because they can’t isolate a ‘virus’. Even so many ‘science’ papers claimed to have isolated the ‘Covid virus’ until they were questioned and had to admit they hadn’t. A reply from the Robert Koch Institute in Germany was typical: ‘I am not aware of a paper which purified isolated SARS-CoV-2.’ So what the hell was Christian Drosten and his gang using to design the ‘Covid’ testing protocol that has produced all the illusory Covid’ cases and ‘Covid’ deaths when the head of the Chinese version of the CDC admitted there was a problem right from the start in that the ‘virus’ had never been isolated/purified? Breathe deeply: What they are calling ‘Covid’ is actually created by a *computer program* i.e. *they made it up* – er, that’s it. They took lung fluid, with many sources of genetic material, from one single person alleged to be infected with Covid-19 by a PCR test which they *claimed*, without clear evidence, contained a ‘virus’. They used several computer programs to create a model of a theoretical virus genome sequence from more than fifty-six million small sequences of RNA, each of an unknown source, assembling them like a puzzle with no known solution. The computer filled in the gaps with sequences from bits in the gene bank to make it look like a bat SARS-like coronavirus! A wave of the magic wand and poof, an in silico (computer-generated) genome, a scientific fantasy, was created. UK health researcher Dr Kevin Corbett made the same point with this analogy:

> … It’s like giving you a few bones and saying that’s your fish. It could be any fish. Not even a skeleton. Here’s a few fragments of bones. That’s your fish … It’s all from gene bank and the bits of the virus sequence that weren’t there they made up.
>
> They synthetically created them to fill in the blanks. That’s what genetics is; it’s a code. So it’s ABBBCCDDD and you’re missing some what you think is EEE so you put it in. It’s all

> synthetic. You just manufacture the bits that are missing. This is the end result of the geneticization of virology. This is basically a computer virus.

Further confirmation came in an email exchange between British citizen journalist Frances Leader and the government's Medicines & Healthcare Products Regulatory Agency (the Gates-funded MHRA) which gave emergency permission for untested 'Covid vaccines' to be used. The agency admitted that the 'vaccine' is not based on an isolated 'virus', but comes from a *computer-generated model*. Frances Leader was naturally banned from Cult-owned fascist Twitter for making this exchange public. The process of creating computer-generated alleged 'viruses' is called 'in silico' or 'in silicon' – computer chips – and the term 'in silico' is believed to originate with biological experiments using only a computer in 1989. 'Vaccines' involved with 'Covid' are also produced 'in silico' or by computer not a natural process. If the original 'virus' is nothing more than a made-up computer model how can there be 'new variants' of something that never existed in the first place? They are not new 'variants'; they are new *computer models* only minutely different to the original program and designed to further terrify the population into having the 'vaccine' and submitting to fascism. You want a 'new variant'? Click, click, enter – there you go. Tell the medical profession that you have discovered a 'South African variant', 'UK variants' or a 'Brazilian variant' and in the usual HIV-causes-AIDS manner they will unquestioningly repeat it with no evidence whatsoever to support these claims. They will go on television and warn about the dangers of 'new variants' while doing nothing more than repeating what they have been told to be true and knowing that any deviation from that would be career suicide. Big-time insiders will know it's a hoax, but much of the medical community is clueless about the way they are being played and themselves play the public without even being aware they are doing so. What an interesting 'coincidence' that AstraZeneca and Oxford University were conducting 'Covid vaccine trials' in the three countries – the UK, South Africa and Brazil – where the first three 'variants' were claimed to have 'broken out'.

Here’s your ‘virus’ – it’s a unicorn

Dr Andrew Kaufman presented a brilliant analysis describing how the ‘virus’ was imagined into fake existence when he dissected an article published by *Nature* and written by 19 authors detailing *alleged* ‘sequencing of a complete viral genome’ of the ‘new SARS-CoV-2 virus’. This computer-modelled *in silico* genome was used as a template for all subsequent genome sequencing experiments that resulted in the so-called variants which he said now number more than *6,000*. The fake genome was constructed from more than 56 million individual short strands of RNA. Those little pieces were assembled into longer pieces by finding areas of overlapping sequences. The computer programs created over two million possible combinations from which the authors simply chose the longest one. They then compared this to a ‘bat virus’ and the computer ‘alignment’ rearranged the sequence and filled in the gaps! They called this computer-generated abomination the ‘complete genome’. Dr Tom Cowan, a fellow medical author and collaborator with Kaufman, said such computer-generation constitutes scientific fraud and he makes this superb analogy:

> Here is an equivalency: A group of researchers claim to have found a unicorn because they found a piece of a hoof, a hair from a tail, and a snippet of a horn. They then add that information into a computer and program it to re-create the unicorn, and they then claim this computer re-creation is the real unicorn. Of course, they had never actually seen a unicorn so could not possibly have examined its genetic makeup to compare their samples with the actual unicorn’s hair, hooves and horn.
>
> The researchers claim they decided which is the real genome of SARS-CoV-2 by ‘consensus’, sort of like a vote. Again, different computer programs will come up with different versions of the imaginary ‘unicorn’, so they come together as a group and decide which is the real imaginary unicorn.

This is how the ‘virus’ that has transformed the world was brought into fraudulent ‘existence’. Extraordinary, yes, but as the Nazis said the bigger the lie the more will believe it. Cowan, however, wasn’t finished and he went on to identify what he called the real blockbuster in the paper. He quotes this section from a paper written by virologists and published by the CDC and then explains what it means:

> Therefore, we examined the capacity of SARS-CoV-2 to infect and replicate in several common primate and human cell lines, including human adenocarcinoma cells (A549), human liver cells (HUH 7.0), and human embryonic kidney cells (HEK-293T). In addition to Vero E6 and Vero CCL81 cells. ... Each cell line was inoculated at high multiplicity of infection and examined 24h post-infection.
>
> No CPE was observed in any of the cell lines except in Vero cells, which grew to greater than 10 to the 7th power at 24 h post-infection. In contrast, HUH 7.0 and 293T showed only modest viral replication, and A549 cells were incompatible with SARS CoV-2 infection.

Cowan explains that when virologists attempt to prove infection they have three possible 'hosts' or models on which they can test. The first was humans. Exposure to humans was generally not done for ethical reasons and has never been done with SARS-CoV-2 or any coronavirus. The second possible host was animals. Cowan said that forgetting for a moment that they never actually use purified virus when exposing animals they do use solutions that they *claim* contain the virus. Exposure to animals has been done with SARS-CoV-2 in an experiment involving mice and this is what they found: *None of the wild (normal) mice got sick*. In a group of genetically-modified mice, a statistically insignificant number lost weight and had slightly bristled fur, but they experienced nothing like the illness called 'Covid-19'. Cowan said the third method – the one they mostly rely on – is to inoculate solutions they *say* contain the virus onto a variety of tissue cultures. This process had never been shown to kill tissue *unless* the sample material was starved of nutrients and poisoned as *part of the process*. Yes, incredibly, in tissue experiments designed to show the 'virus' is responsible for killing the tissue they starve the tissue of nutrients and add toxic drugs including antibiotics and they do not have control studies to see if it's the starvation and poisoning that is degrading the tissue rather than the 'virus' they allege to be in there somewhere. You want me to pinch you? Yep, I understand. Tom Cowan said this about the whole nonsensical farce as he explains what that quote from the CDC paper really means:

> The shocking thing about the above quote is that using their own methods, the virologists found that solutions containing SARS-CoV-2 – even in high amounts – were NOT, I repeat NOT, infective to any of the three human tissue cultures they tested. In plain English, this means they proved, on their terms, that this 'new coronavirus' is not infectious to human beings. It is ONLY infective to monkey kidney cells, and only then when you add two potent drugs (gentamicin and amphotericin), known to be toxic to kidneys, to the mix.
>
> My friends, read this again and again. These virologists, published by the CDC, performed a clear proof, on their terms, showing that the SARS-CoV-2 virus is harmless to human beings. That is the only possible conclusion, but, unfortunately, this result is not even mentioned in their conclusion. They simply say they can provide virus stocks cultured only on monkey Vero cells, thanks for coming.

Cowan concluded: 'If people really understood how this "science" was done, I would hope they would storm the gates and demand honesty, transparency and truth.' Dr Michael Yeadon, former Vice President and Chief Scientific Adviser at drug giant Pfizer has been a vocal critic of the 'Covid vaccine' and its potential for multiple harm. He said in an interview in April, 2021, that 'not one [vaccine] has the virus. He was asked why vaccines normally using a 'dead' version of a disease to activate the immune system were not used for 'Covid' and instead we had the synthetic methods of the 'mRNA Covid vaccine'. Yeadon said that to do the former 'you'd have to have some of [the virus] wouldn't you?' He added: 'No-one's got any – seriously.' Yeadon said that surely they couldn't have fooled the whole world for a year without having a virus, 'but oddly enough ask around – no one's got it'. He didn't know why with all the 'great labs' around the world that the virus had not been isolated – 'Maybe they've been too busy running bad PCR tests and vaccines that people don't need.' What is today called 'science' is not 'science' at all. Science is no longer what is, but whatever people can be manipulated to *believe* that it is. Real science has been hijacked by the Cult to dispense and produce the 'expert scientists' and contentions that suit the agenda of the Cult. How big-time this has happened with the 'Covid' hoax which is entirely based on fake science delivered by fake 'scientists' and fake 'doctors'. The human-caused climate change hoax is also entirely based on fake science delivered by fake 'scientists' and fake 'climate experts'. In both cases real scientists, climate experts and doctors have their views suppressed and deleted by the Cult-owned science establishment, media and Silicon Valley. This is the ‘science’ that politicians claim to be ‘following’ and a common denominator of ‘Covid’ and climate are Cult psychopaths Bill Gates and his mate Klaus Schwab at the Gates-funded World Economic Forum. But, don’t worry, it’s all just a coincidence and absolutely nothing to worry about. *Zzzzzzzz*.

What is a ‘virus’ REALLY?

Dr Tom Cowan is one of many contesting the very existence of viruses let alone that they cause disease. This is understandable when there is no scientific evidence for a disease-causing ‘virus’. German virologist Dr Stefan Lanka won a landmark case in 2017 in the German Supreme Court over his contention that there is no such thing as a measles virus. He had offered a big prize for anyone who could prove there is and Lanka won his case when someone sought to claim the money. There is currently a prize of more than 225,000 euros on offer from an Isolate Truth Fund for anyone who can prove the isolation of SARS-CoV-2 and its genetic substance. Lanka wrote in an article headed ‘The Misconception Called Virus’ that scientists think a ‘virus’ is causing tissue to become diseased and degraded when in fact it is the *processes they are using* which do that – not a ‘virus’. Lanka has done an important job in making this point clear as Cowan did in his analysis of the CDC paper. Lanka says that all claims about viruses as disease-causing pathogens are wrong and based on ‘easily recognisable, understandable and verifiable misinterpretations.’ Scientists believed they were working with ‘viruses’ in their laboratories when they were really working with ‘typical particles of specific dying tissues or cells …’ Lanka said that the tissue decaying process claimed to be caused by a ‘virus’ still happens when no alleged ‘virus’ is involved. It’s the *process* that does the damage and not a ‘virus’. The genetic sample is deprived of nutrients, removed from its energy supply through removal from the body and then doused in toxic antibiotics to remove any bacteria. He confirms again that establishment scientists do not (pinch me) conduct control experiments to see if this is the case and if they did they would see the claims that ‘viruses’ are doing the damage is nonsense. He adds that during the measles ‘virus’ court case he commissioned an independent laboratory to perform just such a control experiment and the result was that the tissues and cells died in the exact same way as with alleged ‘infected’ material. This is supported by a gathering number of scientists, doctors and researchers who reject what is called ‘germ theory’ or the belief in the body being infected by contagious sources emitted by other people. Researchers Dawn Lester and David Parker take the same stance in their highly-detailed and sourced book *What Really Makes You Ill – Why everything you thought you knew about disease is wrong* which was recommended to me by a number of medical professionals genuinely seeking the truth. Lester and Parker say there is no provable scientific evidence to show that a ‘virus’ can be transmitted between people or people and animals or animals and people:

> The definition also claims that viruses are the cause of many diseases, as if this has been definitively proven. But this is not the case; there is no original scientific evidence that definitively demonstrates that any virus is the cause of any disease. The burden of proof for any theory lies with those who proposed it; but none of the existing documents provides ‘proof’ that supports the claim that ‘viruses’ are pathogens.

Dr Tom Cowan employs one of his clever analogies to describe the process by which a ‘virus’ is named as the culprit for a disease when what is called a ‘virus’ is only material released by cells detoxing themselves from infiltration by chemical or radiation poisoning. The tidal wave of technologically-generated radiation in the ‘smart’ modern world plus all the toxic food and drink are causing this to happen more than ever. Deluded ‘scientists’ misread this as a gathering impact of what they wrongly label ‘viruses’.

Paper can infect houses

Cowan said in an article for davidicke.com – with his tongue only mildly in his cheek – that he believed he had made a tremendous discovery that may revolutionise science. He had discovered that small bits of paper are alive, ‘well alive-ish’, can ‘infect’ houses, and then reproduce themselves inside the house. The result was that this explosion of growth in the paper inside the house causes the house to explode, blowing it to smithereens. His evidence for this new theory is that in the past months he had carefully examined many of the houses in his neighbourhood and found almost no scraps of paper on the lawns and surrounds of the house. There was an occasional stray label, but nothing more. Then he would return to these same houses a week or so later and with a few, not all of them, particularly the old and decrepit ones, he found to his shock and surprise they were littered with stray bits of paper. He knew then that the paper had infected these houses, made copies of itself, and blew up the house. A young boy on a bicycle at one of the sites told him he had seen a demolition crew using dynamite to explode the house the previous week, but Cowan dismissed this as the idle thoughts of silly boys because ‘I was on to something big’. He was on to how ‘scientists’ mistake genetic material in the detoxifying process for something they call a ‘virus’. Cowan said of his house and paper story:

> If this sounds crazy to you, it’s because it should. This scenario is obviously nuts. But consider this admittedly embellished, for effect, current viral theory that all scientists, medical doctors and virologists currently believe.

He takes the example of the ‘novel SARS-Cov2’ virus to prove the point. First they take someone with an undefined illness called ‘Covid-19’ and don’t even attempt to find any virus in their sputum. Never mind the scientists still describe how this ‘virus’, which they have not located attaches to a cell receptor, injects its genetic material, in ‘Covid’s’ case, RNA, into the cell. The RNA once inserted exploits the cell to reproduce itself and makes ‘thousands, nay millions, of copies of itself … Then it emerges victorious to claim its next victim’:

If you were to look in the scientific literature for proof, actual scientific proof, that uniform SARS-CoV2 viruses have been properly isolated from the sputum of a sick person, that actual spike proteins could be seen protruding from the virus (which has not been found), you would find that such evidence doesn't exist.

If you go looking in the published scientific literature for actual pictures, proof, that these spike proteins or any viral proteins are ever attached to any receptor embedded in any cell membrane, you would also find that no such evidence exists. If you were to look for a video or documented evidence of the intact virus injecting its genetic material into the body of the cell, reproducing itself and then emerging victorious by budding off the cell membrane, you would find that no such evidence exists.

The closest thing you would find is electron micrograph pictures of cellular particles, possibly attached to cell debris, both of which to be seen were stained by heavy metals, a process that completely distorts their architecture within the living organism. This is like finding bits of paper stuck to the blown-up bricks, thereby proving the paper emerged by taking pieces of the bricks on its way out.

The Enders baloney

Cowan describes the 'Covid' story as being just as make-believe as his paper story and he charts back this fantasy to a Nobel Prize winner called John Enders (1897-1985), an American biomedical scientist who has been dubbed 'The Father of Modern Vaccines'. Enders is claimed to have 'discovered' the process of the viral culture which 'proved' that a 'virus' caused measles. Cowan explains how Enders did this 'by using the EXACT same procedure that has been followed by every virologist to find and characterize every new virus since 1954'. Enders took throat swabs from children with measles and immersed them in 2ml of milk. Penicillin (100u/ml) and the antibiotic streptomycin (50,g/ml) were added and the whole mix was centrifuged – rotated at high speed to separate large cellular debris from small particles and molecules as with milk and cream, for example. Cowan says that if the aim is to find little particles of genetic material ('viruses') in the snot from children with measles it would seem that the last thing you would do is mix the snot with other material – milk –that also has genetic material. 'How are you ever going to know whether whatever you found came from the snot or the milk?' He points out that streptomycin is a 'nephrotoxic' or poisonous-to-the-kidney drug. You will see the relevance of that shortly. Cowan says that it gets worse, much worse, when Enders describes the culture medium upon which the virus 'grows': 'The culture medium consisted of bovine amniotic fluid (90%), beef embryo extract (5%), horse serum (5%), antibiotics and phenol red as an indicator of cell metabolism.' Cowan asks incredulously: 'Did he just say that the culture medium also contained fluids and tissues that are themselves rich sources of genetic material?' The genetic cocktail, or 'medium', is inoculated onto tissue and cells from rhesus monkey *kidney* tissue. This is where the importance of streptomycin comes in and currently-used antimicrobials and other drugs that are *poisonous to kidneys* and used in ALL modern viral cultures (e.g. gentamicin, streptomycin, and amphotericin). Cowan asks: 'How are you ever going to know from this witch's brew where any genetic material comes from as we now have five different sources of rich genetic material in our mix?' Remember, he says, that all genetic material, whether from monkey kidney tissues, bovine serum, milk, etc., is made from the exact same components. The same central question returns: 'How are you possibly going to know that it was the virus that killed the kidney tissue and not the toxic antibiotic and starvation rations on which you are growing the tissue?' John Enders answered the question himself – *you can't*:

> A second agent was obtained from an uninoculated culture of monkey kidney cells. The cytopathic changes [death of the cells] it induced in the unstained preparations could not be distinguished with confidence from the viruses isolated from measles.

The death of the cells ('cytopathic changes') happened in exactly the same manner, whether they inoculated the kidney tissue with the measles snot or not, Cowan says. 'This is evidence that the destruction of the tissue, the very proof of viral causation of illness, was not caused by anything in the snot because they saw the same destructive effect when the snot was not even used … the cytopathic, i.e., cell-killing, changes come from the process of the culture itself, not from any virus in any snot, period.' Enders quotes in his 1957 paper a virologist called Ruckle as reporting similar findings 'and in addition has isolated an agent from monkey kidney tissue that is so far indistinguishable from human measles virus'. In other words, Cowan says, these particles called 'measles viruses' are simply and clearly breakdown products of the starved and poisoned tissue. For measles 'virus' see all 'viruses' including the so-called 'Covid virus'. Enders, the 'Father of Modern Vaccines', also said:

> There is a potential risk in employing cultures of primate cells for the production of vaccines composed of attenuated virus, since the presence of other agents possibly latent in primate tissues cannot be definitely excluded by any known method.

Cowan further quotes from a paper published in the journal *Viruses* in May, 2020, while the 'Covid pandemic' was well underway in the media if not in reality. 'EVs' here refers to particles of genetic debris from our own tissues, such as exosomes of which more in a moment: 'The remarkable resemblance between EVs and viruses has caused quite a few problems in the studies focused on the analysis of EVs released during viral infections.' Later the paper adds that to date a reliable method that can actually guarantee a complete separation (of EVs from viruses) DOES NOT EXIST. This was published at a time when a fairy tale 'virus' was claimed in total certainty to be causing a fairy tale 'viral disease' called 'Covid-19' – a fairy tale that was already well on the way to transforming human society in the image that the Cult has worked to achieve for so long. Cowan concludes his article:

> To summarize, there is no scientific evidence that pathogenic viruses exist. What we think of as 'viruses' are simply the normal breakdown products of dead and dying tissues and cells. When we are well, we make fewer of these particles; when we are starved, poisoned, suffocated by wearing masks, or afraid, we make more.
>
> There is no engineered virus circulating and making people sick. People in laboratories all over the world are making genetically modified products to make people sick. These are called vaccines. There is no virome, no 'ecosystem' of viruses, viruses are not 8%, 50% or 100 % of our genetic material. These are all simply erroneous ideas based on the misconception called a virus.

What is 'Covid'? Load of bollocks

The background described here by Cowan and Lanka was emphasised in the first video presentation that I saw by Dr Andrew Kaufman when he asked whether the ‘Covid virus’ was in truth a natural defence mechanism of the body called ‘exosomes’. These are released by cells when in states of toxicity – see the same themes returning over and over. They are released ever more profusely as chemical and radiation toxicity increases and think of the potential effect therefore of 5G alone as its destructive frequencies infest the human energetic information field with a gathering pace (5G went online in Wuhan in 2019 as the ‘virus’ emerged). I’ll have more about this later. Exosomes transmit a warning to the rest of the body that ‘Houston, we have a problem’. Kaufman presented images of exosomes and compared them with ‘Covid’ under an electron microscope and the similarity was remarkable. They both attach to the same cell receptors (*claimed* in the case of ‘Covid’), contain the same genetic material in the form of RNA or ribonucleic acid, and both are found in ‘viral cell cultures’ with damaged or dying cells. James Hildreth MD, President and Chief Executive Officer of the Meharry Medical College at Johns Hopkins, said: ‘The virus is fully an exosome in every sense of the word.’ Kaufman’s conclusion was that there is no ‘virus’: ‘This entire pandemic is a completely manufactured crisis … there is no evidence of anyone dying from [this] illness.’ Dr Tom Cowan and Sally Fallon Morell, authors of *The Contagion Myth*, published a statement with Dr Kaufman in February, 2021, explaining why the ‘virus’ does not exist and you can read it that in full in the Appendix.

‘Virus’ theory can be traced to the ‘cell theory’ in 1858 of German physician Rudolf Virchow (1821-1920) who contended that disease originates from a single cell infiltrated by a ‘virus’. Dr Stefan Lanka said that findings and insights with respect to the structure, function and central importance of tissues in the creation of life, which were already known in 1858, comprehensively refute the cell theory. Virchow ignored them. We have seen the part later played by John Enders in the 1950s and Lanka notes that infection theories were only established as a global dogma through the policies and eugenics of the Third Reich in Nazi Germany (creation of the same Sabbatian cult behind the 'Covid' hoax). Lanka said: 'Before 1933, scientists dared to contradict this theory; after 1933, these critical scientists were silenced'. Dr Tom Cowan's view is that ill-heath is caused by too much of something, too little of something, or toxification from chemicals and radiation – not contagion. We must also highlight as a major source of the 'virus' theology a man still called the 'Father of Modern Virology' – Thomas Milton Rivers (1888-1962). There is no way given the Cult's long game policy that it was a coincidence for the 'Father of Modern Virology' to be director of the Rockefeller Institute for Medical Research from 1937 to 1956 when he is credited with making the Rockefeller Institute a leader in 'viral research'. Cult Rockefellers were the force behind the creation of Big Pharma 'medicine', established the World Health Organisation in 1948, and have long and close associations with the Gates family that now runs the WHO during the pandemic hoax through mega-rich Cult gofer and psychopath Bill Gates.

Only a Renegade Mind can see through all this bullshit by asking the questions that need to be answered, not taking 'no' or prevarication for an answer, and certainly not hiding from the truth in fear of speaking it. Renegade Minds have always changed the world for the better and they will change this one no matter how bleak it may currently appear to be.

CHAPTER SIX

Sequence of deceit

If you tell the truth, you don't have to remember anything
Mark Twain

Against the background that I have laid out this far the sequence that took us from an invented 'virus' in Cult-owned China in late 2019 to the fascist transformation of human society can be seen and understood in a whole new context.

We were told that a deadly disease had broken out in Wuhan and the world media began its campaign (coordinated by behavioural psychologists as we shall see) to terrify the population into unquestioning compliance. We were shown images of Chinese people collapsing in the street which never happened in the West with what was supposed to be the same condition. In the earliest days when alleged cases and deaths were few the fear register was hysterical in many areas of the media and this would expand into the common media narrative across the world. The real story was rather different, but we were never told that. The Chinese government, one of the Cult's biggest centres of global operation, said they had discovered a new illness with flu-like and pneumonia-type symptoms in a city with such toxic air that it is overwhelmed with flu-like symptoms, pneumonia and respiratory disease. Chinese scientists said it was a new – 'novel' – coronavirus which they called Sars-Cov-2 and that it caused a disease they labelled 'Covid-19'. There was no evidence for this and the 'virus' has never to this day been isolated, purified and its genetic code established from that. It was from the beginning a computer-generated fiction. Stories of Chinese whistleblowers saying the number of deaths was being supressed or that the 'new disease' was related to the Wuhan bio-lab misdirected mainstream and alternative media into cul-de-sacs to obscure the real truth – there was no 'virus'.

Chinese scientists took genetic material from the lung fluid of just a few people and said they had found a 'new' disease when this material had a wide range of content. There was no evidence for a 'virus' for the very reasons explained in the last two chapters. The 'virus' has never been shown to (a) exist and (b) cause any disease. People were diagnosed on symptoms that are so widespread in Wuhan and polluted China and with a PCR test that can't detect infectious disease. On this farce the whole global scam was sold to the rest of the world which would also diagnose respiratory disease as 'Covid-19' from symptoms alone or with a PCR test not testing for a 'virus'. Flu miraculously disappeared *worldwide* in 2020 and into 2021 as it was redesignated 'Covid-19'. It was really the same old flu with its 'flu-like' symptoms attributed to 'flu-like' 'Covid-19'. At the same time with very few exceptions the Chinese response of draconian lockdown and fascism was the chosen weapon to respond across the West as recommended by the Cult-owned Tedros at the Cult-owned World Health Organization run by the Cult-owned Gates. All was going according to plan. Chinese scientists – everything in China is controlled by the Cult-owned government – compared their contaminated RNA lung-fluid material with other RNA sequences and said it appeared to be just under 80 percent identical to the SARS-CoV-1 'virus' claimed to be the cause of the SARS (severe acute respiratory syndrome) 'outbreak' in 2003. They decreed that because of this the 'new virus' had to be related and they called it SARS-CoV-2. There are some serious problems with this assumption and *assumption* was all it was. Most 'factual' science turns out to be assumptions repeated into everyone-knows-that. A match of under 80-percent is meaningless. Dr Kaufman makes the point that there's a *96 percent* genetic correlation between humans and chimpanzees, but 'no one would say our genetic material is part of the chimpanzee family'. Yet the Chinese authorities were claiming that a much lower percentage, less than 80 percent, proved the existence of a new 'coronavirus'. For goodness sake human DNA is 60 percent similar to a *banana*.

You are feeling sleepy

The entire 'Covid' hoax is a global Psyop, a psychological operation to program the human mind into believing and fearing a complete fantasy. A crucial aspect of this was what *appeared* to happen in Italy. It was all very well streaming out daily images of an alleged catastrophe in Wuhan, but to the Western mind it was still on the other side of the world in a very different culture and setting. A reaction of 'this could happen to me and my family' was still nothing like as intense enough for the mind-doctors. The Cult needed a Western example to push people over that edge and it chose Italy, one of its major global locations going back to the Roman Empire. An Italian 'Covid' crisis was manufactured in a particular area called Lombardy which just happens to be notorious for its toxic air and therefore respiratory disease. Wuhan, China, déjà vu. An hysterical media told horror stories of Italians dying from 'Covid' in their droves and how Lombardy hospitals were being overrun by a tidal wave of desperately ill people needing treatment after being struck down by the 'deadly virus'. Here was the psychological turning point the Cult had planned. Wow, if this is happening in Italy, the Western mind concluded, this indeed could happen to me and my family. Another point is that Italian authorities responded by following the Chinese blueprint so vehemently recommended by the Cult-owned World Health Organization. They imposed fascistic lockdowns on the whole country viciously policed with the help of surveillance drones sweeping through the streets seeking out anyone who escaped from mass house arrest. Livelihoods were destroyed and psychology unravelled in the way we have witnessed since in all lockdown countries. Crucial to the plan was that Italy responded in this way to set the precedent of suspending freedom and imposing fascism in a 'Western liberal democracy'. I emphasised in an animated video explanation on davidicke.com posted in the summer of 2020 how important it was to the Cult to expand the Chinese lockdown model across the West. Without this, and the bare-faced lie that non-symptomatic people could still transmit a 'disease' they didn't have, there was no way locking down the whole population, sick and not sick, could be pulled off. At just the right time and with no evidence Cult operatives and gofers claimed that people without symptoms could pass on the 'disease'. In the name of protecting the 'vulnerable' like elderly people, who lockdowns would kill by the tens of thousands, we had for the first time healthy people told to isolate as well as the sick. The great majority of people who tested positive had no symptoms because there was nothing wrong with them. It was just a trick made possible by a test not testing for the 'virus'.

Months after my animated video the Gates-funded Professor Neil Ferguson at the Gates-funded Imperial College confirmed that I was right. He didn't say it in those terms, naturally, but he did say it. Ferguson will enter the story shortly for his outrageously crazy 'computer models' that led to Britain, the United States and many other countries following the Chinese and now Italian methods of response. Put another way, following the Cult script. Ferguson said that SAGE, the UK government's scientific advisory group which has controlled 'Covid' policy from the start, wanted to follow the Chinese lockdown model (while they all continued to work and be paid), but they wondered if they could possibly, in Ferguson's words, 'get away with it in Europe'. 'Get away with it'? Who the hell do these moronic, arrogant people think they are? This appalling man Ferguson said that once Italy went into national lockdown they realised they, too, could mimic China:

> It's a communist one-party state, we said. We couldn't get away with it in Europe, we thought … and then Italy did it. And we realised we could. Behind this garbage from Ferguson is a simple fact: Doing the same as China in every country was the plan from the start and Ferguson's 'models' would play a central role in achieving that. It's just a coincidence, of course, and absolutely nothing to worry your little head about.

Oops, sorry, our mistake

Once the Italian segment of the Psyop had done the job it was designed to do a very different story emerged. Italian authorities revealed that 99 percent of those who had 'died from Covid-19' in Italy had one, two, three, or more 'co-morbidities' or illnesses and health problems that could have ended their life. The US Centers for Disease Control and Prevention (CDC) published a figure of 94 percent for Americans dying of 'Covid' while having other serious medical conditions – on average two to three (some five or six) other potential causes of death. In terms of death from an unproven 'virus' I say it is 100 percent. The other one percent in Italy and six percent in the US would presumably have died from 'Covid's' flu-like symptoms with a range of other possible causes in conjunction with a test not testing for the 'virus'. Fox News reported that even more startling figures had emerged in one US county in which 410 of 422 deaths attributed to 'Covid-19' had other potentially deadly health conditions. The Italian National Health Institute said later that the average age of people dying with a 'Covid-19' diagnosis in Italy was about 81. Ninety percent were over 70 with ten percent over 90. In terms of other reasons to die some 80 percent had two or more chronic diseases with half having three or more including cardiovascular problems, diabetes, respiratory problems and cancer. Why is the phantom 'Covid-19' said to kill overwhelmingly old people and hardly affect the young? Old people continually die of many causes and especially respiratory disease which you can re-diagnose 'Covid-19' while young people die in tiny numbers by comparison and rarely of respiratory disease. Old people 'die of Covid' because they die of other things that can be redesignated 'Covid' and it really is that simple.

Flu has flown

The blueprint was in place. Get your illusory 'cases' from a test not testing for the 'virus' and redesignate other causes of death as 'Covid-19'. You have an instant 'pandemic' from something that is nothing more than a computer-generated fiction. With near-on a billion people having ‘flu-like’ symptoms every year the potential was limitless and we can see why flu quickly and apparently miraculously disappeared *worldwide* by being diagnosed ‘Covid-19’. The painfully bloody obvious was explained away by the childlike media in headlines like this in the UK ‘*Independent*’: ‘Not a single case of flu detected by Public Health England this year as Covid restrictions suppress virus’. I kid you not. The masking, social distancing and house arrest that did not make the ‘Covid virus’ disappear somehow did so with the ‘flu virus’. Even worse the article, by a bloke called Samuel Lovett, suggested that maybe the masking, sanitising and other ‘Covid’ measures should continue to keep the flu away. With a ridiculousness that disturbs your breathing (it’s ‘Covid-19’) the said Lovett wrote: ‘With widespread social distancing and mask-wearing measures in place throughout the UK, the usual routes of transmission for influenza have been blocked.’ He had absolutely no evidence to support that statement, but look at the consequences of him acknowledging the obvious. With flu not disappearing at all and only being relabelled ‘Covid-19’ he would have to contemplate that ‘Covid’ was a hoax on a scale that is hard to imagine. You need guts and commitment to truth to even go there and that’s clearly something Samuel Lovett does not have in abundance. He would never have got it through the editors anyway.

Tens of thousands die in the United States alone every winter from flu including many with pneumonia complications. CDC figures record *45 million* Americans diagnosed with flu in 2017-2018 of which 61,000 died and some reports claim 80,000. Where was the same hysteria then that we have seen with ‘Covid-19’? Some 250,000 Americans are admitted to hospital with pneumonia every year with about 50,000 cases proving fatal. About 65 million suffer respiratory disease every year and three million deaths makes this the third biggest cause of death worldwide. You only have to redesignate a portion of all these people ‘Covid-19’ and you have an instant global pandemic or the *appearance* of one. Why would doctors do this? They are told to do this and all but a few dare not refuse those who must be obeyed. Doctors in general are not researching their own knowledge and instead take it direct and unquestioned from the authorities that own them and their careers. The authorities say they must now diagnose these symptoms 'Covid-19' and not flu, or whatever, and they do it. Dark suits say put 'Covid-19' on death certificates no matter what the cause of death and the doctors do it. Renegade Minds don't fall for the illusion that doctors and medical staff are all highly-intelligent, highly-principled, seekers of medical truth. *Some are*, but not the majority. They are repeaters, gofers, and yes sir, no sir, purveyors of what the system demands they purvey. The 'Covid' con is not merely confined to diseases of the lungs. Instructions to doctors to put 'Covid-19' on death certificates for anyone dying of *anything* within 28 days (or much more) of a positive test not testing for the 'virus' opened the floodgates. The term dying *with* 'Covid' and not *of* 'Covid' was coined to cover the truth. Whether it was a *with* or an *of* they were all added to the death numbers attributed to the 'deadly virus' compiled by national governments and globally by the Gates-funded Johns Hopkins operation in the United States that was so involved in those 'pandemic' simulations. Fraudulent deaths were added to the ever-growing list of fraudulent 'cases' from false positives from a false test. No wonder Professor Walter Ricciardi, scientific advisor to the Italian minister of health, said after the Lombardy hysteria had done its job that 'Covid' death rates were due to Italy having the second oldest population in the world and to *how hospitals record deaths*:

> The way in which we code deaths in our country is very generous in the sense that all the people who die in hospitals with the coronavirus are deemed to be dying of the coronavirus. On re-evaluation by the National Institute of Health, only 12 per cent of death certificates have shown a direct causality from coronavirus, while 88 per cent of patients who have died have at least one pre-morbidity – many had two or three.

This is extraordinary enough when you consider the propaganda campaign to use Italy to terrify the world, but how can they even say twelve percent were genuine when the 'virus' has not been shown to exist, its 'code' is a computer program, and diagnosis comes from a test not testing for it? As in China, and soon the world, 'Covid-19' in

Italy was a redesignation of diagnosis. Lies and corruption were to become the real 'pandemic' fuelled by a pathetically-compliant medical system taking its orders from the tiny few at the top of their national hierarchy who answered to the World Health Organization which answers to Gates and the Cult. Doctors were told – ordered – to diagnose a particular set of symptoms 'Covid-19' and put that on the death certificate for any cause of death if the patient had tested positive with a test not testing for the virus or had 'Covid' symptoms like the flu. The United States even introduced big financial incentives to manipulate the figures with hospitals receiving £4,600 from the Medicare system for diagnosing someone with regular pneumonia, $13,000 if they made the diagnosis from the same symptoms 'Covid-19' pneumonia, and $39, 000 if they put a 'Covid' diagnosed patient on a ventilator that would almost certainly kill them. A few – painfully and pathetically few – medical whistleblowers revealed (before Cult-owned YouTube deleted their videos) that they had been instructed to 'let the patient crash' and put them straight on a ventilator instead of going through a series of far less intrusive and dangerous methods as they would have done before the pandemic hoax began and the financial incentives kicked in. We are talking cold-blooded murder given that ventilators are so damaging to respiratory systems they are usually the last step before heaven awaits. Renegade Minds never fall for the belief that people in white coats are all angels of mercy and cannot be full-on psychopaths. I have explained in detail in *The Answer* how what I am describing here played out across the world coordinated by the World Health Organization through the medical hierarchies in almost every country.

Medical scientist calls it

Information about the non-existence of the 'virus' began to emerge for me in late March, 2020, and mushroomed after that. I was sent an email by Sir Julian Rose, a writer, researcher, and organic farming promotor, from a medical scientist friend of his in the United States. Even at that early stage in March the scientist was able to explain how the ‘Covid’ hoax was being manipulated. He said there were no reliable tests for a specific ‘Covid-19 virus’ and nor were there any reliable agencies or media outlets for reporting numbers of actual ‘Covid-19’ cases. We have seen in the long period since then that he was absolutely right. ‘Every action and reaction to Covid-19 is based on totally flawed data and we simply cannot make accurate assessments,’ he said. Most people diagnosed with ‘Covid-19’ were showing nothing more than cold and flu-like symptoms ‘because most coronavirus strains *are* nothing more than cold/flu-like symptoms’. We had farcical situations like an 84-year-old German man testing positive for ‘Covid-19’ and his nursing home ordered to quarantine only for him to be found to have a common cold. The scientist described back then why PCR tests and what he called the ‘Mickey Mouse test kits’ were useless for what they were claimed to be identifying. ‘The idea these kits can isolate a specific virus like Covid-19 is nonsense,’ he said. Significantly, he pointed out that ‘if you want to create a totally false panic about a totally false pandemic – pick a coronavirus’. This is exactly what the Cult-owned Gates, World Economic Forum and Johns Hopkins University did with their Event 201 ‘simulation’ followed by their real-life simulation called the ‘pandemic’. The scientist said that all you had to do was select the sickest of people with respiratory-type diseases in a single location – ‘say Wuhan’ – and administer PCR tests to them. You can then claim that anyone showing ‘viral sequences’ similar to a coronavirus ‘which will inevitably be quite a few’ is suffering from a ‘new’ disease:

> Since you already selected the sickest flu cases a fairly high proportion of your sample will go on to die. You can then say this ‘new’ virus has a CFR [case fatality rate] higher than the flu and use this to infuse more concern and do more tests which will of course produce more ‘cases’, which expands the testing, which produces yet more ‘cases’ and so on and so on. Before long you have your ‘pandemic’, and all you have done is use a simple test kit trick to convert the worst flu and pneumonia cases into something new that doesn’t ACTUALLY EXIST [my emphasis].

He said that you then ‘just run the same scam in other countries’ and make sure to keep the fear message running high ‘so that people will feel panicky and less able to think critically’. The only problem to overcome was the fact *there is no* actual new deadly pathogen and only regular sick people. This meant that deaths from the ‘new deadly pathogen’ were going to be way too low for a real new deadly virus pandemic, but he said this could be overcome in the following ways – all of which would go on to happen:

> 1. You can claim this is just the beginning and more deaths are imminent [you underpin this with fantasy ‘computer projections’]. Use this as an excuse to quarantine everyone and then claim the quarantine prevented the expected millions of dead.
>
> 2. You can [say that people] ‘minimizing’ the dangers are irresponsible and bully them into not talking about numbers.
>
> 3. You can talk crap about made up numbers hoping to blind people with pseudoscience.
>
> 4. You can start testing well people (who, of course, will also likely have shreds of coronavirus [RNA] in them) and thus inflate your ‘case figures’ with ‘asymptomatic carriers’ (you will of course have to spin that to sound deadly even though any virologist knows the more symptom-less cases you have the less deadly is your pathogen).

The scientist said that if you take these simple steps ‘you can have your own entirely manufactured pandemic up and running in weeks’. His analysis made so early in the hoax was brilliantly prophetic of what would actually unfold. Pulling all the information together in these recent chapters we have this is simple 1, 2, 3, of how you can delude virtually the entire human population into believing in a ‘virus’ that doesn’t exist:

- A ‘Covid case’ is someone who tests positive with a test not testing for the ‘virus’.

- A ‘Covid death’ is someone who dies of *any cause* within 28 days (or much longer) of testing positive with a test not testing for the ‘virus.

- Asymptomatic means there is nothing wrong with you, but they claim you can pass on what you don’t have to justify locking down (quarantining) healthy people in totality.

The foundations of the hoax are that simple. A study involving ten million people in Wuhan, published in November, 2020, demolished the whole lie about those without symptoms passing on the 'virus'. They found '300 asymptomatic cases' and traced their contacts to find that not one of them was detected with the 'virus'. 'Asymptomatic' patients and their contacts were isolated for no less than two weeks and nothing changed. I know it's all crap, but if you are going to claim that those without symptoms can transmit 'the virus' then you must produce evidence for that and they never have. Even World Health Organization official Dr Maria Van Kerkhove, head of the emerging diseases and zoonosis unit, said as early as June, 2020, that she doubted the validity of asymptomatic transmission. She said that 'from the data we have, it still seems to be rare that an asymptomatic person actually transmits onward to a secondary individual' and by 'rare' she meant that she couldn't cite any case of asymptomatic transmission.

The Ferguson factor

The problem for the Cult as it headed into March, 2020, when the script had lockdown due to start, was that despite all the manipulation of the case and death figures they still did not have enough people alleged to have died from 'Covid' to justify mass house arrest. This was overcome in the way the scientist described: 'You can claim this is just the beginning and more deaths are imminent … Use this as an excuse to quarantine everyone and then claim the quarantine prevented the expected millions of dead.' Enter one Professor Neil Ferguson, the Gates-funded 'epidemiologist' at the Gates-funded Imperial College in London. Ferguson is Britain's Christian Drosten in that he has a dire record of predicting health outcomes, but is still called upon to advise government on the next health outcome when another 'crisis' comes along. This may seem to be a strange and ridiculous thing to do. Why would you keep turning for policy guidance to people who have a history of being monumentally wrong? Ah, but it makes sense from the Cult point of view. These ‘experts’ keep on producing predictions that suit the Cult agenda for societal transformation and so it was with Neil Ferguson as he revealed his horrific (and clearly insane) computer model predictions that allowed lockdowns to be imposed in Britain, the United States and many other countries. Ferguson does not have even an A-level in biology and would appear to have no formal training in computer modelling, medicine or epidemiology, according to Derek Winton, an MSc in Computational Intelligence. He wrote an article somewhat aghast at what Ferguson did which included taking no account of respiratory disease ‘seasonality’ which means it is far worse in the winter months. Who would have thought that respiratory disease could be worse in the winter? Well, certainly not Ferguson.

The massively China-connected Imperial College and its bizarre professor provided the excuse for the long-incubated Chinese model of human control to travel westward at lightning speed. Imperial College confirms on its website that it collaborates with the Chinese Research Institute; publishes more than 600 research papers every year with Chinese research institutions; has 225 Chinese staff; 2,600 Chinese students – the biggest international group; 7,000 former students living in China which is the largest group outside the UK; and was selected for a tour by China’s President Xi Jinping during his state visit to the UK in 2015. The college takes major donations from China and describes itself as the UK’s number one university collaborator with Chinese research institutions. The China communist/fascist government did not appear phased by the woeful predictions of Ferguson and Imperial when during the lockdown that Ferguson induced the college signed a five-year collaboration deal with China tech giant Huawei that will have Huawei’s indoor 5G network equipment installed at the college’s West London tech campus along with an ‘AI cloud platform’. The deal includes Chinese sponsorship of Imperial’s Venture Catalyst entrepreneurship competition. Imperial is an example of the enormous influence the Chinese government has within British and North American universities and research centres – and further afield. Up to 200 academics from more than a dozen UK universities are being investigated on suspicion of ‘unintentionally’ helping the Chinese government build weapons of mass destruction by ‘transferring world-leading research in advanced military technology such as aircraft, missile designs and cyberweapons’. Similar scandals have broken in the United States, but it’s all a coincidence. Imperial College serves the agenda in many other ways including the promotion of every aspect of the United Nations Agenda 21/2030 (the Great Reset) and produced computer models to show that human-caused ‘climate change’ is happening when in the real world it isn’t. Imperial College is driving the climate agenda as it drives the ‘Covid’ agenda (both Cult hoaxes) while Patrick Vallance, the UK government’s Chief Scientific Adviser on ‘Covid’, was named Chief Scientific Adviser to the UN ‘climate change’ conference known as COP26 hosted by the government in Glasgow, Scotland. ‘Covid’ and ‘climate’ are fundamentally connected.

Professor Woeful

From Imperial’s bosom came Neil Ferguson still advising government despite his previous disasters and it was announced early on that he and other key people like UK Chief Medical Adviser Chris Whitty had caught the ‘virus’ as the propaganda story was being sold. Somehow they managed to survive and we had Prime Minister Boris Johnson admitted to hospital with what was said to be a severe version of the ‘virus’ in this same period. His whole policy and demeanour changed when he returned to Downing Street. It’s a small world with these government advisors – especially in their communal connections to Gates – and Ferguson had partnered with Whitty to write a paper called ‘Infectious disease: Tough choices to reduce Ebola transmission’ which involved another scare-story that didn’t happen. Ferguson’s ‘models’ predicted that up to150, 000 could die from ‘mad cow disease’, or BSE, and its version in sheep if it was transmitted to humans. BSE was not transmitted and instead triggered by an organophosphate pesticide used to treat a pest on cows. Fewer than 200 deaths followed from the human form. Models by Ferguson and his fellow incompetents led to the unnecessary culling of millions of pigs, cattle and sheep in the foot and mouth outbreak in 2001 which destroyed the lives and livelihoods of farmers and their families who had often spent decades building their herds and flocks. Vast numbers of these animals did not have foot and mouth and had no contact with the infection. Another 'expert' behind the cull was Professor Roy Anderson, a computer modeller at Imperial College specialising in the epidemiology of *human*, not animal, disease. Anderson has served on the Bill and Melinda Gates Grand Challenges in Global Health advisory board and chairs another Gates-funded organisation. Gates is everywhere.

In a precursor to the 'Covid' script Ferguson backed closing schools 'for prolonged periods' over the swine flu 'pandemic' in 2009 and said it would affect a third of the world population if it continued to spread at the speed he claimed to be happening. His mates at Imperial College said much the same and a news report said: 'One of the authors, the epidemiologist and disease modeller Neil Ferguson, who sits on the World Health Organisation's emergency committee for the outbreak, said the virus had "full pandemic potential".' Professor Liam Donaldson, the Chris Whitty of his day as Chief Medical Officer, said the worst case could see 30 percent of the British people infected by swine flu with 65,000 dying. Ferguson and Donaldson were indeed proved correct when at the end of the year the number of deaths attributed to swine flu was 392. The term 'expert' is rather liberally applied unfortunately, not least to complete idiots. Swine flu 'projections' were great for GlaxoSmithKline (GSK) as millions rolled in for its Pandemrix influenza vaccine which led to brain damage with children most affected. The British government (taxpayers) paid out more than £60 million in compensation after GSK was given immunity from prosecution. Yet another 'Covid' déjà vu. Swine flu was supposed to have broken out in Mexico, but Dr Wolfgang Wodarg, a German doctor, former member of parliament and critic of the 'Covid' hoax, observed 'the spread of swine flu' in Mexico City at the time. He said: ‘What we experienced in Mexico City was a very mild flu which did not kill more than usual – which killed even fewer people than usual.’ Hyping the fear against all the facts is not unique to ‘Covid’ and has happened many times before. Ferguson is reported to have over-estimated the projected death toll of bird flu (H5N1) by some three million-fold, but bird flu vaccine makers again made a killing from the scare. This is some of the background to the Neil Ferguson who produced the perfectly-timed computer models in early 2020 predicting that half a million people would die in Britain without draconian lockdown and 2.2 million in the United States. Politicians panicked, people panicked, and lockdowns of alleged short duration were instigated to ‘flatten the curve’ of cases gleaned from a test not testing for the ‘virus’. I said at the time that the public could forget the ‘short duration’ bit. This was an agenda to destroy the livelihoods of the population and force them into mass control through dependency and there was going to be nothing ‘short’ about it. American researcher Daniel Horowitz described the consequences of the ‘models’ spewed out by Gates-funded Ferguson and Imperial College:

> What led our government and the governments of many other countries into panic was a single Imperial College of UK study, funded by global warming activists, that predicted 2.2 million deaths if we didn’t lock down the country. In addition, the reported 8-9% death rate in Italy scared us into thinking there was some other mutation of this virus that they got, which might have come here.
>
> Together with the fact that we were finally testing and had the ability to actually report new cases, we thought we were headed for a death spiral. But again … we can’t flatten a curve if we don’t know when the curve started.

How about it *never* started?

Giving them what they want

An investigation by German news outlet *Welt Am Sonntag (World on Sunday)* revealed how in March, 2020, the German government gathered together ‘leading scientists from several research institutes and universities’ and ‘together, they were to produce a [modelling] paper that would serve as legitimization for further tough political measures'. The Cult agenda was justified by computer modelling not based on evidence or reality; it was specifically constructed to justify the Cult demand for lockdowns all over the world to destroy the independent livelihoods of the global population. All these modellers and everyone responsible for the 'Covid' hoax have a date with a trial like those in Nuremberg after World War Two when Nazis faced the consequences of their war crimes. These corrupt-beyond-belief 'modellers' wrote the paper according to government instructions and it said that that if lockdown measures were lifted then up to one million Germans would die from 'Covid-19' adding that some would die 'agonizingly at home, gasping for breath' unable to be treated by hospitals that couldn't cope. All lies. No matter – it gave the Cult all that it wanted. What did long-time government 'modeller' Neil Ferguson say? If the UK and the United States didn't lockdown half a million would die in Britain and 2.2 million Americans. Anyone see a theme here? 'Modellers' are such a crucial part of the lockdown strategy that we should look into their background and follow the money. Researcher Rosemary Frei produced an excellent article headlined 'The Modelling-paper Mafiosi'. She highlights a guy called John Edmunds, a British epidemiologist, and professor in the Faculty of Epidemiology and Population Health at the London School of Hygiene & Tropical Medicine. He studied at Imperial College. Edmunds is a member of government 'Covid' advisory bodies which have been dictating policy, the New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG) and the Scientific Advisory Group for Emergencies (SAGE).

Ferguson, another member of NERVTAG and SAGE, led the way with the original 'virus' and Edmunds has followed in the 'variant' stage and especially the so-called UK or Kent variant known as the 'Variant of Concern' (VOC) B.1.1.7. He said in a co-written report for the Centre for Mathematical modelling of Infectious Diseases at the London School of Hygiene and Tropical Medicine, with input from the Centre's 'Covid-19' Working Group, that there was 'a realistic possibility that VOC B.1.1.7 is associated with an increased risk of death compared to non-VOC viruses'. Fear, fear, fear, get the vaccine, fear, fear, fear, get the vaccine. Rosemary Frei reveals that almost all the paper's authors and members of the modelling centre's 'Covid-19' Working Group receive funding from the Bill and Melinda Gates Foundation and/or the associated Gates-funded Wellcome Trust. The paper was published by e-journal *Medr χiv* which only publishes papers not peer-reviewed and the journal was established by an organisation headed by Facebook's Mark Zuckerberg and his missus. What a small world it is. Frei discovered that Edmunds is on the Scientific Advisory Board of the Coalition for Epidemic Preparedness Innovations (CEPI) which was established by the Bill and Melinda Gates Foundation, Klaus Schwab's Davos World Economic Forum and Big Pharma giant Wellcome. CEPI was 'launched in Davos [in 2017] to develop vaccines to stop future epidemics', according to its website. 'Our mission is to accelerate the development of vaccines against emerging infectious diseases and enable equitable access to these vaccines for people during outbreaks.' What kind people they are. Rosemary Frei reveals that Public Health England (PHE) director Susan Hopkins is an author of her organisation's non-peer-reviewed reports on 'new variants'. Hopkins is a professor of infectious diseases at London's Imperial College which is gifted tens of millions of dollars a year by the Bill and Melinda Gates Foundation. Gates-funded modelling disaster Neil Ferguson also co-authors Public Health England reports and he spoke in December, 2020, about the potential danger of the B.1.1.7. 'UK variant' promoted by Gates-funded modeller John Edmunds. When I come to the 'Covid vaccines' the 'new variants' will be shown for what they are – bollocks.

Connections, connections

All these people and modellers are lockdown-obsessed or, put another way, they demand what the Cult demands. Edmunds said in January, 2021, that to ease lockdowns too soon would be a disaster and they had to 'vaccinate much, much, much more widely than the elderly'. Rosemary Frei highlights that Edmunds is married to Jeanne Pimenta who is described in a LinkedIn profile as director of epidemiology at GlaxoSmithKline (GSK) and she held shares in the company. Patrick Vallance, co-chair of SAGE and the government's Chief Scientific Adviser, is a former executive of GSK and has a deferred bonus of shares in the company worth £600,000. GSK has serious business connections with Bill Gates and is collaborating with mRNA-'vaccine' company CureVac to make 'vaccines' for the new variants that Edmunds is talking about. GSK is planning a 'Covid vaccine' with drug giant Sanofi. Puppet Prime Minister Boris Johnson announced in the spring of 2021 that up to 60 million vaccine doses were to be made at the GSK facility at Barnard Castle in the English North East. Barnard Castle, with a population of just 6,000, was famously visited in breach of lockdown rules in April, 2020, by Johnson aide Dominic Cummings who said that he drove there 'to test his eyesight' before driving back to London. Cummings would be better advised to test his integrity – not that it would take long. The GSK facility had nothing to do with his visit then although I'm sure Patrick Vallance would have been happy to arrange an introduction and some tea and biscuits. Ruthless psychopath Gates has made yet another fortune from vaccines in collaboration with Big Pharma companies and gushes at the phenomenal profits to be made from vaccines – more than a 20-to-1 return as he told one interviewer. Gates also tweeted in December, 2019, with the foreknowledge of what was coming: 'What's next for our foundation? I'm particularly excited about what the next year could mean for one of the best buys in global health: vaccines.'

Modeller John Edmunds is a big promotor of vaccines as all these people appear to be. He's the dean of the London School of Hygiene & Tropical Medicine's Faculty of Epidemiology and Population Health which is primarily funded by the Bill and Melinda Gates Foundation and the Gates-established and funded GAVI vaccine alliance which is the Gates vehicle to vaccinate the world. The organisation Doctors Without Borders has described GAVI as being 'aimed more at supporting drug-industry desires to promote new products than at finding the most efficient and sustainable means for fighting the diseases of poverty'. But then that's why the psychopath Gates created it. John Edmunds said in a video that the London School of Hygiene & Tropical Medicine is involved in every aspect of vaccine development including large-scale clinical trials. He contends that mathematical modelling can show that vaccines protect individuals and society. That's on the basis of shit in and shit out, I take it. Edmunds serves on the UK Vaccine Network as does Ferguson and the government's foremost 'Covid' adviser, the grim-faced, dark-eyed Chris Whitty. The Vaccine Network says it works 'to support the government to identify and shortlist targeted investment opportunities for the most promising vaccines and vaccine technologies that will help combat infectious diseases with epidemic potential, and to address structural issues related to the UK's broader vaccine infrastructure'. Ferguson is acting Director of the Imperial College Vaccine Impact Modelling Consortium which has funding from the Bill and Melina Gates Foundation and the Gates-created GAVI 'vaccine alliance'. Anyone wonder why these characters see vaccines as the answer to every problem? Ferguson is wildly enthusiastic in his support for GAVI's campaign to vaccine children en masse in poor countries. You would expect someone like Gates who has constantly talked about the need to reduce the population to want to fund vaccines to keep more people alive. I'm sure that's why he does it. The John Edmunds London School of Hygiene & Tropical Medicine (LSHTM) has a Vaccines Manufacturing Innovation Centre which develops, tests and commercialises vaccines. Rosemary Frei writes:

> The vaccines centre also performs affiliated activities like combating 'vaccine hesitancy'. The latter includes the Vaccine Confidence Project. The project's stated purpose is, among other things, 'to provide analysis and guidance for early response and engagement with the public to ensure sustained confidence in vaccines and immunisation'. The Vaccine Confidence Project's director is LSHTM professor Heidi Larson. For more than a decade she's been researching how to combat vaccine hesitancy.

How the bloody hell can blokes like John Edmunds and Neil Ferguson with those connections and financial ties model 'virus' case and death projections for the government and especially in a way that gives their paymasters like Gates exactly what they want? It's insane, but this is what you find throughout the world.

'Covid' is not dangerous, oops, wait, yes it is

Only days before Ferguson's nightmare scenario made Jackboot Johnson take Britain into a China-style lockdown to save us from a deadly 'virus' the UK government website gov.uk was reporting something very different to Ferguson on a page of official government guidance for 'high consequence infectious diseases (HCID)'. It said this about 'Covid-19':

> As of 19 March 2020, COVID-19 *is no longer considered to be a high consequence infectious diseases (HCID) in the UK* [my emphasis]. The 4 nations public health HCID group made an interim recommendation in January 2020 to classify COVID-19 as an HCID. This was based on consideration of the UK HCID criteria about the virus and the disease with information available during the early stages of the outbreak.
>
> Now that more is known about COVID-19, the public health bodies in the UK have reviewed the most up to date information about COVID-19 against the UK HCID criteria. They have determined that several features have now changed; in particular, more information is available about mortality rates (low overall), and there is now greater clinical awareness and a specific and sensitive laboratory test, the availability of which continues to increase. The Advisory Committee on Dangerous Pathogens (ACDP) is also of the opinion that COVID-19 should no longer be classified as an HCID.

Soon after the government had been exposed for downgrading the risk they upgraded it again and everyone was back to singing from the same Cult hymn book. Ferguson and his fellow Gates clones indicated that lockdowns and restrictions would have to continue until a Gates-funded vaccine was developed. Gates said the same because Ferguson and his like were repeating the Gates script which is the Cult script. 'Flatten the curve' became an ongoing nightmare of continuing lockdowns with periods in between of severe restrictions in pursuit of destroying independent incomes and had nothing to do with protecting health about which the Cult gives not a shit. Why wouldn't Ferguson be pushing a vaccine 'solution' when he's owned by vaccine-obsessive Gates who makes a fortune from them and when Ferguson heads the Vaccine Impact Modelling Consortium at Imperial College funded by the Gates Foundation and GAVI, the ‘vaccine alliance’, created by Gates as his personal vaccine promotion operation? To compound the human catastrophe that Ferguson’s ‘models’ did so much to create he was later exposed for breaking his own lockdown rules by having sexual liaisons with his married girlfriend Antonia Staats at his home while she was living at another location with her husband and children. Staats was a ‘climate’ activist and senior campaigner at the Soros-funded Avaaz which I wouldn’t trust to tell me that grass is green. Ferguson had to resign as a government advisor over this hypocrisy in May, 2020, but after a period of quiet he was back being quoted by the ridiculous media on the need for more lockdowns and a vaccine rollout. Other government-advising ‘scientists’ from Imperial College’ held the fort in his absence and said lockdown could be indefinite until a vaccine was found. The Cult script was being sung by the payrolled choir. I said there was no intention of going back to ‘normal’ when the ‘vaccine’ came because the ‘vaccine’ is part of a very different agenda that I will discuss in Human 2.0. Why would the Cult want to let the world go back to normal when destroying that normal forever was the whole point of what was happening? House arrest, closing businesses and schools through lockdown, (un)social distancing and masks all followed the Ferguson fantasy models. Again as I predicted (these people are so predictable) when the ‘vaccine’ arrived we were told that house arrest, lockdown, (un)social distancing and masks would still have to continue. I will deal with the masks in the next chapter because they are of fundamental importance.

Where’s the ‘pandemic’?

Any mildly in-depth assessment of the figures revealed what was really going on. Cult-funded and controlled organisations still have genuine people working within them such is the number involved. So it is with Genevieve Briand, assistant program director of the Applied Economics master’s degree program at Johns Hopkins

University. She analysed the impact that 'Covid-19' had on deaths from *all* causes in the United States using official data from the CDC for the period from early February to early September, 2020. She found that allegedly 'Covid' *related*-deaths exceeded those from heart disease which she found strange with heart disease always the biggest cause of fatalities. Her research became even more significant when she noted the sudden decline in 2020 of *all* non-'Covid' deaths: 'This trend is completely contrary to the pattern observed in all previous years … the total decrease in deaths by other causes almost exactly equals the increase in deaths by Covid-19.' This was such a game, set and match in terms of what was happening that Johns Hopkins University deleted the article on the grounds that it 'was being used to support false and dangerous inaccuracies about the impact of the pandemic'. No – because it exposed the scam from official CDC figures and this was confirmed when those figures were published in January, 2021. Here we can see the effect of people dying from heart attacks, cancer, road accidents and gunshot wounds – *anything* – having 'Covid-19' on the death certificate along with those diagnosed from 'symptoms' who had even not tested positive with a test not testing for the 'virus'. I am not kidding with the gunshot wounds, by the way. Brenda Bock, coroner in Grand County, Colorado, revealed that two gunshot victims tested positive for the 'virus' within the previous 30 days and were therefore classified as 'Covid deaths'. Bock said: 'These two people had tested positive for Covid, but that's not what killed them. A gunshot wound is what killed them.' She said she had not even finished her investigation when the state listed the gunshot victims as deaths due to the 'virus'. The death and case figures for 'Covid-19' are an absolute joke and yet they are repeated like parrots by the media, politicians and alleged medical 'experts'. The official Cult narrative is the only show in town.

Genevieve Briand found that deaths from all causes were not exceptional in 2020 compared with previous years and a Spanish magazine published figures that said the same about Spain which was a 'Covid' propaganda hotspot at one point. *Discovery Salud*, a health and medicine magazine, quoted government figures which showed how 17,000 *fewer* people died in Spain in 2020 than in 2019 and more than 26,000 fewer than in 2018. The age-standardised mortality rate for England and Wales when age distribution is taken into account was significantly lower in 2020 than the 1970s, 80s and 90s, and was only the ninth highest since 2000. Where is the 'pandemic'?

Post mortems and autopsies virtually disappeared for 'Covid' deaths amid claims that 'virus-infected' bodily fluids posed a risk to those carrying out the autopsy. This was rejected by renowned German pathologist and forensic doctor Klaus Püschel who said that he and his staff had by then done 150 autopsies on 'Covid' patients with no problems at all. He said they were needed to know why some 'Covid' patients suffered blood clots and not severe respiratory infections. The 'virus' is, after all, called SARS or 'severe acute respiratory syndrome'. I highlighted in the spring of 2020 this phenomenon and quoted New York intensive care doctor Cameron Kyle-Sidell who posted a soon deleted YouTube video to say that they had been told to prepare to treat an infectious disease called 'Covid-19', but that was not what they were dealing with. Instead he likened the lung condition of the most severely ill patients to what you would expect with cabin depressurisation in a plane at 30,000 feet or someone dropped on the top of Everest without oxygen or acclimatisation. I have never said this is not happening to a small minority of alleged 'Covid' patients – I am saying this is not caused by a phantom 'contagious virus'. Indeed Kyle-Sidell said that 'Covid-19' was not the disease they were told was coming their way. 'We are operating under a medical paradigm that is untrue,' he said, and he believed they were treating the wrong disease: 'These people are being slowly starved of oxygen.' Patients would take off their oxygen masks in a state of fear and stress and while they were blue in the face on the brink of death. They did not look like patients dying of pneumonia. You can see why they don't want autopsies when their virus doesn't exist and there is another condition in some people that they don't wish to be uncovered. I should add here that the 5G system of millimetre waves was being rapidly introduced around the world in 2020 and even more so now as they fire 5G at the Earth from satellites. At 60 gigahertz within the 5G range that frequency interacts with the oxygen molecule and stops people breathing in sufficient oxygen to be absorbed into the bloodstream. They are installing 5G in schools and hospitals. The world is not mad or anything. 5G can cause major changes to the lungs and blood as I detail in *The Answer* and these consequences are labelled 'Covid-19', the alleged symptoms of which can be caused by 5G and other electromagnetic frequencies as cells respond to radiation poisoning.

The 'Covid death' scam

Dr Scott Jensen, a Minnesota state senator and medical doctor, exposed 'Covid' Medicare payment incentives to hospitals and death certificate manipulation. He said he was sent a seven-page document by the US Department of Health 'coaching' him on how to fill out death certificates which had never happened before. The document said that he didn't need to have a laboratory test for 'Covid-19' to put that on the death certificate and that shocked him when death certificates are supposed to be about facts. Jensen described how doctors had been 'encouraged, if not pressured' to make a diagnosis of 'Covid-19' if they thought it was probable or *'presumed'*. No positive test was necessary – not that this would have mattered anyway. He said doctors were told to diagnose 'Covid' by symptoms when these were the same as colds, allergies, other respiratory problems, and certainly with influenza which 'disappeared' in the 'Covid' era. A common sniffle was enough to get the dreaded verdict. Ontario authorities decreed that a single care home resident with *one* symptom from a long list must lead to the isolation of the entire home. Other courageous doctors like Jensen made the same point about death figure manipulation and how deaths by other causes were falling while 'Covid-19 deaths' were rising at the same rate due to re-diagnosis. Their videos rarely survive long on YouTube with its Cult-supporting algorithms courtesy of CEO Susan Wojcicki and her bosses at Google. Figure-tampering was so glaring and ubiquitous that even officials were letting it slip or outright saying it. UK chief scientific adviser Patrick Vallance said on one occasion that ‘Covid’ on the death certificate doesn’t mean ‘Covid’ was the cause of death (so why the hell is it there?) and we had the rare sight of a BBC reporter telling the truth when she said: ‘Someone could be successfully treated for Covid, in say April, discharged, and then in June, get run over by a bus and die … That person would still be counted as a Covid death in England.’ Yet the BBC and the rest of the world media went on repeating the case and death figures as if they were real. Illinois Public Health Director Dr Ngozi Ezike revealed the deceit while her bosses must have been clenching their buttocks:

> If you were in a hospice and given a few weeks to live and you were then found to have Covid that would be counted as a Covid death. [There might be] a clear alternate cause, but it is still listed as a Covid death. So everyone listed as a Covid death doesn’t mean that was the cause of the death, but that they had Covid at the time of death.

Yes, a ‘Covid virus’ never shown to exist and tested for with a test not testing for the ‘virus’. In the first period of the pandemic hoax through the spring of 2020 the process began of designating almost everything a ‘Covid’ death and this has continued ever since. I sat in a restaurant one night listening to a loud conversation on the next table where a family was discussing in bewilderment how a relative who had no symptoms of ‘Covid’, and had died of a long-term problem, could have been diagnosed a death by the ‘virus’. I could understand their bewilderment. If they read this book they will know why this medical fraud has been perpetrated the world over.

Some media truth shock

The media ignored the evidence of death certificate fraud until eventually one columnist did speak out when she saw it first-hand. Bel Mooney is a long-time national newspaper journalist in Britain currently working for the *Daily Mail*. Her article on February 19th, 2021, carried this headline: ‘My dad Ted passed three Covid tests and died of a chronic illness yet he's officially one of Britain's 120,000 victims of the virus and is far from alone ... so how many more are there?' She told how her 99-year-old father was in a care home with a long-standing chronic obstructive pulmonary disease and vascular dementia. Maybe, but he was still aware enough to tell her from the start that there was no 'virus' and he refused the 'vaccine' for that reason. His death was not unexpected given his chronic health problems and Mooney said she was shocked to find that 'Covid-19' was declared the cause of death on his death certificate. She said this was a 'bizarre and unacceptable untruth' for a man with long-time health problems who had tested negative twice at the home for the 'virus'. I was also shocked by this story although not by what she said. I had been highlighting the death certificate manipulation for ten months. It was the confirmation that a professional full-time journalist only realised this was going on when it affected her directly and neither did she know that whether her dad tested positive or negative was irrelevant with the test not testing for the 'virus'. Where had she been? She said she did not believe in 'conspiracy theories' without knowing I'm sure that this and 'conspiracy theorists' were terms put into widespread circulation by the CIA in the 1960s to discredit those who did not accept the ridiculous official story of the Kennedy assassination. A blanket statement of 'I don't believe in conspiracy theories' is always bizarre. The dictionary definition of the term alone means the world is drowning in conspiracies. What she said was even more daft when her dad had just been affected by the 'Covid' conspiracy. Why else does she think that 'Covid-19' was going on the death certificates of people who died of something else?

To be fair once she saw from personal experience what was happening she didn't mince words. Mooney was called by the care home on the morning of February 9th to be told her father had died in his sleep. When she asked for the official cause of death what came back was 'Covid-19'. Mooney challenged this and was told there had been deaths from Covid on the dementia floor (confirmed by a test not testing for the 'virus') so they considered it 'reasonable to assume'. 'But doctor,' Mooney rightly protested, 'an assumption isn't a diagnosis.' She said she didn't blame the perfectly decent and sympathetic doctor – 'he was just doing his job'. Sorry, but that's *bullshit*. He wasn't doing his job at all. He was putting a false cause of death on the death certificate and that is a criminal offence for which he should be brought to account and the same with the millions of doctors worldwide who have done the same. They were not doing their job they were following orders and that must not wash at new Nuremberg trials any more than it did at the first ones. Mooney's doctor was 'assuming' (presuming) as he was told to, but 'just following orders' makes no difference to his actions. A doctor's job is to serve the patient and the truth, not follow orders, but that's what they have done all over the world and played a central part in making the 'Covid' hoax possible with all its catastrophic consequences for humanity. Shame on them and they must answer for their actions. Mooney said her disquiet worsened when she registered her father's death by telephone and was told by the registrar there had been very many other cases like hers where 'the deceased' had not tested positive for 'Covid' yet it was recorded as the cause of death. The test may not matter, but those involved at their level *think* it matters and it shows a callous disregard for accurate diagnosis. The pressure to do this is coming from the top of the national 'health' pyramids which in turn obey the World Health Organization which obeys Gates and the Cult. Mooney said the registrar agreed that this must distort the national figures adding that 'the strangest thing is that every winter we record countless deaths from flu, and this winter there have been none. Not one!' She asked if the registrar thought deaths from flu were being misdiagnosed and lumped together with 'Covid' deaths. The answer was a 'puzzled yes'. Mooney said that the funeral director said the same about 'Covid' deaths which had nothing to do with 'Covid'. They had lost count of the number of families upset by this and other funeral companies in different countries have had the same experience. Mooney wrote:

> The nightly shroud-waving and shocking close-ups of pain imposed on us by the TV news bewildered and terrified the population into eager compliance with lockdowns. We were invited to 'save the NHS' and to grieve for strangers – the real-life loved ones behind those shocking death counts. Why would the public imagine what I now fear, namely that the way Covid-19 death statistics are compiled might make the numbers seem greater than they are?

Oh, just a little bit – like 100 percent.

Do the maths

Mooney asked why a country would wish to skew its mortality figures by wrongly certifying deaths? What had been going on? Well, if you don't believe in conspiracies you will never find the answer which is that *it's a conspiracy*. She did, however, describe what she had discovered as a 'national scandal'. In reality it's a global scandal and happening everywhere. Pillars of this conspiracy were all put into place before the button was pressed with the Drosten PCR protocol and high amplifications to produce the cases and death certificate changes to secure illusory 'Covid' deaths. Mooney notes that normally two doctors were needed to certify a death, with one having to know the patient, and how the rules were changed in the spring of 2020 to allow one doctor to do this. In the same period 'Covid deaths' were decreed to be all cases where Covid-19 was put on the death certificate even without a positive test or any symptoms. Mooney asked: 'How many of the 30,851 (as of January 15) care home resident deaths with Covid-19 on the certificate (32.4 per cent of all deaths so far) were based on an assumption, like that of my father? And what has that done to our national psyche?'All of them is the answer to the first question and it has devastated and dismantled the national psyche, actually the global psyche, on a colossal scale. In the UK case and death data is compiled by organisations like Public Health England (PHE) and the Office for National Statistics (ONS). Mooney highlights the insane policy of counting a death from any cause as 'Covid-19' if this happens within 28 days of a positive test (with a test not testing for the 'virus') and she points out that ONS statistics reflect deaths 'involving Covid' 'or due to Covid' which meant in practice any death where 'Covid-19' was mentioned on the death certificate. She described the consequences of this fraud:

> Most people will accept the narrative they are fed, so panicky governments here and in Europe witnessed the harsh measures enacted in totalitarian China and jumped into lockdown. Headlines about Covid deaths tolled like the knell that would bring doomsday to us all. Fear stalked our empty streets. Politicians parroted the frankly ridiculous aim of 'zero Covid' and shut down the economy, while most British people agreed that lockdown was essential and (astonishingly to me, as a patriotic Brit) even wanted more restrictions.
>
> For what? Lies on death certificates? Never mind the grim toll of lives ruined, suicides, schools closed, rising inequality, depression, cancelled hospital treatments, cancer patients in a torture of waiting, poverty, economic devastation, loneliness, families kept apart, and so on. How many lives have been lost as a direct result of lockdown?

She said that we could join in a national chorus of shock and horror at reaching the 120,000 death toll which was surely certain to have been totally skewed all along, but what about the human cost of lockdown justified by these 'death figures'? *The British Medical Journal* had reported a 1,493 percent increase in cases of children taken to Great Ormond Street Hospital with abusive head injuries alone and then there was the effect on families:

> Perhaps the most shocking thing about all this is that families have been kept apart – and obeyed the most irrational, changing rules at the whim of government – because they believed in the statistics. They succumbed to fear, which his generation rejected in that war fought for freedom. Dad (God rest his soul) would be angry. And so am I.

Another theme to watch is that in the winter months when there are more deaths from all causes they focus on 'Covid' deaths and in the summer when the British Lung Foundation says respiratory disease plummets by 80 percent they rage on about 'cases'. Either way fascism on population is always the answer.

Nazi eugenics in the 21st century

Elderly people in care homes have been isolated from their families month after lonely month with no contact with relatives and grandchildren who were banned from seeing them. We were told that lockdown fascism was to ‘protect the vulnerable’ like elderly people. At the same time Do Not Resuscitate (DNR) orders were placed on their medical files so that if they needed resuscitation it wasn’t done and ‘Covid-19’ went on their death certificates. Old people were not being ‘protected’ they were being culled – murdered in truth. DNR orders were being decreed for disabled and young people with learning difficulties or psychological problems. The UK Care Quality Commission, a non-departmental body of the Department of Health and Social Care, found that 34 percent of those working in health and social care were pressured into placing ‘do not attempt cardiopulmonary resuscitation’ orders on ‘Covid’ patients who suffered from disabilities and learning difficulties without involving the patient or their families in the decision. UK judges ruled that an elderly woman with dementia should have the DNA-manipulating ‘Covid vaccine’ against her son’s wishes and that a man with severe learning difficulties should have the jab despite his family’s objections. Never mind that many had already died. The judiciary always supports doctors and government in fascist dictatorships. They wouldn’t dare do otherwise. A horrific video was posted showing fascist officers from Los Angeles police forcibly giving the ‘Covid’ shot to women with special needs who were screaming that they didn’t want it. The same fascists are seen giving the jab to a sleeping elderly woman in a care home. This is straight out of the Nazi playbook. Hitler’s Nazis committed mass murder of the mentally ill and physically disabled throughout Germany and occupied territories in the programme that became known as Aktion T4, or just T4. Sabbatian-controlled Hitler and his grotesque crazies set out to kill those they considered useless and unnecessary. The Reich Committee for the Scientific Registering of Hereditary and Congenital Illnesses registered the births of babies identified by physicians to have ‘defects’. By 1941 alone more than 5,000 children were murdered by the state and it is estimated that in total the number of innocent people killed in Aktion T4 was between 275,000 and 300,000. Parents were told their children had been sent away for ‘special treatment’ never to return. It is rather pathetic to see claims about plans for new extermination camps being dismissed today when the same force behind current events did precisely that 80 years ago. Margaret Sanger was a Cult operative who used 'birth control' to sanitise her programme of eugenics. Organisations she founded became what is now Planned Parenthood. Sanger proposed that 'the whole dysgenic population would have its choice of segregation or sterilization'. These included epileptics, 'feeble-minded', and prostitutes. Sanger opposed charity because it perpetuated 'human waste'. She reveals the Cult mentality and if anyone thinks that extermination camps are a 'conspiracy theory' their naivety is touching if breathtakingly stupid.

If you don't believe that doctors can act with callous disregard for their patients it is worth considering that doctors and medical staff agreed to put government-decreed DNR orders on medical files and do nothing when resuscitation is called for. I don't know what you call such people in your house. In mine they are Nazis from the Josef Mengele School of Medicine. Phenomenal numbers of old people have died worldwide from the effects of lockdown, depression, lack of treatment, the 'vaccine' (more later) and losing the will to live. A common response at the start of the manufactured pandemic was to remove old people from hospital beds and transfer them to nursing homes. The decision would result in a mass cull of elderly people in those homes through lack of treatment – *not* 'Covid'. Care home whistleblowers have told how once the 'Covid' era began doctors would not come to their homes to treat patients and they were begging for drugs like antibiotics that often never came. The most infamous example was ordered by New York governor Andrew Cuomo, brother of a moronic CNN host, who amazingly was given an Emmy Award for his handling of the 'Covid crisis' by the ridiculous Wokers that hand them out. Just how ridiculous could be seen in February, 2021, when a Department of Justice and FBI investigation began into how thousands of old people in New York died in nursing homes after being discharged from hospital to make way for 'Covid' patients on Cuomo's say-so – and how he and his staff covered up these facts. This couldn't have happened to a nicer psychopath. Even then there was a 'Covid' spin. Reports said that thousands of old people who tested positive for 'Covid' in hospital were transferred to nursing homes to both die of 'Covid' and transmit it to others. No – they were in hospital because they were ill and the fact that they tested positive with a test not testing for the 'virus' is irrelevant. They were ill often with respiratory diseases ubiquitous in old people near the end of their lives. Their transfer out of hospital meant that their treatment stopped and many would go on to die.

They're old. Who gives a damn?

I have exposed in the books for decades the Cult plan to cull the world's old people and even to introduce at some point what they call a 'demise pill' which at a certain age everyone would take and be out of here by law. In March, 2021, Spain legalised euthanasia and assisted suicide following the Netherlands, Belgium, Luxembourg and Canada on the Tiptoe to the demise pill. Treatment of old people by many 'care' homes has been a disgrace in the 'Covid' era. There are many, many, caring staff – I know some. There have, however, been legions of stories about callous treatment of old people and their families. Police were called when families came to take their loved ones home in the light of isolation that was killing them. They became prisoners of the state. Care home residents in insane, fascist Ontario, Canada, were not allowed to leave their *room* once the 'Covid' hoax began. UK staff have even wheeled elderly people away from windows where family members were talking with them. Oriana Criscuolo from Stockport in the English North West dropped off some things for her 80-year-old father who has Parkinson's disease and dementia and she wanted to wave to him through a ground-floor window. She was told that was 'illegal'. When she went anyway they closed the curtains in the middle of the day. Oriana said:

> It's just unbelievable. I cannot understand how care home staff – people who are being paid to care – have become so uncaring. Their behaviour is inhumane and cruel. It's beyond belief.

She was right and this was not a one-off. What a way to end your life in such loveless circumstances. UK registered nurse Nicky Millen, a proper old school nurse for 40 years, said that when she started her career care was based on dignity, choice, compassion and empathy. Now she said ‘the things that are important to me have gone out of the window.’ She was appalled that people were dying without their loved ones and saying goodbye on iPads. Nicky described how a distressed 89-year-old lady stroked her face and asked her ‘how many paracetamol would it take to finish me off’. Life was no longer worth living while not seeing her family. Nicky said she was humiliated in front of the ward staff and patients for letting the lady stroke her face and giving her a cuddle. Such is the dehumanisation that the ‘Covid’ hoax has brought to the surface. Nicky worked in care homes where patients told her they were being held prisoner. ‘I want to live until I die’, one said to her. ‘I had a lady in tears because she hadn’t seen her great-grandson.’ Nicky was compassionate old school meeting psychopathic New Normal. She also said she had worked on a ‘Covid’ ward with no ‘Covid’ patients. Jewish writer Shai Held wrote an article in March, 2020, which was headlined ‘The Staggering, Heartless Cruelty Toward the Elderly’. What he described was happening from the earliest days of lockdown. He said ‘the elderly’ were considered a group and not unique individuals (the way of the Woke). Shai Held said:

> Notice how the all-too-familiar rhetoric of dehumanization works: ‘The elderly’ are bunched together as a faceless mass, all of them considered culprits and thus effectively deserving of the suffering the pandemic will inflict upon them. Lost entirely is the fact that the elderly are individual human beings, each with a distinctive face and voice, each with hopes and dreams, memories and regrets, friendships and marriages, loves lost and loves sustained.

‘The elderly’ have become another dehumanised group for which anything goes and for many that has resulted in cold disregard for their rights and their life. The distinctive face that Held talks about is designed to be deleted by masks until everyone is part of a faceless mass.

‘War-zone’ hospitals myth

Again and again medical professionals have told me what was really going on and how hospitals ‘overrun like war zones’ according to the media were virtually empty. The mantra from medical whistleblowers was please don’t use my name or my career is over. Citizen journalists around the world sneaked into hospitals to film evidence exposing the ‘war-zone’ lie. They really *were* largely empty with closed wards and operating theatres. I met a hospital worker in my town on the Isle of Wight during the first lockdown in 2020 who said the only island hospital had never been so quiet. Lockdown was justified by the psychopaths to stop hospitals being overrun. At the same time that the island hospital was near-empty the military arrived here to provide *extra beds*. It was all propaganda to ramp up the fear to ensure compliance with fascism as were never-used temporary hospitals with thousands of beds known as Nightingales and never-used make-shift mortuaries opened by the criminal UK government. A man who helped to install those extra island beds attributed to the army said they were never used and the hospital was empty. Doctors and nurses ‘stood around talking or on their phones, wandering down to us to see what we were doing’. There were no masks or social distancing. He accused the useless local island paper, the *County Press*, of ‘pumping the fear as if our hospital was overrun and we only have one so it should have been’. He described ambulances parked up with crews outside in deck chairs. When his brother called an ambulance he was told there was a two-hour backlog which he called ‘bullshit’. An old lady on the island fell ‘and was in a bad way’, but a caller who rang for an ambulance was told the situation wasn’t urgent enough. Ambulance stations were working under capacity while people would hear ambulances with sirens blaring driving through the streets. When those living near the stations realised what was going on they would follow them as they left, circulated around an urban area with the sirens going, and then came back without stopping. All this was to increase levels of fear and the same goes for the ‘ventilator shortage crisis’ that cost tens of millions for hastily produced ventilators never to be used.

Ambulance crews that agreed to be exploited in this way for fear propaganda might find themselves a mirror. I wish them well with that. Empty hospitals were the obvious consequence of treatment and diagnoses of non-'Covid' conditions cancelled and those involved handed a death sentence. People have been dying at home from undiagnosed and untreated cancer, heart disease and other life-threatening conditions to allow empty hospitals to deal with a 'pandemic' that wasn't happening.

Death of the innocent

'War-zones' have been laying off nursing staff, even doctors where they can. There was no work for them. Lockdown was justified by saving lives and protecting the vulnerable they were actually killing with DNR orders and preventing empty hospitals being 'overrun'. In Britain the mantra of stay at home to 'save the NHS' was everywhere and across the world the same story was being sold when it was all lies. Two California doctors, Dan Erickson and Artin Massihi at Accelerated Urgent Care in Bakersfield, held a news conference in April, 2020, to say that intensive care units in California were 'empty, essentially', with hospitals shutting floors, not treating patients and laying off doctors. The California health system was working at minimum capacity 'getting rid of doctors because we just don't have the volume'. They said that people with conditions such as heart disease and cancer were not coming to hospital out of fear of 'Covid-19'. Their video was deleted by Susan Wojcicki's Cult-owned YouTube after reaching five million views. Florida governor Ron Desantis, who rejected the severe lockdowns of other states and is being targeted for doing so, said that in March, 2020, every US governor was given models claiming they would run out of hospital beds in days. That was never going to happen and the 'modellers' knew it. Deceit can be found at every level of the system. Urgent children's operations were cancelled including fracture repairs and biopsies to spot cancer. Eric Nicholls, a consultant paediatrician, said 'this is obviously concerning and we need to return to normal operating and to increase capacity as soon as possible'. Psychopaths in power were rather less concerned *because* they are psychopaths. Deletion of urgent care and diagnosis has been happening all over the world and how many kids and others have died as a result of the actions of these cold and heartless lunatics dictating 'health' policy? The number must be stratospheric. Richard Sullivan, professor of cancer and global health at King's College London, said people feared 'Covid' more than cancer such was the campaign of fear. 'Years of lost life will be quite dramatic', Sullivan said, with 'a huge amount of avoidable mortality'. Sarah Woolnough, executive director for policy at Cancer Research UK, said there had been a 75 percent drop in urgent referrals to hospitals by family doctors of people with suspected cancer. Sullivan said that 'a lot of services have had to scale back – we've seen a dramatic decrease in the amount of elective cancer surgery'. Lockdown deaths worldwide has been absolutely fantastic with the *New York Post* reporting how data confirmed that 'lockdowns end more lives than they save':

> There was a sharp decline in visits to emergency rooms and an increase in fatal heart attacks because patients didn't receive prompt treatment. Many fewer people were screened for cancer. Social isolation contributed to excess deaths from dementia and Alzheimer's.
>
> Researchers predicted that the social and economic upheaval would lead to tens of thousands of "deaths of despair" from drug overdoses, alcoholism and suicide. As unemployment surged and mental-health and substance-abuse treatment programs were interrupted, the reported levels of anxiety, depression and suicidal thoughts increased dramatically, as did alcohol sales and fatal drug overdoses.

This has been happening while nurses and other staff had so much time on their hands in the 'war-zones' that Tic-Tok dancing videos began appearing across the Internet with medical staff dancing around in empty wards and corridors as people died at home from causes that would normally have been treated in hospital.

Mentions in dispatches

One brave and truth-committed whistleblower was Louise Hampton, a call handler with the UK NHS who made a viral Internet video saying she had done 'fuck all' during the 'pandemic' which was ‘a load of bollocks’. She said that ‘Covid-19’ was rebranded flu and of course she lost her job. This is what happens in the medical and endless other professions now when you tell the truth. Louise filmed inside ‘war-zone’ accident and emergency departments to show they were empty and I mean *empty* as in no one there. The mainstream media could have done the same and blown the gaff on the whole conspiracy. They haven’t to their eternal shame. Not that most ‘journalists’ seem capable of manifesting shame as with the psychopaths they slavishly repeat without question. The relative few who were admitted with serious health problems were left to die alone with no loved ones allowed to see them because of ‘Covid’ rules and they included kids dying without the comfort of mum and dad at their bedside while the evil behind this couldn’t give a damn. It was all good fun to them. A Scottish NHS staff nurse publicly quit in the spring of 2021 saying: ‘I can no longer be part of the lies and the corruption by the government.’ She said hospitals ‘aren’t full, the beds aren’t full, beds have been shut, wards have been shut’. Hospitals were never busy throughout ‘Covid’. The staff nurse said that Nicola Sturgeon, tragically the leader of the Scottish government, was on television saying save the hospitals and the NHS – ‘but the beds are empty’ and ‘we’ve not seen flu, we always see flu every year’. She wrote to government and spoke with her union Unison (the unions are Cult-compromised and *useless*, but nothing changed. Many of her colleagues were scared of losing their jobs if they spoke out as they wanted to. She said nursing staff were being affected by wearing masks all day and ‘my head is splitting every shift from wearing a mask’. The NHS is part of the fascist tyranny and must be dismantled so we can start again with human beings in charge. (Ironically, hospitals were reported to be busier again when official ‘Covid’ cases *fell* in spring/summer of 2021 and many other conditions required treatment at the same time as *the fake vaccine rollout*.)

I will cover the ‘Covid vaccine’ scam in detail later, but it is another indicator of the sickening disregard for human life that I am highlighting here. The DNA-manipulating concoctions do not fulfil the definition of a ‘vaccine’, have never been used on humans before and were given only emergency approval because trials were not completed and they continued using the unknowing public. The result was what a NHS senior nurse with responsibility for ‘vaccine’ procedure said was ‘genocide’. She said the ‘vaccines’ were not ‘vaccines’. They had not been shown to be safe and claims about their effectiveness by drug companies were ‘poetic licence’. She described what was happening as a ‘horrid act of human annihilation’. The nurse said that management had instigated a policy of not providing a Patient Information Leaflet (PIL) before people were ‘vaccinated’ even though health care professionals are supposed to do this according to protocol. Patients should also be told that they are taking part in an ongoing clinical trial. Her challenges to what is happening had seen her excluded from meetings and ridiculed in others. She said she was told to ‘watch my step … or I would find myself surplus to requirements’. The nurse, who spoke anonymously in fear of her career, said she asked her NHS manager why he/she was content with taking part in genocide against those having the ‘vaccines’. The reply was that everyone had to play their part and to ‘put up, shut up, and get it done’. Government was ‘leaning heavily’ on NHS management which was clearly leaning heavily on staff. This is how the global ‘medical’ hierarchy operates and it starts with the Cult and its World Health Organization.

She told the story of a doctor who had the Pfizer jab and when questioned had no idea what was in it. The doctor had never read the literature. We have to stop treating doctors as intellectual giants when so many are moral and medical pygmies. The doctor did not even know that the ‘vaccines’ were not fully approved or that their trials were ongoing. They were, however, asking their patients if they minded taking part in follow-ups for research purposes – yes, the *ongoing clinical trial*. The nurse said the doctor’s ignorance was not rare and she had spoken to a hospital consultant who had the jab without any idea of the background or that the ‘trials’ had not been completed. Nurses and pharmacists had shown the same ignorance.

'My NHS colleagues have forsaken their duty of care, broken their code of conduct – Hippocratic Oath – and have been brainwashed just the same as the majority of the UK public through propaganda …' She said she had not been able to recruit a single NHS colleague, doctor, nurse or pharmacist to stand with her and speak out. Her union had refused to help. She said that if the genocide came to light she would not hesitate to give evidence at a Nuremberg-type trial against those in power who could have affected the outcomes but didn't.

And all for what?

To put the nonsense into perspective let's say the 'virus' does exist and let's go completely crazy and accept that the official manipulated figures for cases and deaths are accurate. *Even then* a study by Stanford University epidemiologist Dr John Ioannidis published on the World Health Organization website produced an average infection to fatality rate of … *0.23 percent!* Ioannidis said: 'If one could sample equally from all locations globally, the median infection fatality rate might even be substantially lower than the 0.23% observed in my analysis.' For healthy people under 70 it was … *0.05 percent*! This compares with the 3.4 percent claimed by the Cult-owned World Health Organization when the hoax was first played and maximum fear needed to be generated. An updated Stanford study in April, 2021, put the 'infection' to 'fatality' rate at just 0.15 percent. Another team of scientists led by Megan O'Driscoll and Henrik Salje studied data from 45 countries and published their findings on the Nature website. For children and young people the figure is so small it virtually does not register although authorities will be hyping dangers to the young when they introduce DNA-manipulating 'vaccines' for children. The O'Driscoll study produced an average infection-fatality figure of 0.003 for children from birth to four; 0.001 for 5 to 14; 0.003 for 15 to 19; and it was still only 0.456 up to 64. To claim that children must be 'vaccinated' to protect them from 'Covid' is an obvious lie and so there must be another reason and there is. What's more the average age of a 'Covid' death is akin to the average age that people die in general. The average age of death in England is about 80 for men and 83 for women. The average age of death from alleged 'Covid' is between 82 and 83. California doctors, Dan Erickson and Artin Massihi, said at their April media conference that projection models of millions of deaths had been 'woefully inaccurate'. They produced detailed figures showing that Californians had a 0.03 chance of dying from 'Covid' based on the number of people who tested positive (with a test not testing for the 'virus'). Erickson said there was a 0.1 percent chance of dying from 'Covid' in the *state* of New York, not just the city, and a 0.05 percent chance in Spain, a centre of 'Covid-19' hysteria at one stage. The Stanford studies supported the doctors' data with fatality rate estimates of 0.23 and 0.15 percent. How close are these figures to my estimate of *zero*? Death-rate figures claimed by the World Health Organization at the start of the hoax were some 15 times higher. The California doctors said there was no justification for lockdowns and the economic devastation they caused. Everything they had ever learned about quarantine was that you quarantine the *sick* and not the healthy. They had never seen this before and it made no medical sense.

Why in the in the light of all this would governments and medical systems the world over say that billions must go under house arrest; lose their livelihood; in many cases lose their mind, their health and their life; force people to wear masks dangerous to health and psychology; make human interaction and even family interaction a criminal offence; ban travel; close restaurants, bars, watching live sport, concerts, theatre, and any activity involving human togetherness and discourse; and closing schools to isolate children from their friends and cause many to commit suicide in acts of hopelessness and despair? The California doctors said lockdown consequences included increased child abuse, partner abuse, alcoholism, depression, and other impacts they were seeing every day. Who would do that to the entire human race if not mentally-ill psychopaths of almost unimaginable extremes like Bill Gates? We must face the reality of what we are dealing with and come out of denial. Fascism and tyranny are made possible only by the target population submitting and acquiescing to fascism and tyranny. The whole of human history shows that to be true. Most people naively and unquestioning believed what they were told about a 'deadly virus' and meekly and weakly submitted to house arrest. Those who didn't believe it – at least in total – still submitted in fear of the consequences of not doing so. For the rest who wouldn't submit draconian fines have been imposed, brutal policing by psychopaths *for* psychopaths, and condemnation from the meek and weak who condemn the Pushbackers on behalf of the very force that has them, too, in its gunsights. 'Pathetic' does not even begin to suffice. Britain's brainless 'Health' Secretary Matt Hancock warned anyone lying to border officials about returning from a list of 'hotspot' countries could face a jail sentence of up to ten years which is more than for racially-aggravated assault, incest and attempting to have sex with a child under 13. Hancock is a lunatic, but he has the state apparatus behind him in a Cult-led chain reaction and the same with UK 'Vaccine Minister' Nadhim Zahawi, a prominent member of the mega-Cult secret society, Le Cercle, which featured in my earlier books. The Cult enforces its will on governments and medical systems; government and medical systems enforce their will on business and police; business enforces its will on staff who enforce it on customers; police enforce the will of the Cult on the population and play their essential part in creating a world of fascist control that their own children and grandchildren will have to live in their entire lives. It is a hierarchical pyramid of imposition and acquiescence and, yes indeedy, of clinical insanity.

Does anyone bright enough to read this book have to ask what the answer is? I think not, but I will reveal it anyway in the fewest of syllables: Tell the psychos and their moronic lackeys to fuck off and let's get on with our lives. We are many – They are few.

CHAPTER SEVEN

War on your mind

One believes things because one has been conditioned to believe them
Aldous Huxley, *Brave New World*

I have described the 'Covid' hoax as a 'Psyop' and that is true in every sense and on every level in accordance with the definition of that term which is psychological warfare. Break down the 'Covid pandemic' to the foundation themes and it is psychological warfare on the human individual and collective mind.

The same can be said for the entire human belief system involving every subject you can imagine. Huxley was right in his contention that people believe what they are conditioned to believe and this comes from the repetition throughout their lives of the same falsehoods. They spew from government, corporations, media and endless streams of 'experts' telling you what the Cult wants you to believe and often believing it themselves (although *far* from always). 'Experts' are rewarded with 'prestigious' jobs and titles and as agents of perceptual programming with regular access to the media. The Cult has to control the narrative – control *information* – or they lose control of the vital, crucial, without-which-they-cannot-prevail public perception of reality. The foundation of that control today is the Internet made possible by the Defense Advanced Research Projects Agency (DARPA), the incredibly sinister technological arm of the Pentagon. The Internet is the result of military technology.

DARPA openly brags about establishing the Internet which has been a long-term project to lasso the minds of the global population. I have said for decades the plan is to control information to such an extreme that eventually no one would see or hear anything that the Cult does not approve. We are closing in on that end with ferocious censorship since the 'Covid' hoax began and in my case it started back in the 1990s in terms of books and speaking venues. I had to create my own publishing company in 1995 precisely because no one else would publish my books even then. I think they're all still running.

Cult Internet

To secure total control of information they needed the Internet in which pre-programmed algorithms can seek out 'unclean' content for deletion and even stop it being posted in the first place. The Cult had to dismantle print and non-Internet broadcast media to ensure the transfer of information to the appropriate-named 'Web' – a critical expression of the *Cult* web. We've seen the ever-quickening demise of traditional media and control of what is left by a tiny number of corporations operating worldwide. Independent journalism in the mainstream is already dead and never was that more obvious than since the turn of 2020. The Cult wants all information communicated via the Internet to globally censor and allow the plug to be pulled any time. Lockdowns and forced isolation has meant that communication between people has been through electronic means and no longer through face-to-face discourse and discussion. Cult psychopaths have targeted the bars, restaurants, sport, venues and meeting places in general for this reason. None of this is by chance and it's to stop people gathering in any kind of privacy or number while being able to track and monitor all Internet communications and block them as necessary. Even private messages between individuals have been censored by these fascists that control Cult fronts like Facebook, Twitter, Google and YouTube which are all officially run by Sabbatian place-people and from the background by higher-level Sabbatian place people.

Facebook, Google, Amazon and their like were seed-funded and supported into existence with money-no-object infusions of funds either directly or indirectly from DARPA and CIA technology arm In-Q-Tel. The Cult plays the long game and prepares very carefully for big plays like 'Covid'. Amazon is another front in the psychological war and pretty much controls the global market in book sales and increasingly publishing. Amazon's limitless funds have deleted fantastic numbers of independent publishers to seize global domination on the way to deciding which books can be sold and circulated and which cannot. Moves in that direction are already happening. Amazon's leading light Jeff Bezos is the grandson of Lawrence Preston Gise who worked with DARPA predecessor ARPA. Amazon has big connections to the CIA and the Pentagon. The plan I have long described went like this:

1. Employ military technology to establish the Internet.

2. Sell the Internet as a place where people can freely communicate without censorship and allow that to happen until the Net becomes the central and irreversible pillar of human society. If the Internet had been highly censored from the start many would have rejected it.

3. Fund and manipulate major corporations into being to control the circulation of information on your Internet using cover stories about geeks in garages to explain how they came about. Give them unlimited funds to expand rapidly with no need to make a profit for years while non-Cult companies who need to balance the books cannot compete. You know that in these circumstances your Googles, YouTubes, Facebooks and Amazons are going to secure near monopolies by either crushing or buying up the opposition.

4. Allow freedom of expression on both the Internet and communication platforms to draw people in until the Internet is the central and irreversible pillar of human society and your communication corporations have reached a stage of near monopoly domination.

5. Then unleash your always-planned frenzy of censorship on the basis of 'where else are you going to go?' and continue to expand that until nothing remains that the Cult does not want its human targets to see.

The process was timed to hit the 'Covid' hoax to ensure the best chance possible of controlling the narrative which they knew they had to do at all costs. They were, after all, about to unleash a 'deadly virus' that didn't really exist. If you do that in an environment of free-flowing information and opinion you would be dead in the water before you could say Gates is a psychopath. The network was in place through which the Cult-created-and-owned World Health Organization could dictate the 'Covid' narrative and response policy slavishly supported by Cult-owned Internet communication giants and mainstream media while those telling a different story were censored. Google, YouTube, Facebook and Twitter openly announced that they would do this. What else would we expect from Cult-owned operations like Facebook which former executives have confirmed set out to make the platform more addictive than cigarettes and coldly manipulates emotions of its users to sow division between people and groups and scramble the minds of the young? If Zuckerberg lives out the rest of his life without going to jail for crimes against humanity, and most emphatically against the young, it will be a travesty of justice. Still, no matter, cause and effect will catch up with him eventually and the same with Sergey Brin and Larry Page at Google with its CEO Sundar Pichai who fix the Google search results to promote Cult narratives and hide the opposition. Put the same key words into Google and other search engines like DuckDuckGo and you will see how different results can be. Wikipedia is another intensely biased 'encyclopaedia' which skews its content to the Cult agenda. YouTube links to Wikipedia's version of 'Covid' and 'climate change' on video pages in which experts in their field offer a different opinion (even that is increasingly rare with Wojcicki censorship). Into this 'Covid' silence-them network must be added government media censors, sorry 'regulators', such as Ofcom in the UK which imposed tyrannical restrictions on British broadcasters that had the effect of banning me from ever appearing. Just to debate with me about my evidence and views on 'Covid' would mean breaking the fascistic impositions of Ofcom and its CEO career government bureaucrat Melanie Dawes. Gutless British broadcasters tremble at the very thought of fascist Ofcom.

Psychos behind 'Covid'

The reason for the ‘Covid’ catastrophe in all its facets and forms can be seen by whom and what is driving the policies worldwide in such a coordinated way. Decisions are not being made to protect health, but to target psychology. The dominant group guiding and ‘advising’ government policy are not medical professionals. They are psychologists and behavioural scientists. Every major country has its own version of this phenomenon and I’ll use the British example to show how it works. In many ways the British version has been affecting the wider world in the form of the huge behaviour manipulation network in the UK which operates in other countries. The network involves private companies, government, intelligence and military. The Cabinet Office is at the centre of the government ‘Covid’ Psyop and part-owns, with ‘innovation charity’ Nesta, the Behavioural Insights Team (BIT) which claims to be independent of government but patently isn’t. The BIT was established in 2010 and its job is to manipulate the psyche of the population to acquiesce to government demands and so much more. It is also known as the ‘Nudge Unit’, a name inspired by the 2009 book by two ultra-Zionists, Cass Sunstein and Richard Thaler, called *Nudge: Improving Decisions About Health, Wealth, and Happiness*. The book, as with the Behavioural Insights Team, seeks to ‘nudge’ behaviour (manipulate it) to make the public follow patterns of action and perception that suit those in authority (the Cult). Sunstein is so skilled at this that he advises the World Health Organization and the UK Behavioural Insights Team and was Administrator of the White House Office of Information and Regulatory Affairs in the Obama administration. Biden appointed him to the Department of Homeland Security – another ultra-Zionist in the fold to oversee new immigration laws which is another policy the Cult wants to control. Sunstein is desperate to silence anyone exposing conspiracies and co-authored a 2008 report on the subject in which suggestions were offered to ban ‘conspiracy theorizing’ or impose ‘some kind of tax, financial or otherwise, on those who disseminate such theories’. I guess a psychiatrist’s chair is out of the question?

Sunstein's mate Richard Thaler, an 'academic affiliate' of the UK Behavioural Insights Team, is a proponent of 'behavioural economics' which is defined as the study of 'the effects of psychological, cognitive, emotional, cultural and social factors on the decisions of individuals and institutions'. Study the effects so they can be manipulated to be what you want them to be. Other leading names in the development of behavioural economics are ultra-Zionists Daniel Kahneman and Robert J. Shiller and they, with Thaler, won the Nobel Memorial Prize in Economic Sciences for their work in this field. The Behavioural Insights Team is operating at the heart of the UK government and has expanded globally through partnerships with several universities including Harvard, Oxford, Cambridge, University College London (UCL) and Pennsylvania. They claim to have 'trained' (reframed) 20,000 civil servants and run more than 750 projects involving 400 randomised controlled trials in dozens of countries' as another version of mind reframers Common Purpose. BIT works from its office in New York with cities and their agencies, as well as other partners, across the United States and Canada – this is a company part-owned by the British government Cabinet Office. An executive order by President Cult-servant Obama established a US Social and Behavioral Sciences Team in 2015. They all have the same reason for being and that's to brainwash the population directly and by brainwashing those in positions of authority.

'Covid' mind game

Another prime aspect of the UK mind-control network is the 'independent' [joke] Scientific Pandemic Insights Group on Behaviours (SPI-B) which 'provides behavioural science advice aimed at anticipating and helping people adhere to interventions that are recommended by medical or epidemiological experts'. That means manipulating public perception and behaviour to do whatever government tells them to do. It's disgusting and if they really want the public to be 'safe' this lot should all be under lock and key. According to the government website SPI-B consists of ‘behavioural scientists, health and social psychologists, anthropologists and historians’ and advises the Whitty-Vallance-led Scientific Advisory Group for Emergencies (SAGE) which in turn advises the government on ‘the science’ (it doesn’t) and ‘Covid’ policy. When politicians say they are being guided by ‘the science’ this is the rabble in each country they are talking about and that ‘science’ is dominated by behaviour manipulators to enforce government fascism through public compliance. The Behaviour Insight Team is headed by psychologist David Solomon Halpern, a visiting professor at King’s College London, and connects with a national and global web of other civilian and military organisations as the Cult moves towards its goal of fusing them into one fascistic whole in every country through its ‘Fusion Doctrine’. The behaviour manipulation network involves, but is not confined to, the Foreign Office; National Security Council; government communications headquarters (GCHQ); MI5; MI6; the Cabinet Office-based Media Monitoring Unit; and the Rapid Response Unit which ‘monitors digital trends to spot emerging issues; including misinformation and disinformation; and identifies the best way to respond’.

There is also the 77th Brigade of the UK military which operates like the notorious Israeli military’s Unit 8200 in manipulating information and discussion on the Internet by posing as members of the public to promote the narrative and discredit those who challenge it. Here we have the military seeking to manipulate *domestic* public opinion while the Nazis in government are fine with that. Conservative Member of Parliament Tobias Ellwood, an advocate of lockdown and control through ‘vaccine passports’, is a Lieutenant Colonel reservist in the 77th Brigade which connects with the military operation jHub, the ‘innovation centre’ for the Ministry of Defence and Strategic Command. jHub has also been involved with the civilian National Health Service (NHS) in ‘symptom tracing’ the population. The NHS is a key part of this mind control network and produced a document in December, 2020, explaining to staff how to use psychological manipulation with different groups and ages to get them to have the DNA-manipulating ‘Covid vaccine’ that’s designed to cumulatively rewrite human genetics. The document, called ‘Optimising Vaccination Roll Out – Do’s and Dont’s for all messaging, documents and “communications” in the widest sense’, was published by NHS England and the NHS Improvement *Behaviour Change Unit* in partnership with Public Health England and Warwick Business School. I hear the mantra about ‘save the NHS’ and ‘protect the NHS’ when we need to scrap the NHS and start again. The current version is far too corrupt, far too anti-human and totally compromised by Cult operatives and their assets. UK government broadcast media censor Ofcom will connect into this web – as will the BBC with its tremendous Ofcom influence – to control what the public see and hear and dictate mass perception. Nuremberg trials must include personnel from all these organisations.

The fear factor

The ‘Covid’ hoax has led to the creation of the UK Cabinet Office-connected Joint Biosecurity Centre (JBC) which is officially described as providing ‘expert advice on pandemics’ using its independent [all Cult operations are ‘independent’] analytical function to provide real-time analysis about infection outbreaks to identify and respond to outbreaks of Covid-19’. Another role is to advise the government on a response to spikes in infections – ‘for example by closing schools or workplaces in local areas where infection levels have risen’. Put another way, promoting the Cult agenda. The Joint Biosecurity Centre is modelled on the Joint Terrorism Analysis Centre which analyses intelligence to set ‘terrorism threat levels’ and here again you see the fusion of civilian and military operations and intelligence that has led to military intelligence producing documents about ‘vaccine hesitancy’ and how it can be combated. Domestic civilian matters and opinions should not be the business of the military. The Joint Biosecurity Centre is headed by Tom Hurd, director general of the Office for Security and Counter-Terrorism from the establishment-to-its-fingertips Hurd family. His father is former Foreign Secretary Douglas Hurd. How coincidental that Tom

Hurd went to the elite Eton College and Oxford University with Boris Johnson. Imperial College with its ridiculous computer modeller Neil Ferguson will connect with this gigantic web that will itself interconnect with similar set-ups in other major and not so major countries. Compared with this Cult network the politicians, be they Boris Johnson, Donald Trump or Joe Biden, are bit-part players ‘following the science’. The network of psychologists was on the ‘Covid’ case from the start with the aim of generating maximum fear of the ‘virus’ to ensure compliance by the population. A government behavioural science group known as SPI-B produced a paper in March, 2020, for discussion by the main government science advisory group known as SAGE. It was headed ‘Options for increasing adherence to social distancing measures’ and it said the following in a section headed ‘Persuasion’:

- A substantial number of people still do not feel sufficiently personally threatened; it could be that they are reassured by the low death rate in their demographic group, although levels of concern may be rising. Having a good understanding of the risk has been found to be positively associated with adoption of COVID-19 social distancing measures in Hong Kong.

- The perceived level of personal threat needs to be increased among those who are complacent, using hard-hitting evaluation of options for increasing social distancing emotional messaging. To be effective this must also empower people by making clear the actions they can take to reduce the threat.

- Responsibility to others: There seems to be insufficient understanding of, or feelings of responsibility about, people’s role in transmitting the infection to others … Messaging about actions need to be framed positively in terms of protecting oneself and the community, and increase confidence that they will be effective.

- Some people will be more persuaded by appeals to play by the rules, some by duty to the community, and some to personal risk.

> All these different approaches are needed. The messaging also needs to take account of the realities of different people's lives. Messaging needs to take account of the different motivational levers and circumstances of different people.

All this could be achieved the SPI-B psychologists said by *using the media to increase the sense of personal threat* which translates as terrify the shit out of the population, including children, so they all do what we want. That's not happened has it? Those excuses for 'journalists' who wouldn't know journalism if it bit them on the arse (the great majority) have played their crucial part in serving this Cult-government Psyop to enslave their own kids and grandkids. How they live with themselves I have no idea. The psychological war has been underpinned by constant government 'Covid' propaganda in almost every television and radio ad break, plus the Internet and print media, which has pounded out the fear with taxpayers footing the bill for their own programming. The result has been people terrified of a 'virus' that doesn't exist or one with a tiny fatality rate even if you believe it does. People walk down the street and around the shops wearing face-nappies damaging their health and psychology while others report those who refuse to be that naïve to the police who turn up in their own face-nappies. I had a cameraman come to my flat and he was so frightened of 'Covid' he came in wearing a mask and refused to shake my hand in case he caught something. He had – naïveitis – and the thought that he worked in the mainstream media was both depressing and made his behaviour perfectly explainable. The fear which has gripped the minds of so many and frozen them into compliance has been carefully cultivated by these psychologists who are really psychopaths. If lives get destroyed and a lot of young people commit suicide it shows our plan is working. SPI-B then turned to compulsion on the public to comply. 'With adequate preparation, rapid change can be achieved', it said. Some countries had introduced mandatory self-isolation on a wide scale without evidence of major public unrest and a large majority of the UK's population appeared to be supportive of more coercive measures with 64 percent of adults saying they would support putting London under a lockdown (watch the 'polls' which are designed to make people believe that public opinion is in favour or against whatever the subject in hand).

For 'aggressive protective measures' to be effective, the SPI-B paper said, special attention should be devoted to those population groups that are more at risk. Translated from the Orwellian this means making the rest of population feel guilty for not protecting the 'vulnerable' such as old people which the Cult and its agencies were about to kill on an industrial scale with lockdown, lack of treatment and the Gates 'vaccine'. Psychopath psychologists sold their guilt-trip so comprehensively that Los Angeles County Supervisor Hilda Solis reported that children were apologising (from a distance) to their parents and grandparents for bringing 'Covid' into their homes and getting them sick. '… These apologies are just some of the last words that loved ones will ever hear as they die alone,' she said. Gut-wrenchingly Solis then used this childhood tragedy to tell children to stay at home and 'keep your loved ones alive'. Imagine heaping such potentially life-long guilt on a kid when it has absolutely nothing to do with them. These people are deeply disturbed and the psychologists behind this even more so.

Uncivil war – divide and rule

Professional mind-controllers at SPI-B wanted the media to increase a sense of responsibility to others (do as you're told) and promote 'positive messaging' for those actions while in contrast to invoke 'social disapproval' by the unquestioning, obedient, community of anyone with a mind of their own. Again the compliant Goebbels-like media obliged. This is an old, old, trick employed by tyrannies the world over throughout human history. You get the target population to keep the target population in line – *your* line. SPI-B said this could 'play an important role in preventing anti-social behaviour or discouraging failure to enact pro-social behaviour'. For 'anti-social' in the Orwellian parlance of SPI-B see any behaviour that government doesn't approve. SPI-B recommendations said that 'social disapproval' should be accompanied by clear messaging and promotion of strong collective identity – hence the government and celebrity mantra of ‘we’re all in this together’. Sure we are. The mind doctors have such contempt for their targets that they think some clueless comedian, actor or singer telling them to do what the government wants will be enough to win them over. We have had UK comedian Lenny Henry, actor Michael Caine and singer Elton John wheeled out to serve the propagandists by urging people to have the DNA-manipulating ‘Covid’ non-‘vaccine’. The role of Henry and fellow black celebrities in seeking to coax a ‘vaccine’ reluctant black community into doing the government’s will was especially stomach-turning. An emotion-manipulating script and carefully edited video featuring these black ‘celebs’ was such an insult to the intelligence of black people and where’s the self-respect of those involved selling their souls to a fascist government agenda? Henry said he heard black people’s ‘legitimate worries and concerns’, but people must ‘trust the facts’ when they were doing exactly that by not having the ‘vaccine’. They had to include the obligatory reference to Black Lives Matter with the line … ‘Don’t let coronavirus cost even more black lives – because we matter’. My god, it was pathetic. ‘I know the vaccine is safe and what it does.’ How? ‘I’m a comedian and it says so in my script.’

SPI-B said social disapproval needed to be carefully managed to avoid victimisation, scapegoating and misdirected criticism, but they knew that their ‘recommendations’ would lead to exactly that and the media were specifically used to stir-up the divide-and-conquer hostility. Those who conform like good little baa, baas, are praised while those who have seen through the tidal wave of lies are ‘Covidiots’. The awake have been abused by the fast asleep for not conforming to fascism and impositions that the awake know are designed to endanger their health, dehumanise them, and tear asunder the very fabric of human society. We have had the curtain-twitchers and morons reporting neighbours and others to the face-nappied police for breaking ‘Covid rules’ with fascist police delighting in posting links and phone numbers where this could be done. The Cult cannot impose its will without a compliant police and military or a compliant population willing to play their part in enslaving themselves and their kids. The words of a pastor in Nazi Germany are so appropriate today:

> First they came for the socialists and I did not speak out because I was not a socialist.
>
> Then they came for the trade unionists and I did not speak out because I was not a trade unionist.
>
> Then they came for the Jews and I did not speak out because I was not a Jew.
>
> Then they came for me and there was no one left to speak for me.

Those who don't learn from history are destined to repeat it and so many are.

'Covid' rules: Rewiring the mind

With the background laid out to this gigantic national and global web of psychological manipulation we can put 'Covid' rules into a clear and sinister perspective. Forget the claims about protecting health. 'Covid' rules are about dismantling the human mind, breaking the human spirit, destroying self-respect, and then putting Humpty Dumpty together again as a servile, submissive slave. Social isolation through lockdown and distancing have devastating effects on the human psyche as the psychological psychopaths well know and that's the real reason for them. Humans need contact with each other, discourse, closeness and touch, or they eventually, and literarily, go crazy. Masks, which I will address at some length, fundamentally add to the effects of isolation and the Cult agenda to dehumanise and de-individualise the population. To do this while knowing – in fact *seeking* – this outcome is the very epitome of evil and psychologists involved in this *are* the epitome of evil. They must like all the rest of the Cult demons and their assets stand trial for crimes against humanity on a scale that defies the imagination. Psychopaths in uniform use isolation to break enemy troops and agents and make them subservient and submissive to tell what they know. The technique is rightly considered a form of torture and torture is most certainly what has been imposed on the human population.

Clinically-insane American psychologist Harry Harlow became famous for his isolation experiments in the 1950s in which he separated baby monkeys from their mothers and imprisoned them for months on end in a metal container or ‘pit of despair’. They soon began to show mental distress and depression as any idiot could have predicted. Harlow put other monkeys in steel chambers for three, six or twelve months while denying them any contact with animals or humans. He said that the effects of total social isolation for six months were ‘so devastating and debilitating that we had assumed initially that twelve months of isolation would not produce any additional decrement’; but twelve months of isolation ‘almost obliterated the animals socially’. This is what the Cult and its psychopaths are doing to you and your children. Even monkeys in partial isolation in which they were not allowed to form relationships with other monkeys became ‘aggressive and hostile, not only to others, but also towards their own bodies’. We have seen this in the young as a consequence of lockdown. UK government psychopaths launched a public relations campaign telling people not to hug each other even after they received the ‘Covid-19 vaccine’ which we were told with more lies would allow a return to ‘normal life’. A government source told *The Telegraph*: ‘It will be along the lines that it is great that you have been vaccinated, but if you are going to visit your family and hug your grandchildren there is a chance you are going to infect people you love.’ The source was apparently speaking from a secure psychiatric facility. Janet Lord, director of Birmingham University’s Institute of Inflammation and Ageing, said that parents and grandparents should avoid hugging their children. Well, how can I put it, Ms Lord? Fuck off. Yep, that’ll do.

Destroying the kids – where are the parents?

Observe what has happened to people enslaved and isolated by lockdown as suicide and self-harm has soared worldwide, particularly among the young denied the freedom to associate with their friends. A study of 49,000 people in English-speaking countries concluded that almost half of young adults are at clinical risk of mental health disorders. A national survey in America of 1,000 currently enrolled high school and college students found that 5 percent reported attempting suicide during the pandemic. Data from the US CDC's National Syndromic Surveillance Program from January 1st to October 17th, 2020, revealed a *31 percent* increase in mental health issues among adolescents aged 12 to 17 compared with 2019. The CDC reported that America in general suffered the biggest drop in life expectancy since World War Two as it fell by a year in the first half of 2020 as a result of 'deaths of despair' – overdoses and suicides. Deaths of despair have leapt by more than 20 percent during lockdown and include the highest number of fatal overdoses ever recorded in a single year – 81,000. Internet addiction is another consequence of being isolated at home which lowers interest in physical activities as kids fall into inertia and what's the point? Children and young people are losing hope and giving up on life, sometimes literally. A 14-year-old boy killed himself in Maryland because he had 'given up' when his school district didn't reopen; an 11-year-old boy shot himself during a zoom class; a teenager in Maine succumbed to the isolation of the 'pandemic' when he ended his life after experiencing a disrupted senior year at school. Children as young as nine have taken their life and all these stories can be repeated around the world. Careers are being destroyed before they start and that includes those in sport in which promising youngsters have not been able to take part. The plan of the psycho-psychologists is working all right. Researchers at Cambridge University found that lockdowns cause significant harm to children's mental health. Their study was published in the *Archives of Disease in Childhood*, and followed 168 children aged between 7 and 11. The researchers concluded:

> During the UK lockdown, children's depression symptoms have increased substantially, relative to before lockdown. The scale of this effect has direct relevance for the continuation of different elements of lockdown policy, such as complete or partial school closures …

> … Specifically, we observed a statistically significant increase in ratings of depression, with a medium-to-large effect size. Our findings emphasise the need to incorporate the potential impact of lockdown on child mental health in planning the ongoing response to the global pandemic and the recovery from it.

Not a chance when the Cult's psycho-psychologists were getting exactly what they wanted. The UK's Royal College of Paediatrics and Child Health has urged parents to look for signs of eating disorders in children and young people after a three to four fold increase. Specialists say the 'pandemic' is a major reason behind the rise. You don't say. The College said isolation from friends during school closures, exam cancellations, loss of extra-curricular activities like sport, and an increased use of social media were all contributory factors along with fears about the virus (psycho-psychologists again), family finances, and students being forced to quarantine. Doctors said young people were becoming severely ill by the time they were seen with 'Covid' regulations reducing face-to-face consultations. Nor is it only the young that have been devastated by the psychopaths. Like all bullies and cowards the Cult is targeting the young, elderly, weak and infirm. A typical story was told by a British lady called Lynn Parker who was not allowed to visit her husband in 2020 for the last ten and half months of his life 'when he needed me most' between March 20th and when he died on December 19th. This vacates the criminal and enters the territory of evil. The emotional impact on the immune system alone is immense as are the number of people of all ages worldwide who have died as a result of Cult-demanded, Gates-demanded, lockdowns.

Isolation is torture

The experience of imposing solitary confinement on millions of prisoners around the world has shown how a large percentage become 'actively psychotic and/or acutely suicidal'. Social isolation has been found to trigger 'a specific psychiatric syndrome, characterized by hallucinations; panic attacks; overt paranoia; diminished impulse control; hypersensitivity to external stimuli; and difficulties with thinking, concentration and memory'. Juan Mendez, a United Nations rapporteur (investigator), said that isolation is a form of torture. Research has shown that even after isolation prisoners find it far more difficult to make social connections and I remember chatting to a shop assistant after one lockdown who told me that when her young son met another child again he had no idea how to act or what to do. Hannah Flanagan, Director of Emergency Services at Journey Mental Health Center in Dane County, Wisconsin, said: 'The specificity about Covid social distancing and isolation that we've come across as contributing factors to the suicides are really new to us this year.' But they are not new to those that devised them. They are getting the effect they want as the population is psychologically dismantled to be rebuilt in a totally different way. Children and the young are particularly targeted. They will be the adults when the full-on fascist AI-controlled technocracy is planned to be imposed and they are being prepared to meekly submit. At the same time older people who still have a memory of what life was like before – and how fascist the new normal really is – are being deleted. You are going to see efforts to turn the young against the old to support this geriatric genocide. Hannah Flanagan said the big increase in suicide in her county proved that social isolation is not only harmful, but deadly. Studies have shown that isolation from others is one of the main risk factors in suicide and even more so with women. Warnings that lockdown could create a 'perfect storm' for suicide were ignored. After all this was one of the *reasons* for lockdown. Suicide, however, is only the most extreme of isolation consequences. There are many others. Dr Dhruv Khullar, assistant professor of healthcare policy at Weill Cornell Medical College, said in a *New York Times* article in 2016 long before the fake 'pandemic':

> A wave of new research suggests social separation is bad for us. Individuals with less social connection have disrupted sleep patterns, altered immune systems, more inflammation and higher levels of stress hormones. One recent study found that isolation increases the risk of heart disease by 29 percent and stroke by 32 percent. Another analysis that pooled data from 70 studies and 3.4 million people found that socially isolated individuals had a 30 percent higher risk of dying in the next seven years, and that this effect was largest in middle age.

> Loneliness can accelerate cognitive decline in older adults, and isolated individuals are twice as likely to die prematurely as those with more robust social interactions. These effects start early: Socially isolated children have significantly poorer health 20 years later, even after controlling for other factors. All told, loneliness is as important a risk factor for early death as obesity and smoking.

There you have proof from that one article alone four years before 2020 that those who have enforced lockdown, social distancing and isolation knew what the effect would be and that is even more so with professional psychologists that have been driving the policy across the globe. We can go back even further to the years 2000 and 2003 and the start of a major study on the effects of isolation on health by Dr Janine Gronewold and Professor Dirk M. Hermann at the University Hospital in Essen, Germany, who analysed data on 4,316 people with an average age of 59 who were recruited for the long-term research project. They found that socially isolated people are more than 40 percent more likely to have a heart attack, stroke, or other major cardiovascular event and nearly 50 percent more likely to die from any cause. Given the financial Armageddon unleashed by lockdown we should note that the study found a relationship between increased cardiovascular risk and lack of financial support. After excluding other factors social isolation was still connected to a 44 percent increased risk of cardiovascular problems and a 47 percent increased risk of death by any cause. Lack of financial support was associated with a 30 percent increase in the risk of cardiovascular health events. Dr Gronewold said it had been known for some time that feeling lonely or lacking contact with close friends and family can have an impact on physical health and the study had shown that having strong social relationships is of high importance for heart health. Gronewold said they didn't understand yet why people who are socially isolated have such poor health outcomes, but this was obviously a worrying finding, particularly during these times of prolonged social distancing. Well, it can be explained on many levels. You only have to identify the point in the body where people feel loneliness and missing people they are parted from – it's in the centre of the chest where they feel the ache of loneliness and the ache of missing people. 'My heart aches for you' … 'My heart aches for some company.' I will explain this more in the chapter Escaping Wetiko, but when you realise that the body is the mind – they are expressions of each other – the reason why state of the mind dictates state of the body becomes clear.

American psychologist Ranjit Powar was highlighting the effects of lockdown isolation as early as April, 2020. She said humans have evolved to be social creatures and are wired to live in interactive groups. Being isolated from family, friends and colleagues could be unbalancing and traumatic for most people and could result in short or even long-term psychological and physical health problems. An increase in levels of anxiety, aggression, depression, forgetfulness and hallucinations were possible psychological effects of isolation. 'Mental conditions may be precipitated for those with underlying pre-existing susceptibilities and show up in many others without any pre-condition.' Powar said personal relationships helped us cope with stress and if we lost this outlet for letting off steam the result can be a big emotional void which, for an average person, was difficult to deal with. 'Just a few days of isolation can cause increased levels of anxiety and depression' – so what the hell has been the effect on the global population of *18 months* of this at the time of writing? Powar said: 'Add to it the looming threat of a dreadful disease being repeatedly hammered in through the media and you have a recipe for many shades of mental and physical distress.' For those with a house and a garden it is easy to forget that billions have had to endure lockdown isolation in tiny overcrowded flats and apartments with nowhere to go outside. The psychological and physical consequences of this are unimaginable and with lunatic and abusive partners and parents the consequences have led to tremendous increases in domestic and child abuse and alcoholism as people seek to shut out the horror. Ranjit Powar said:

> Staying in a confined space with family is not all a rosy picture for everyone. It can be extremely oppressive and claustrophobic for large low-income families huddled together in small single-room houses. Children here are not lucky enough to have many board/electronic games or books to keep them occupied.

> Add to it the deep insecurity of running out of funds for food and basic necessities. On the other hand, there are people with dysfunctional family dynamics, such as domineering, abusive or alcoholic partners, siblings or parents which makes staying home a period of trial. Incidence of suicide and physical abuse against women has shown a worldwide increase. Heightened anxiety and depression also affect a person's immune system, making them more susceptible to illness.

To think that Powar's article was published on April 11th, 2020.

Six-feet fantasy

Social (unsocial) distancing demanded that people stay six feet or two metres apart. UK government advisor Robert Dingwall from the New and Emerging Respiratory Virus Threats Advisory Group said in a radio interview that the two-metre rule was 'conjured up out of nowhere' and was not based on science. No, it was not based on *medical* science, but it didn't come out of nowhere. The distance related to *psychological* science. Six feet/two metres was adopted in many countries and we were told by people like the criminal Anthony Fauci and his ilk that it was founded on science. Many schools could not reopen because they did not have the space for six-feet distancing. Then in March, 2021, after a year of six-feet 'science', a study published in the *Journal of Infectious Diseases* involving more than 500,000 students and almost 100,000 staff over 16 weeks revealed no significant difference in 'Covid' cases between six feet and three feet and Fauci changed his tune. Now three feet was okay. There is no difference between six feet and three *inches* when there is no 'virus' and they got away with six feet for psychological reasons for as long as they could. I hear journalists and others talk about 'unintended consequences' of lockdown. They are not *unintended* at all; they have been coldly-calculated for a specific outcome of human control and that's why super-psychopaths like Gates have called for them so vehemently. Super-psychopath psychologists have demanded them and psychopathic or clueless, spineless, politicians have gone along with them by 'following the science'. But it's not science at all. 'Science' is not what is; it's only what people can be manipulated to believe it is. The whole 'Covid' catastrophe is founded on mind control. Three word or three statement mantras issued by the UK government are a well-known mind control technique and so we've had 'Stay home/protect the NHS/save lives', 'Stay alert/control the virus/save lives' and 'hands/face/space'. One of the most vocal proponents of extreme 'Covid' rules in the UK has been Professor Susan Michie, a member of the British Communist Party, who is not a medical professional. Michie is the director of the Centre for Behaviour Change at University College London. She is a *behavioural psychologist* and another filthy rich 'Marxist' who praised China's draconian lockdown. She was known by fellow students at Oxford University as 'Stalin's nanny' for her extreme Marxism. Michie is an influential member of the UK government's Scientific Advisory Group for Emergencies (SAGE) and behavioural manipulation groups which have dominated 'Covid' policy. She is a consultant adviser to the World Health Organization on 'Covid-19' and behaviour. Why the hell are lockdowns anything to do with her when they are claimed to be about health? Why does a behavioural psychologist from a group charged with changing the behaviour of the public want lockdown, human isolation and mandatory masks? Does that question really need an answer? Michie *absolutely* has to explain herself before a Nuremberg court when humanity takes back its world again and even more so when you see the consequences of masks that she demands are compulsory. This is a Michie classic:

> The benefits of getting primary school children to wear masks is that regardless of what little degree of transmission is occurring in those age groups it could help normalise the practice. Young children wearing masks may be more likely to get their families to accept masks.

Those words alone should carry a prison sentence when you ponder on the callous disregard for children involved and what a statement it makes about the mind and motivations of Susan Michie. What a lovely lady and what she said there encapsulates the mentality of the psychopaths behind the 'Covid' horror. Let us compare what Michie said with a countrywide study in Germany published at researchsquare.com involving 25,000 school children and 17,854 health complaints submitted by parents. Researchers found that masks are harming children physically, psychologically, and behaviourally with 24 health issues associated with mask wearing. They include: shortness of breath (29.7%); dizziness (26.4%); increased headaches (53%); difficulty concentrating (50%); drowsiness or fatigue (37%); and malaise (42%). Nearly a third of children experienced more sleep issues than before and a quarter developed new fears. Researchers found health issues and other impairments in 68 percent of masked children covering their faces for an average of 4.5 hours a day. Hundreds of those taking part experienced accelerated respiration, tightness in the chest, weakness, and short-term impairment of consciousness. A reminder of what Michie said again:

> The benefits of getting primary school children to wear masks is that regardless of what little degree of transmission is occurring in those age groups it could help normalise the practice. Young children wearing masks may be more likely to get their families to accept masks.

Psychopaths in government and psychology now have children and young people – plus all the adults – wearing masks for hours on end while clueless teachers impose the will of the psychopaths on the young they should be protecting. What the hell are parents doing?

Cult lab rats

We have some schools already imposing on students microchipped buzzers that activate when they get ‘too close’ to their pals in the way they do with lab rats. How apt. To the Cult and its brain-dead servants our children *are* lab rats being conditioned to be unquestioning, dehumanised slaves for the rest of their lives. Children and young people are being weaned and frightened away from the most natural human instincts including closeness and touch. I have tracked in the books over the years how schools were banning pupils from greeting each other with a hug and the whole Cult-induced Me Too movement has terrified men and boys from a relaxed and natural interaction with female friends and work colleagues to the point where many men try never to be in a room alone with a woman that's not their partner. Airhead celebrities have as always played their virtue-signalling part in making this happen with their gross exaggeration. For every monster like Harvey Weinstein there are at least tens of thousands of men that don't treat women like that; but everyone must be branded the same and policy changed for them as well as the monster. I am going to be using the word 'dehumanise' many times in this chapter because that is what the Cult is seeking to do and it goes very deep as we shall see. Don't let them kid you that social distancing is planned to end one day. That's not the idea. We are seeing more governments and companies funding and producing wearable gadgets to keep people apart and they would not be doing that if this was meant to be short-term. A tech start-up company backed by GCHQ, the British Intelligence and military surveillance headquarters, has created a social distancing wrist sensor that alerts people when they get too close to others. The CIA has also supported tech companies developing similar devices. The wearable sensor was developed by Tended, one of a number of start-up companies supported by GCHQ (see the CIA and DARPA). The device can be worn on the wrist or as a tag on the waistband and will vibrate whenever someone wearing the device breaches social distancing and gets anywhere near natural human contact. The company had a lucky break in that it was developing a distancing sensor when the 'Covid' hoax arrived which immediately provided a potentially enormous market. How fortunate. The government in big-time Cult-controlled Ontario in Canada is investing $2.5 million in wearable contact tracing technology that 'will alert users if they may have been exposed to the Covid-19 in the workplace and will beep or vibrate if they are within six feet of another person'. Facedrive Inc., the technology company behind this, was founded in 2016 with funding from the Ontario Together Fund and obviously they, too, had a prophet on the board of directors. The human surveillance and control technology is called TraceSCAN and would be worn by the human cyborgs in places such as airports, workplaces, construction sites, care homes and … *schools*.

I emphasise schools with children and young people the prime targets. You know what is planned for society as a whole if you keep your eyes on the schools. They have always been places where the state program the next generation of slaves to be its compliant worker-ants – or Woker-ants these days; but in the mist of the 'Covid' madness they have been transformed into mind laboratories on a scale never seen before. Teachers and head teachers are just as programmed as the kids – often more so. Children are kept apart from human interaction by walk lanes, classroom distancing, staggered meal times, masks, and the rolling-out of buzzer systems. Schools are now physically laid out as a laboratory maze for lab-rats. Lunatics at a school in Anchorage, Alaska, who should be prosecuted for child abuse, took away desks and forced children to kneel (know your place) on a mat for five hours a day while wearing a mask and using their chairs as a desk. How this was supposed to impact on a 'virus' only these clinically insane people can tell you and even then it would be clap-trap. The school banned recess (interaction), art classes (creativity), and physical exercise (getting body and mind moving out of inertia). Everyone behind this outrage should be in jail or better still a mental institution. The behavioural manipulators are all for this dystopian approach to schools. Professor Susan Michie, the mind-doctor and British Communist Party member, said it was wrong to say that schools were safe. They had to be made so by 'distancing', masks and ventilation (sitting all day in the cold). I must ask this lady round for dinner on a night I know I am going to be out and not back for weeks. She probably wouldn't be able to make it, anyway, with all the visits to her own psychologist she must have block-booked.

Masking identity

I know how shocking it must be for you that a behaviour manipulator like Michie wants everyone to wear masks which have long been a feature of mind-control programs like the infamous MKUltra in the United States, but, there we are. We live and learn. I spent many years from 1996 to right across the millennium researching mind control in detail on both sides of the Atlantic and elsewhere. I met a large number of mind-control survivors and many had been held captive in body and mind by MKUltra. MK stands for mind-control, but employs the German spelling in deference to the Nazis spirited out of Germany at the end of World War Two by Operation Paperclip in which the US authorities, with help from the Vatican, transported Nazi mind-controllers and engineers to America to continue their work. Many of them were behind the creation of NASA and they included Nazi scientist and SS officer Wernher von Braun who swapped designing V-2 rockets to bombard London with designing the Saturn V rockets that powered the NASA moon programme's Apollo craft. I think I may have mentioned that the Cult has no borders. Among Paperclip escapees was Josef Mengele, the Angel of Death in the Nazi concentration camps where he conducted mind and genetic experiments on children often using twins to provide a control twin to measure the impact of his 'work' on the other. If you want to observe the Cult mentality in all its extremes of evil then look into the life of Mengele. I have met many people who suffered mercilessly under Mengele in the United States where he operated under the name Dr Greene and became a stalwart of MKUltra programming and torture. Among his locations was the underground facility in the Mojave Desert in California called the China Lake Naval Weapons Station which is almost entirely below the surface. My books *The Biggest Secret*, *Children of the Matrix* and *The Perception Deception* have the detailed background to MKUltra.

The best-known MKUltra survivor is American Cathy O'Brien. I first met her and her late partner Mark Phillips at a conference in Colorado in 1996. Mark helped her escape and deprogram from decades of captivity in an offshoot of MKUltra known as Project Monarch in which 'sex slaves' were provided for the rich and famous including Father George Bush, Dick Cheney and the Clintons. Read Cathy and Mark's book *Trance-Formation of America* and if you are new to this you will be shocked to the core. I read it in 1996 shortly before, with the usual synchronicity of my life, I found myself given a book table at the conference right next to hers. MKUltra never ended despite being very publicly exposed (only a small part of it) in the 1970s and continues in other guises. I am still in touch with Cathy. She contacted me during 2020 after masks became compulsory in many countries to tell me how they were used as part of MKUltra programming. I had been observing 'Covid regulations' and the relationship between authority and public for months. I saw techniques that I knew were employed on individuals in MKUltra being used on the global population. I had read many books and manuals on mind control including one called *Silent Weapons for Quiet Wars* which came to light in the 1980s and was a guide on how to perceptually program on a mass scale. 'Silent Weapons' refers to mind-control. I remembered a line from the manual as governments, medical authorities and law enforcement agencies have so obviously talked to – or rather at – the adult population since the 'Covid' hoax began as if they are children. The document said:

> If a person is spoken to by a T.V. advertiser as if he were a twelve-year-old, then, due to suggestibility, he will, with a certain probability, respond or react to that suggestion with the uncritical response of a twelve-year-old and will reach in to his economic reservoir and deliver its energy to buy that product on impulse when he passes it in the store.

That's why authority has spoken to adults like children since all this began.

Why did Michael Jackson wear masks?

Every aspect of the 'Covid' narrative has mind-control as its central theme. Cathy O'Brien wrote an article for davidicke.com about the connection between masks and mind control. Her daughter Kelly who I first met in the 1990s was born while Cathy was still held captive in MKUltra. Kelly was forced to wear a mask as part of her programming from the age of *two* to dehumanise her, target her sense of individuality and reduce the amount of oxygen her brain and body received. *Bingo*. This is the real reason for compulsory masks, why they have been enforced en masse, and why they seek to increase the number they demand you wear. First one, then two, with one disgraceful alleged 'doctor' recommending four which is nothing less than a death sentence. Where and how often they must be worn is being expanded for the purpose of mass mind control and damaging respiratory health which they can call 'Covid-19'. Canada's government headed by the man-child Justin Trudeau, says it's fine for children of two and older to wear masks. An insane 'study' in Italy involving just 47 children concluded there was no problem for babies as young as *four months* wearing them. Even after people were 'vaccinated' they were still told to wear masks by the criminal that is Anthony Fauci. Cathy wrote that mandating masks is allowing the authorities literally to control the air we breathe which is what was done in MKUltra. You might recall how the singer Michael Jackson wore masks and there is a reason for that. He was subjected to MKUltra mind control through Project Monarch and his psyche was scrambled by these simpletons. Cathy wrote:

> In MKUltra Project Monarch mind control, Michael Jackson had to wear a mask to silence his voice so he could not reach out for help. Remember how he developed that whisper voice when he wasn't singing? Masks control the mind from the outside in, like the redefining of words is doing. By controlling what we can and cannot say for fear of being labeled racist or beaten, for example, it ultimately controls thought that drives our words and ultimately actions (or lack thereof).
>
> Likewise, a mask muffles our speech so that we are not heard, which controls voice … words … mind. This is Mind Control. Masks are an obvious mind control device, and I am disturbed so many people are complying on a global scale. Masks depersonalize while making a person feel as though they have no voice. It is a barrier to others. People who would never choose to comply but are forced to wear a mask in order to keep their job, and ultimately their family fed, are compromised. They often feel shame and are subdued. People have stopped talking with each other while media controls the narrative.

The 'no voice' theme has often become literal with train passengers told not to speak to each other in case they pass on the 'virus', singing banned for the same reason and bonkers California officials telling people riding roller coasters that they cannot shout and scream. Cathy said she heard every day from healed MKUltra survivors who cannot wear a mask without flashing back on ways their breathing was controlled – 'from ball gags and penises to water boarding'. She said that through the years when she saw images of people in China wearing masks 'due to pollution' that it was really to control their oxygen levels. 'I knew it was as much of a population control mechanism of depersonalisation as are burkas', she said. Masks are another Chinese communist/fascist method of control that has been swept across the West as the West becomes China at lightning speed since we entered 2020.

Mask-19

There are other reasons for mandatory masks and these include destroying respiratory health to call it 'Covid-19' and stunting brain development of children and the young. Dr Margarite Griesz-Brisson MD, PhD, is a Consultant Neurologist and Neurophysiologist and the Founder and Medical Director of the London Neurology and Pain Clinic. Her CV goes down the street and round the corner. She is clearly someone who cares about people and won't parrot the propaganda. Griesz-Brisson has a PhD in pharmacology, with special interest in neurotoxicology, environmental medicine, neuroregeneration and neuroplasticity (the way the brain can change in the light of information received). She went public in October, 2020, with a passionate warning about the effects of mask-wearing laws:

> The reinhalation of our exhaled air will without a doubt create oxygen deficiency and a flooding of carbon dioxide. We know that the human brain is very sensitive to oxygen deprivation. There are nerve cells for example in the hippocampus that can't be longer than 3 minutes without oxygen – they cannot survive. The acute warning symptoms are headaches, drowsiness, dizziness, issues in concentration, slowing down of reaction time – reactions of the cognitive system.

Oh, I know, let's tell bus, truck and taxi drivers to wear them and people working machinery. How about pilots, doctors and police? Griesz-Brisson makes the important point that while the symptoms she mentions may fade as the body readjusts this does not alter the fact that people continue to operate in oxygen deficit with long list of potential consequences. She said it was well known that neurodegenerative diseases take years or decades to develop. 'If today you forget your phone number, the breakdown in your brain would have already started 20 or 30 years ago.' She said degenerative processes in your brain are getting amplified as your oxygen deprivation continues through wearing a mask. Nerve cells in the brain are unable to divide themselves normally in these circumstances and lost nerve cells will no longer be regenerated. 'What is gone is gone.' Now consider that people like shop workers and *schoolchildren* are wearing masks for hours every day. What in the name of sanity is going to be happening to them? 'I do not wear a mask, I need my brain to think', Griesz-Brisson said, 'I want to have a clear head when I deal with my patients and not be in a carbon dioxide-induced anaesthesia'. If you are told to wear a mask anywhere ask the organisation, police, store, whatever, for their risk assessment on the dangers and negative effects on mind and body of enforcing mask-wearing. They won't have one because it has never been done not even by government. All of them must be subject to class-action lawsuits as the consequences come to light. They don't do mask risk assessments for an obvious reason. They know what the conclusions would be and independent scientific studies that *have* been done tell a horror story of consequences.

'Masks are criminal'

Dr Griesz-Brisson said that for children and adolescents, masks are an absolute no-no. They had an extremely active and adaptive immune system and their brain was incredibly active with so much to learn. 'The child's brain, or the youth's brain, is thirsting for oxygen.' The more metabolically active an organ was, the more oxygen it required; and in children and adolescents every organ was metabolically active. Griesz-Brisson said that to deprive a child's or adolescent's brain of oxygen, or to restrict it in any way, was not only dangerous to their health, it was absolutely criminal. 'Oxygen deficiency inhibits the development of the brain, and the damage that has taken place as a result CANNOT be reversed.' Mind manipulators of MKUltra put masks on two-year-olds they wanted to neurologically rewire and you can see why. Griesz-Brisson said a child needs the brain to learn and the brain needs oxygen to function. 'We don't need a clinical study for that. This is simple, indisputable physiology.' Consciously and purposely induced oxygen deficiency was an absolutely deliberate health hazard, and an absolute medical contraindication which means that 'this drug, this therapy, this method or measure should not be used, and is not allowed to be used'. To coerce an entire population to use an absolute medical contraindication by force, she said, there had to be definite and serious reasons and the reasons must be presented to competent interdisciplinary and independent bodies to be verified and authorised. She had this warning of the consequences that were coming if mask wearing continued:

> When, in ten years, dementia is going to increase exponentially, and the younger generations couldn't reach their god-given potential, it won't help to say 'we didn't need the masks'. I know how damaging oxygen deprivation is for the brain, cardiologists know how damaging it is for the heart, pulmonologists know how damaging it is for the lungs. Oxygen deprivation damages every single organ. Where are our health departments, our health insurance, our medical associations? It would have been their duty to be vehemently against the lockdown and to stop it and stop it from the very beginning.
>
> Why do the medical boards issue punishments to doctors who give people exemptions? Does the person or the doctor seriously have to prove that oxygen deprivation harms people? What kind of medicine are our doctors and medical associations representing? Who is responsible for this crime? The ones who want to enforce it? The ones who let it happen and play along, or the ones who don't prevent it?

All of the organisations and people she mentions there either answer directly to the Cult or do whatever hierarchical levels above them tell them to do. The outcome of both is the same. 'It's not about masks, it's not about viruses, it's certainly not about your health', Griesz-Brisson said. 'It is about much, much more. I am not participating. I am not afraid.' They were taking our air to breathe and there was no unfounded medical exemption from face masks. Oxygen deprivation was dangerous for every single brain. It had to be the free decision of every human being whether they want to wear a mask that was absolutely ineffective to protect themselves from a virus. She ended by rightly identifying where the responsibility lies for all this:

> The imperative of the hour is personal responsibility. We are responsible for what we think, not the media. We are responsible for what we do, not our superiors. We are responsible for our health, not the World Health Organization. And we are responsible for what happens in our country, not the government.

Halle-bloody-lujah.

But surgeons wear masks, right?

Independent studies of mask-wearing have produced a long list of reports detailing mental, emotional and physical dangers. What a definition of insanity to see police officers imposing mask-wearing on the public which will cumulatively damage their health while the police themselves wear masks that will cumulatively damage *their* health. It's utter madness and both public and police do this because 'the government says so' – yes a government of brain-donor idiots like UK Health Secretary Matt Hancock reading the 'follow the science' scripts of psychopathic, lunatic psychologists. The response you get from Stockholm syndrome sufferers defending the very authorities that are destroying them and their families is that 'surgeons wear masks'. This is considered the game, set and match that they must work and don't cause oxygen deficit. Well, actually, scientific studies have shown that they *do* and oxygen levels are monitored in operating theatres to compensate. Surgeons wear masks to stop spittle and such like dropping into open wounds – not to stop 'viral particles' which are so miniscule they can only be seen through an electron microscope. Holes in the masks are significantly bigger than 'viral particles' and if you sneeze or cough they will breach the mask. I watched an incredibly disingenuous 'experiment' that claimed to prove that masks work in catching 'virus' material from the mouth and nose. They did this with a slow motion camera and the mask did block big stuff which stayed inside the mask and against the face to be breathed in or cause infections on the face as we have seen with many children. ‘Viral particles’, however, would never have been picked up by the camera as they came through the mask when they are far too small to be seen. The ‘experiment’ was therefore disingenuous *and* useless.

Studies have concluded that wearing masks in operating theatres (and thus elsewhere) make no difference to preventing infection while the opposite is true with toxic shite building up in the mask and this had led to an explosion in tooth decay and gum disease dubbed by dentists ‘mask mouth’. You might have seen the Internet video of a furious American doctor urging people to take off their masks after a four-year-old patient had been rushed to hospital the night before and nearly died with a lung infection that doctors sourced to mask wearing. A study in the journal *Cancer Discovery* found that inhalation of harmful microbes can contribute to advanced stage lung cancer in adults and long-term use of masks can help breed dangerous pathogens. Microbiologists have said frequent mask wearing creates a moist environment in which microbes can grow and proliferate before entering the lungs. The Canadian Agency for Drugs and Technologies in Health, or CADTH, a Canadian national organisation that provides research and analysis to healthcare decision-makers, said this as long ago as 2013 in a report entitled ‘Use of Surgical Masks in the Operating Room: A Review of the Clinical Effectiveness and Guidelines’. It said:

- No evidence was found to support the use of surgical face masks to reduce the frequency of surgical site infections
- No evidence was found on the effectiveness of wearing surgical face masks to protect staff from infectious material in the operating room.
- Guidelines recommend the use of surgical face masks by staff in the operating room to protect both operating room staff and patients (despite the lack of evidence).

We were told that the world could go back to 'normal' with the arrival of the 'vaccines'. When they came, fraudulent as they are, the story changed as I knew that it would. We are in the midst of transforming 'normal', not going back to it. Mary Ramsay, head of immunisation at Public Health England, echoed the words of US criminal Anthony Fauci who said masks and other regulations must stay no matter if people are vaccinated. The Fauci idiot continued to wear two masks – different colours so both could be clearly seen – after he *claimed* to have been vaccinated. Senator Rand Paul told Fauci in one exchange that his double-masks were 'theatre' and he was right. It's all theatre. Mary Ramsay back-tracked on the vaccine-return-to-normal theme when she said the public may need to wear masks and social-distance for years despite the jabs. 'People have got used to those lower-level restrictions now, and [they] can live with them', she said telling us what the idea has been all along. 'The vaccine does not give you a pass, even if you have had it, you must continue to follow all the guidelines' said a Public Health England statement which reneged on what we had been told before and made having the 'vaccine' irrelevant to 'normality' even by the official story. Spain's fascist government trumped everyone by passing a law mandating the wearing of masks on the beach and even when swimming in the sea. The move would have devastated what's left of the Spanish tourist industry, posed potential breathing dangers to swimmers and had Northern European sunbathers walking around with their forehead brown and the rest of their face white as a sheet. The ruling was so crazy that it had to be retracted after pressure from public and tourist industry, but it confirmed where the Cult wants to go with masks and how clinically insane authority has become. The determination to make masks permanent and hide the serious dangers to body and mind can be seen in the censorship of scientist Professor Denis Rancourt by Bill Gates-funded academic publishing website ResearchGate over his papers exposing the dangers and uselessness of masks. Rancourt said:

> ResearchGate today has permanently locked my account, which I have had since 2015. Their reasons graphically show the nature of their attack against democracy, and their corruption of

> science … By their obscene non-logic, a scientific review of science articles reporting on harms caused by face masks has a 'potential to cause harm'. No criticism of the psychological device (face masks) is tolerated, if the said criticism shows potential to influence public policy.

This is what happens in a fascist world.

Where are the 'greens' (again)?

Other dangers of wearing masks especially regularly relate to the inhalation of minute plastic fibres into the lungs and the deluge of discarded masks in the environment and oceans. Estimates predicted that more than 1.5 billion disposable masks will end up in the world's oceans every year polluting the water with tons of plastic and endangering marine wildlife. Studies project that humans are using 129 billion face masks each month worldwide – about three million a minute. Most are disposable and made from plastic, non-biodegradable microfibers that break down into smaller plastic particles that become widespread in ecosystems. They are littering cities, clogging sewage channels and turning up in bodies of water. I have written in other books about the immense amounts of microplastics from endless sources now being absorbed into the body. Rolf Halden, director of the Arizona State University (ASU) Biodesign Center for Environmental Health Engineering, was the senior researcher in a 2020 study that analysed 47 human tissue samples and found microplastics in all of them. 'We have detected these chemicals of plastics in every single organ that we have investigated', he said. I wrote in *The Answer* about the world being deluged with microplastics. A study by the Worldwide Fund for Nature (WWF) found that people are consuming on average every week some 2,000 tiny pieces of plastic mostly through water and also through marine life and the air. Every year humans are ingesting enough microplastics to fill a heaped dinner plate and in a life-time of 79 years it is enough to fill two large waste bins. Marco Lambertini, WWF International director general said: 'Not only are plastics polluting our oceans and waterways and killing marine life – it's in all of us and we can't escape consuming plastics,' American geologists found tiny plastic fibres, beads and shards in rainwater samples collected from the remote slopes of the Rocky Mountain National Park near Denver, Colorado. Their report was headed: ‘It is raining plastic.’ Rachel Adams, senior lecturer in Biomedical Science at Cardiff Metropolitan University, said that among health consequences are internal inflammation and immune responses to a ‘foreign body’. She further pointed out that microplastics become carriers of toxins including mercury, pesticides and dioxins (a known cause of cancer and reproductive and developmental problems). These toxins accumulate in the fatty tissues once they enter the body through microplastics. Now this is being compounded massively by people putting plastic on their face and throwing it away.

Workers exposed to polypropylene plastic fibres known as ‘flock’ have developed ‘flock worker’s lung’ from inhaling small pieces of the flock fibres which can damage lung tissue, reduce breathing capacity and exacerbate other respiratory problems. *Now* … commonly used surgical masks have three layers of melt-blown textiles made of … polypropylene. We have billions of people putting these microplastics against their mouth, nose and face for hours at a time day after day in the form of masks. How does anyone think that will work out? I mean – what could possibly go wrong? We posted a number of scientific studies on this at davidicke.com, but when I went back to them as I was writing this book the links to the science research website where they were hosted were dead. Anything that challenges the official narrative in any way is either censored or vilified. The official narrative is so unsupportable by the evidence that only deleting the truth can protect it. A study by Chinese scientists still survived – with the usual twist which it why it was still active, I guess. Yes, they found that virtually all the masks they tested increased the daily intake of microplastic fibres, but people should still wear them because the danger from the ‘virus’ was worse said the crazy ‘team’ from the Institute of Hydrobiology in Wuhan. Scientists first discovered microplastics in lung tissue of some patients who died of lung cancer in the 1990s. Subsequent studies have confirmed the potential health damage with the plastic degrading slowly and remaining in the lungs to accumulate in volume. Wuhan researchers used a machine simulating human breathing to establish that masks shed up to nearly 4,000 microplastic fibres in a month with reused masks producing more. Scientists said some masks are laced with toxic chemicals and a variety of compounds seriously restricted for both health and environmental reasons. They include cobalt (used in blue dye) and formaldehyde known to cause watery eyes, burning sensations in the eyes, nose, and throat, plus coughing, wheezing and nausea. No – that must be 'Covid-19'.

Mask 'worms'

There is another and potentially even more sinister content of masks. Mostly new masks of different makes filmed under a microscope around the world have been found to contain strange black fibres or 'worms' that appear to move or 'crawl' by themselves and react to heat and water. The nearest I have seen to them are the self-replicating fibres that are pulled out through the skin of those suffering from Morgellons disease which has been connected to the phenomena of 'chemtrails' which I will bring into the story later on. Morgellons fibres continue to grow outside the body and have a form of artificial intelligence. Black 'worm' fibres in masks have that kind of feel to them and there is a nanotechnology technique called 'worm micelles' which carry and release drugs or anything else you want to deliver to the body. For sure the suppression of humanity by mind altering drugs is the Cult agenda big time and the more excuses they can find to gain access to the body the more opportunities there are to make that happen whether through 'vaccines' or masks pushed against the mouth and nose for hours on end.

So let us summarise the pros and cons of masks:

Against masks: Breathing in your own carbon dioxide; depriving the body and brain of sufficient oxygen; build-up of toxins in the mask that can be breathed into the lungs and cause rashes on the face and 'mask-mouth'; breathing microplastic fibres and toxic chemicals into the lungs; dehumanisation and deleting individualisation by literally making people faceless; destroying human emotional interaction through facial expression and deleting parental connection with their babies which look for guidance to their facial expression.

For masks: They don't protect you from a 'virus' that doesn't exist and even if it did 'viral' particles are so minute they are smaller than the holes in the mask.

Governments, police, supermarkets, businesses, transport companies, and all the rest who seek to impose masks have done no risk assessment on their consequences for health and psychology and are now open to group lawsuits when the impact becomes clear with a cumulative epidemic of respiratory and other disease. Authorities will try to exploit these effects and hide the real cause by dubbing them 'Covid-19'. Can you imagine setting out to force the population to wear health-destroying masks without doing any assessment of the risks? It is criminal and it is evil, but then how many people targeted in this way, who see their children told to wear them all day at school, have asked for a risk assessment? Billions can't be imposed upon by the few unless the billions allow it. Oh, yes, with just a tinge of irony, 85 percent of all masks made worldwide come from *China*.

Wash your hands in toxic shite

'Covid' rules include the use of toxic sanitisers and again the health consequences of constantly applying toxins to be absorbed through the skin is obvious to any level of Renegade Mind. America's Food and Drug Administration (FDA) said that sanitisers are drugs and issued a warning about 75 dangerous brands which contain methanol used in antifreeze and can cause death, kidney damage and blindness. The FDA circulated the following warning even for those brands that it claims to be safe:

> Store hand sanitizer out of the reach of pets and children, and children should use it only with adult supervision. Do not drink hand sanitizer. This is particularly important for young children, especially toddlers, who may be attracted by the pleasant smell or brightly colored bottles of hand sanitizer.
>
> Drinking even a small amount of hand sanitizer can cause alcohol poisoning in children. (However, there is no need to be concerned if your children eat with or lick their hands after using hand sanitizer.) During this coronavirus pandemic, poison control centers have had an increase in calls about accidental ingestion of hand sanitizer, so it is important that adults monitor young children's use.
>
> Do not allow pets to swallow hand sanitizer. If you think your pet has eaten something potentially dangerous, call your veterinarian or a pet poison control center right away. Hand sanitizer is flammable and should be stored away from heat and flames. When using hand sanitizer, rub your hands until they feel completely dry before performing activities that may involve heat, sparks, static electricity, or open flames.

There you go, perfectly safe, then, and that's without even a mention of the toxins absorbed through the skin. Come on kids – sanitise your hands everywhere you go. It will save you from the 'virus'. Put all these elements together of the 'Covid' normal and see how much health and psychology is being cumulatively damaged, even devastated, to 'protect your health'. Makes sense, right? They are only imposing these things because they care, right? *Right?*

Submitting to insanity

Psychological reframing of the population goes very deep and is done in many less obvious ways. I hear people say how contradictory and crazy 'Covid' rules are and how they are ever changing. This is explained away by dismissing those involved as idiots. It is a big mistake. The Cult is delighted if its cold calculation is perceived as incompetence and idiocy when it is anything but. Oh, yes, there are idiots within the system – lots of them – but they are *administering* the Cult agenda, mostly unknowingly. They are not deciding and dictating it. The bulwark against tyranny is self-respect, always has been, always will be. It is self-respect that has broken every tyranny in history. By its very nature self-respect will not bow to oppression and its perpetrators. There is so little self-respect that it's always the few that overturn dictators. Many may eventually follow, but the few with the iron spines (self-respect) kick it off and generate the momentum. The Cult targets self-respect in the knowledge that once this has gone only submission remains. Crazy, contradictory, ever-changing 'Covid' rules are systematically applied by psychologists to delete self-respect. They *want* you to see that the rules make no sense. It is one thing to decide to do something when *you* have made the choice based on evidence and logic. You still retain your self-respect. It is quite another when you can see what you are being told to do is insane, ridiculous and makes no sense, and *yet you still do it*. Your self-respect is extinguished and this has been happening as ever more obviously stupid and nonsensical things have been demanded and the great majority have complied even when they can see they are stupid and nonsensical.

People walk around in face-nappies knowing they are damaging their health and make no difference to a 'virus'. They do it in fear of not doing it. I know it's daft, but I'll do it anyway. When that happens something dies inside of you and submissive reframing has begun. Next there's a need to hide from yourself that you have conceded your self-respect and you convince yourself that you have not really submitted to fear and intimidation. You begin to believe that you are complying with craziness because it's the right thing to do. When first you concede your self-respect of 2+2 = 4 to 2+2 = 5 you *know* you are compromising your self-respect. Gradually to avoid facing that fact you begin to *believe* that 2+2=5. You have been reframed and I have been watching this process happening in the human psyche on an industrial scale. The Cult is working to break your spirit and one of its major tools in that war is humiliation. I read how former American soldier Bradley Manning (later Chelsea Manning after a sex-change) was treated after being jailed for supplying WikiLeaks with documents exposing the enormity of government and elite mendacity. Manning was isolated in solitary confinement for eight months, put under 24-hour surveillance, forced to hand over clothing before going to bed, and stand naked for every roll call. This is systematic humiliation. The introduction of anal swab ‘Covid’ tests in China has been done for the same reason to delete self-respect and induce compliant submission. Anal swabs are mandatory for incoming passengers in parts of China and American diplomats have said they were forced to undergo the indignity which would have been calculated humiliation by the Cult-owned Chinese government that has America in its sights.

Government-people: An abusive relationship

Spirit-breaking psychological techniques include giving people hope and apparent respite from tyranny only to take it away again. This happened in the UK during Christmas, 2020, when the psycho-psychologists and their political lackeys announced an easing of restrictions over the holiday only to reimpose them almost immediately on the basis of yet another lie. There is a big psychological difference between getting used to oppression and being given hope of relief only to have that dashed. Psychologists know this and we have seen the technique used repeatedly. Then there is traumatising people before you introduce more extreme regulations that require compliance. A perfect case was the announcement by the dark and sinister Whitty and Vallance in the UK that ‘new data’ predicted that 4,000 could die every day over the winter of 2020/2021 if we did not lockdown again. I think they call it lying and after traumatising people with that claim out came Jackboot Johnson the next day with new curbs on human freedom. Psychologists know that a frightened and traumatised mind becomes suggestable to submission and behaviour reframing. Underpinning all this has been to make people fearful and suspicious of each other and see themselves as a potential danger to others. In league with deleted self-respect you have the perfect psychological recipe for self-loathing. The relationship between authority and public is now demonstrably the same as that of subservience to an abusive partner. These are signs of an abusive relationship explained by psychologist Leslie Becker-Phelps:

Psychological and emotional abuse: Undermining a partner's self-worth with verbal attacks, name-calling, and belittling. Humiliating the partner in public, unjustly accusing them of having an affair, or interrogating them about their every behavior. Keeping partner confused or off balance by saying they were just kidding or blaming the partner for 'making' them act this way … Feigning in public that they care while turning against them in private. This leads to victims frequently feeling confused, incompetent, unworthy, hopeless, and chronically self-doubting. [Apply these techniques to how governments have treated the population since New Year, 2020, and the parallels are obvious.]

Physical abuse: The abuser might physically harm their partner in a range of ways, such as grabbing, hitting, punching, or shoving them. They might throw objects at them or harm them with a weapon. [Observe the physical harm imposed by masks, lockdown, and so on.]

Threats and intimidation: One way abusers keep their partners in line is by instilling fear. They might be verbally threatening, or give threatening looks or gestures. Abusers often make it known that they are tracking their partner's every move. They might destroy their partner's possessions, threaten to harm them, or threaten to harm their family members. Not surprisingly, victims of this abuse often feel anxiety, fear, and panic. [No words necessary.]

Isolation: Abusers often limit their partner's activities, forbidding them to talk or interact with friends or family. They might limit access to a car or even turn off their phone. All of this might be done by physically holding them against their will, but is often accomplished through psychological abuse and intimidation. The more isolated a person feels, the fewer resources they have to help gain perspective on their situation and to escape from it. [No words necessary.]

Economic abuse: Abusers often make their partners beholden to them for money by controlling access to funds of any kind. They might prevent their partner from getting a job or withhold access to money they earn from a job. This creates financial dependency that makes leaving the relationship very difficult. [See destruction of livelihoods and the proposed meagre 'guaranteed income' so long as you do whatever you are told.]

Using children: An abuser might disparage their partner's parenting skills, tell their children lies about their partner, threaten to take custody of their children, or threaten to harm their children. These tactics instil fear and often elicit compliance. [See reframed social service mafia and how children are being mercilessly abused by the state over 'Covid' while their parents look on too frightened to do anything.]

A further recurring trait in an abusive relationship is the abused blaming themselves for their abuse and making excuses for the abuser. We have the public blaming each other for lockdown abuse by government and many making excuses for the government while attacking those who challenge the government. How often we have heard authorities say that rules are being imposed or reimposed only because people have refused to 'behave' and follow the rules. We don't want to do it – it's *you*.

Renegade Minds are an antidote to all of these things. They will never concede their self-respect no matter what the circumstances. Even when apparent humiliation is heaped upon them they laugh in its face and reflect back the humiliation on the abuser where it belongs. Renegade Minds will never wear masks they know are only imposed to humiliate, suppress and damage both physically and psychologically. Consequences will take care of themselves and they will never break their spirit or cause them to concede to tyranny. UK newspaper columnist Peter Hitchens was one of the few in the mainstream media to speak out against lockdowns and forced vaccinations. He then announced he had taken the jab. He wanted to see family members abroad and he believed vaccine passports were inevitable even though they had not yet been introduced. Hitchens has a questioning and critical mind, but not a Renegade one. If he had no amount of pressure would have made him concede. Hitchens excused his action by saying that the battle has been lost. Renegade Minds never accept defeat when freedom is at stake and even if they are the last one standing the self-respect of not submitting to tyranny is more important than any outcome or any consequence.

That's why Renegade Minds are the only minds that ever changed anything worth changing.

CHAPTER EIGHT

‘Reframing’ insanity

Insanity is relative. It depends on who has who locked in what cage
Ray Bradbury

‘Reframing’ a mind means simply to change its perception and behaviour. This can be done subconsciously to such an extent that subjects have no idea they have been ‘reframed’ while to any observer changes in behaviour and attitudes are obvious.

Human society is being reframed on a ginormous scale since the start of 2020 and here we have the reason why psychologists rather than doctors have been calling the shots. Ask most people who have succumbed to ‘Covid’ reframing if they have changed and most will say ‘no’; but they *have* and fundamentally. The Cult’s long-game has been preparing for these times since way back and crucial to that has been to prepare both population and officialdom mentally and emotionally. To use the mind-control parlance they had to reframe the population with a mentality that would submit to fascism and reframe those in government and law enforcement to impose fascism or at least go along with it. The result has been the fact-deleted mindlessness of ‘Wokeness’ and officialdom that has either enthusiastically or unquestioningly imposed global tyranny demanded by reframed politicians on behalf of psychopathic and deeply evil cultists. ‘Cognitive reframing’ identifies and challenges the way someone sees the world in the form of situations, experiences and emotions and then restructures those perceptions to view the same set of circumstances in a different way. This can have benefits if the attitudes are personally destructive while on the other side it has the potential for individual and collective mind control which the subject has no idea has even happened.

Cognitive therapy was developed in the 1960s by Aaron T. Beck who was born in Rhode Island in 1921 as the son of Jewish immigrants from the Ukraine. He became interested in the techniques as a treatment for depression. Beck's daughter Judith S. Beck is prominent in the same field and they founded the Beck Institute for Cognitive Behavior Therapy in Philadelphia in 1994. Cognitive reframing, however, began to be used worldwide by those with a very dark agenda. The Cult reframes politicians to change their attitudes and actions until they are completely at odds with what they once appeared to stand for. The same has been happening to government administrators at all levels, law enforcement, military and the human population. Cultists love mind control for two main reasons: It allows them to control what people think, do and say to secure agenda advancement and, by definition, it calms their legendary insecurity and fear of the unexpected. I have studied mind control since the time I travelled America in 1996. I may have been talking to next to no one in terms of an audience in those years, but my goodness did I gather a phenomenal amount of information and knowledge about so many things including the techniques of mind control. I have described this in detail in other books going back to *The Biggest Secret* in 1998. I met a very large number of people recovering from MKUltra and its offshoots and successors and I began to see how these same techniques were being used on the population in general. This was never more obvious than since the 'Covid' hoax began.

Reframing the enforcers

I have observed over the last two decades and more the very clear transformation in the dynamic between the police, officialdom and the public. I tracked this in the books as the relationship mutated from one of serving the public to seeing them as almost the enemy and certainly a lower caste. There has always been a class divide based on income and always been some psychopathic, corrupt, and big-I-am police officers. This was different. Wholesale change was unfolding in the collective dynamic; it was less about money and far more about position and perceived power. An us-and-them was emerging. Noses were lifted skyward by government administration and law enforcement and their attitude to the public they were *supposed* to be serving changed to one of increasing contempt, superiority and control. The transformation was so clear and widespread that it had to be planned. Collective attitudes and dynamics do not change naturally and organically that quickly on that scale. I then came across an organisation in Britain called Common Purpose created in the late 1980s by Julia Middleton who would work in the office of Deputy Prime Minister John Prescott during the long and disastrous premiership of war criminal Tony Blair. When Blair speaks the Cult is speaking and the man should have been in jail a long time ago. Common Purpose proclaims itself to be one of the biggest 'leadership development' organisations in the world while functioning as a *charity* with all the financial benefits which come from that. It hosts 'leadership development' courses and programmes all over the world and claims to have 'brought together' what it calls 'leaders' from more than 100 countries on six continents. The modus operandi of Common Purpose can be compared with the work of the UK government's reframing network that includes the Behavioural Insights Team 'nudge unit' and 'Covid' reframing specialists at SPI-B. WikiLeaks described Common Purpose long ago as 'a hidden virus in our government and schools' which is unknown to the general public: 'It recruits and trains "leaders" to be loyal to the directives of Common Purpose and the EU, instead of to their own departments, which they then undermine or subvert, the NHS [National Health Service] being an example.' This is a vital point to understand the 'Covid' hoax. The NHS, and its equivalent around the world, has been utterly reframed in terms of administrators and much of the medical personnel with the transformation underpinned by recruitment policies. The outcome has been the criminal and psychopathic behaviour of the

NHS over 'Covid' and we have seen the same in every other major country. WikiLeaks said Common Purpose trainees are 'learning to rule without regard to democracy' and to usher in a police state (current events explained). Common Purpose operated like a 'glue' and had members in the NHS, BBC, police, legal profession, church, many of Britain's 7,000 quangos, local councils, the Civil Service, government ministries and Parliament, and controlled many RDA's (Regional Development Agencies). Here we have one answer for how and why British institutions and their like in other countries have changed so negatively in relation to the public. This further explains how and why the beyond-disgraceful reframed BBC has become a propaganda arm of 'Covid' fascism. They are all part of a network pursuing the same goal.

By 2019 Common Purpose was quoting a figure of 85,000 'leaders' that had attended its programmes. These 'students' of all ages are known as Common Purpose 'graduates' and they consist of government, state and local government officials and administrators, police chiefs and officers, and a whole range of others operating within the national, local and global establishment. Cressida Dick, Commissioner of the London Metropolitan Police, is the Common Purpose graduate who was the 'Gold Commander' that oversaw what can only be described as the murder of Brazilian electrician Jean Charles de Menezes in 2005. He was held down by psychopathic police and shot seven times in the head by a psychopathic lunatic after being mistaken for a terrorist when he was just a bloke going about his day. Dick authorised officers to pursue and keep surveillance on de Menezes and ordered that he be stopped from entering the underground train system. Police psychopaths took her at her word clearly. She was 'disciplined' for this outrage by being *promoted* – eventually to the top of the 'Met' police where she has been a disaster. Many Chief Constables controlling the police in different parts of the UK are and have been Common Purpose graduates. I have heard the 'graduate' network described as a sort of Mafia or secret society operating within the fabric of government at all levels pursuing a collective policy ingrained at Common Purpose training events. Founder Julia Middleton herself has said:

> Locally and internationally, Common Purpose graduates will be 'lighting small fires' to create change in their organisations and communities … The Common Purpose effect is best illustrated by the many stories of small changes brought about by leaders, who themselves have changed.

A Common Purpose mission statement declared:

> Common Purpose aims to improve the way society works by expanding the vision, decision-making ability and influence of all kinds of leaders. The organisation runs a variety of educational programmes for leaders of all ages, backgrounds and sectors, in order to provide them with the inspirational, information and opportunities they need to change the world.

Yes, but into what? Since 2020 the answer has become clear.

NLP and the Delphi technique

Common Purpose would seem to be a perfect name or would common programming be better? One of the foundation methods of reaching 'consensus' (group think) is by setting the agenda theme and then encouraging, cajoling or pressuring everyone to agree a 'consensus' in line with the core theme promoted by Common Purpose. The methodology involves the 'Delphi technique', or an adaption of it, in which opinions are expressed that are summarised by a 'facilitator or change agent' at each stage. Participants are 'encouraged' to modify their views in the light of what others have said. Stage by stage the former individual opinions are merged into group consensus which just happens to be what Common Purpose wants them to believe. A key part of this is to marginalise anyone refusing to concede to group think and turn the group against them to apply pressure to conform. We are seeing this very technique used on the general population to make 'Covid' group-thinkers hostile to those who have seen through the bullshit. People can be reframed by using perception manipulation methods such as Neuro-Linguistic Programming (NLP) in which you change perception with the use of carefully constructed language. An NLP website described the technique this way:

> … A method of influencing brain behaviour (the 'neuro' part of the phrase) through the use of language (the 'linguistic' part) and other types of communication to enable a person to 'recode' the way the brain responds to stimuli (that's the 'programming') and manifest new and better behaviours. Neuro-Linguistic Programming often incorporates hypnosis and self-hypnosis to help achieve the change (or 'programming') that is wanted.

British alternative media operation UKColumn has done very detailed research into Common Purpose over a long period. I quoted co-founder and former naval officer Brian Gerrish in my book *Remember Who You Are*, published in 2011, as saying the following years before current times:

> It is interesting that many of the mothers who have had children taken by the State speak of the Social Services people being icily cool, emotionless and, as two ladies said in slightly different words, '… like little robots'. We know that NLP is cumulative, so people can be given small imperceptible doses of NLP in a course here, another in a few months, next year etc. In this way, major changes are accrued in their personality, but the day by day change is almost unnoticeable.

In these and other ways 'graduates' have had their perceptions uniformly reframed and they return to their roles in the institutions of government, law enforcement, legal profession, military, 'education', the UK National Health Service and the whole swathe of the establishment structure to pursue a common agenda preparing for the 'post-industrial', 'post-democratic' society. I say 'preparing' but we are now there. 'Post-industrial' is code for the Great Reset and 'post-democratic' is 'Covid' fascism. UKColumn has spoken to partners of those who have attended Common Purpose 'training'. They have described how personalities and attitudes of 'graduates' changed very noticeably for the worse by the time they had completed the course. They had been 'reframed' and told they are the 'leaders' – the special ones – who know better than the population. There has also been the very demonstrable recruitment of psychopaths and narcissists into government administration at all levels and law enforcement. If you want psychopathy hire psychopaths and you get a simple cause and effect. If you want administrators, police officers and 'leaders' to perceive the public as lesser beings who don't matter then employ narcissists. These personalities are identified using 'psychometrics' that identifies knowledge, abilities, attitudes and personality traits, mostly through carefully-designed questionnaires and tests. As this policy has passed through the decades we have had power-crazy, power-trippers appointed into law enforcement, security and government administration in preparation for current times and the dynamic between public and law enforcement/officialdom has been transformed. UKColumn's Brian Gerrish said of the narcissistic personality:

> Their love of themselves and power automatically means that they will crush others who get in their way. I received a major piece of the puzzle when a friend pointed out that when they made public officials re-apply for their own jobs several years ago they were also required to do psychometric tests. This was undoubtedly the start of the screening process to get 'their' sort of people in post.

How obvious that has been since 2020 although it was clear what was happening long before if people paid attention to the changing public-establishment dynamic.

Change agents

At the centre of events in 'Covid' Britain is the National Health Service (NHS) which has behaved disgracefully in slavishly following the Cult agenda. The NHS management structure is awash with Common Purpose graduates or 'change agents' working to a common cause. Helen Bevan, a Chief of Service Transformation at the NHS Institute for Innovation and Improvement, co-authored a document called 'Towards a million change agents, a review of the social movements literature: implications for large scale change in the NHS'. The document compared a project management approach to that of change and social movements where 'people change themselves and each other – peer to peer'. Two definitions given for a 'social movement' were:

A group of people who consciously attempt to build a radically new social order; involves people of a broad range of social backgrounds; and deploys politically confrontational and socially disruptive tactics – Cyrus Zirakzadeh 1997

Collective challenges, based on common purposes and social solidarities, in sustained interaction with elites, opponents, and authorities – Sidney Tarrow 1994

Helen Bevan wrote another NHS document in which she defined 'framing' as 'the process by which leaders construct, articulate and put across their message in a powerful and compelling way in order to win people to their cause and call them to action'. I think I could come up with another definition that would be rather more accurate. The National Health Service and institutions of Britain and the wider world have been taken over by reframed 'change agents' and that includes everything from the United Nations to national governments, local councils and social services which have been kidnapping children from loving parents on an extraordinary and gathering scale on the road to the end of parenthood altogether. Children from loving homes are stolen and kidnapped by the state and put into the 'care' (inversion) of the local authority through council homes, foster parents and forced adoption. At the same time children are allowed to be abused without response while many are under council 'care'. UKColumn highlighted the Common Purpose connection between South Yorkshire Police and Rotherham council officers in the case of the scandal in that area of the sexual exploitation of children to which the authorities turned not one blind eye, but both:

> We were alarmed to discover that the Chief Executive, the Strategic Director of Children and Young People's Services, the Manager for the Local Strategic Partnership, the Community Cohesion Manager, the Cabinet Member for Cohesion, the Chief Constable and his predecessor had all attended Leadership training courses provided by the pseudo-charity Common Purpose.

Once 'change agents' have secured positions of hire and fire within any organisation things start to move very quickly. Personnel are then hired and fired on the basis of whether they will work towards the agenda the change agent represents. If they do they are rapidly promoted even though they may be incompetent. Those more qualified and skilled who are pre-Common Purpose 'old school' see their careers stall and even disappear. This has been happening for decades in every institution of state, police, 'health' and social services and all of them have been transformed as a result in their attitudes to their jobs and the public. Medical professions, including nursing, which were once vocations for the caring now employ many cold, callous and couldn't give a shit personality types. The UKColumn investigation concluded:

> By blurring the boundaries between people, professions, public and private sectors, responsibility and accountability, Common Purpose encourages 'graduates' to believe that as new selected leaders, they can work together, outside of the established political and social structures, to achieve a paradigm shift or CHANGE – so called 'Leading Beyond Authority'. In doing so, the allegiance of the individual becomes 'reframed' on CP colleagues and their NETWORK.

Reframing the Face-Nappies

Nowhere has this process been more obvious than in the police where recruitment of psychopaths and development of unquestioning mind-controlled group-thinkers have transformed law enforcement into a politically-correct 'Woke' joke and a travesty of what should be public service. Today they wear their face-nappies like good little gofers and enforce 'Covid' rules which are fascism under another name. Alongside the specifically-recruited psychopaths we have software minds incapable of free thought. Brian Gerrish again:

> An example is the policeman who would not get on a bike for a press photo because he had not done the cycling proficiency course. Normal people say this is political correctness gone mad. Nothing could be further from the truth. The policeman has been reframed, and in his reality it is perfect common sense not to get on the bike 'because he hasn't done the cycling course'.
>
> Another example of this is where the police would not rescue a boy from a pond until they had taken advice from above on the 'risk assessment'. A normal person would have arrived, perhaps thought of the risk for a moment, and dived in. To the police now 'reframed', they followed 'normal' procedure.

There are shocking cases of reframed ambulance crews doing the same. Sheer unthinking stupidity of London Face-Nappies headed by Common Purpose graduate Cressida Dick can be seen in their behaviour at a vigil in March, 2021, for a murdered woman, Sarah Everard. A police officer had been charged with the crime. Anyone with a brain would have left the vigil alone in the circumstances. Instead they 'manhandled' women to stop them breaking 'Covid rules' to betray classic reframing. Minds in the thrall of perception control have no capacity for seeing a situation on its merits and acting accordingly. 'Rules is rules' is their only mind-set. My father used to say that rules and regulations are for the guidance of the intelligent and the blind obedience of the idiot. Most of the intelligent, decent, coppers have gone leaving only the other kind and a few old school for whom the job must be a daily nightmare. The combination of psychopaths and rule-book software minds has been clearly on public display in the 'Covid' era with automaton robots in uniform imposing fascistic 'Covid' regulations on the population without any personal initiative or judging situations on their merits. There are thousands of examples around the world, but I'll make my point with the infamous Derbyshire police in the English East Midlands – the ones who think pouring dye into beauty spots and using drones to track people walking in the countryside away from anyone is called 'policing'. To them there are rules decreed by the government which they have to enforce and in their bewildered state a group gathering in a closed space and someone walking alone in the countryside are the same thing. It is beyond idiocy and enters the realm of clinical insanity.

Police officers in Derbyshire said they were 'horrified' – *horrified* – to find 15 to 20 'irresponsible' kids playing a football match at a closed leisure centre 'in breach of coronavirus restrictions'. When they saw the police the kids ran away leaving their belongings behind and the reframed men and women of Derbyshire police were seeking to establish their identities with a view to fining their parents. The most natural thing for youngsters to do – kicking a ball about – is turned into a criminal activity and enforced by the moronic software programs of Derbyshire police. You find the same mentality in every country. These barely conscious 'horrified' officers said they had to take action because 'we need to ensure these rules are being followed' and 'it is of the utmost importance that you ensure your children are following the rules and regulations for Covid-19'. Had any of them done ten seconds of research to see if this parroting of their masters' script could be supported by any evidence? Nope. Reframed people don't think – others think for them and that's the whole idea of reframing. I have seen police officers one after the other repeating without question word for word what officialdom tells them just as I have seen great swathes of the public doing the same. Ask either for 'their' opinion and out spews what they have been told to think by the official narrative. Police and public may seem to be in different groups, but their mentality is the same. Most people do whatever they are told in fear not doing so or because they believe what officialdom tells them; almost the entirety of the police do what they are told for the same reason. Ultimately it's the tiny inner core of the global Cult that's telling both what to do.

So Derbyshire police were 'horrified'. Oh, really? Why did they think those kids were playing football? It was to relieve the psychological consequences of lockdown and being denied human contact with their friends and interaction, touch and discourse vital to human psychological health. Being denied this month after month has dismantled the psyche of many children and young people as depression and suicide have exploded. Were Derbyshire police *horrified by that*? Are you kidding? Reframed people don't have those mental and emotional processes that can see how the impact on the psychological health of youngsters is far more dangerous than any ‘virus’ even if you take the mendacious official figures to be true. The reframed are told (programmed) how to act and so they do. The Derbyshire Chief Constable in the first period of lockdown when the black dye and drones nonsense was going on was Peter Goodman. He was the man who severed the connection between his force and the Derbyshire Constabulary *Male Voice* Choir when he decided that it was not inclusive enough to allow women to join. The fact it was a male voice choir making a particular sound produced by male voices seemed to elude a guy who terrifyingly ran policing in Derbyshire. He retired weeks after his force was condemned as disgraceful by former Supreme Court Justice Jonathan Sumption for their behaviour over extreme lockdown impositions. Goodman was replaced by his deputy Rachel Swann who was in charge when her officers were ‘horrified’. The police statement over the boys committing the hanging-offence of playing football included the line about the youngsters being ‘irresponsible in the times we are all living through’ missing the point that the real relevance of the ‘times we are all living through’ is the imposition of fascism enforced by psychopaths and reframed minds of police officers playing such a vital part in establishing the fascist tyranny that their own children and grandchildren will have to live in their entire lives. As a definition of insanity that is hard to beat although it might be run close by imposing masks on people that can have a serious effect on their health while wearing a face nappy all day themselves. Once again public and police do it for the same reason – the authorities tell them to and who are they to have the self-respect to say no?

Wokers in uniform

How reframed do you have to be to arrest a *six-year-old* and take him to court for *picking a flower* while waiting for a bus? Brain dead police and officialdom did just that in North Carolina where criminal proceedings happen regularly for children under nine. Attorney Julie Boyer gave the six-year-old crayons and a colouring book during the ‘flower’ hearing while the ‘adults’ decided his fate. County Chief District Court Judge Jay Corpening asked: ‘Should a child that believes in Santa Claus, the Easter Bunny and the tooth fairy be making life-altering decisions?’ Well, of course not, but common sense has no meaning when you have a common purpose and a reframed mind. Treating children in this way, and police operating in American schools, is all part of the psychological preparation for children to accept a police state as normal all their adult lives. The same goes for all the cameras and biometric tracking technology in schools. Police training is focused on reframing them as snowflake Wokers and this is happening in the military. Pentagon top brass said that ‘training sessions on extremism’ were needed for troops who asked why they were so focused on the Capitol Building riot when Black Lives Matter riots were ignored. What’s the difference between them some apparently and rightly asked. Actually, there is a difference. Five people died in the Capitol riot, only one through violence, and that was a police officer shooting an unarmed protestor. BLM riots killed at least 25 people and cost billions. Asking the question prompted the psychopaths and reframed minds that run the Pentagon to say that more ‘education’ (programming) was needed. Troop training is all based on psychological programming to make them fodder for the Cult – ‘Military men are just dumb, stupid animals to be used as pawns in foreign policy’ as Cult-to-his-DNA former Secretary of State Henry Kissinger famously said. Governments see the police in similar terms and it’s time for those among them who can see this to defend the people and stop being enforcers of the Cult agenda upon the people.

The US military, like the country itself, is being targeted for destruction through a long list of Woke impositions. Cult-owned gaga ‘President’ Biden signed an executive order when he took office to allow taxpayer money to pay for transgender surgery for active military personnel and veterans. Are you a man soldier? No, I’m a LGBTQIA+ with a hint of Skoliosexual and Spectrasexual. Oh, good man. Bad choice of words you bigot. The Pentagon announced in March, 2021, the appointment of the first ‘diversity and inclusion officer' for US Special Forces. Richard Torres-Estrada arrived with the publication of a 'D&I Strategic Plan which will guide the enterprise-wide effort to institutionalize and sustain D&I'. If you think a Special Forces 'Strategic Plan' should have something to do with defending America you haven't been paying attention. Defending Woke is now the military's new role. Torres-Estrada has posted images comparing Donald Trump with Adolf Hitler and we can expect no bias from him as a representative of the supposedly non-political Pentagon. Cable news host Tucker Carlson said: 'The Pentagon is now the Yale faculty lounge but with cruise missiles.' Meanwhile Secretary of Defense Lloyd Austin, a board member of weapons-maker Raytheon with stock and compensation interests in October, 2020, worth $1.4 million, said he was purging the military of the 'enemy within' – anyone who isn't Woke and supports Donald Trump. Austin refers to his targets as 'racist extremists' while in true Woke fashion being himself a racist extremist. Pentagon documents pledge to 'eradicate, eliminate and conquer all forms of racism, sexism and homophobia'. The definitions of these are decided by 'diversity and inclusion committees' peopled by those who see racism, sexism and homophobia in every situation and opinion. Woke (the Cult) is dismantling the US military and purging testosterone as China expands its military and gives its troops 'masculinity training'. How do we think that is going to end when this is all Cult coordinated? The US military, like the British military, is controlled by Woke and spineless top brass who just go along with it out of personal career interests.

'Woke' means fast asleep

Mind control and perception manipulation techniques used on individuals to create group-think have been unleashed on the global population in general. As a result many have no capacity to see the obvious fascist agenda being installed all around them or what 'Covid' is really all about. Their brains are firewalled like a computer system not to process certain concepts, thoughts and realisations that are bad for the Cult. The young are most targeted as the adults they will be when the whole fascist global state is planned to be fully implemented. They need to be prepared for total compliance to eliminate all pushback from entire generations. The Cult has been pouring billions into taking complete control of 'education' from schools to universities via its operatives and corporations and not least Bill Gates as always. The plan has been to transform 'education' institutions into programming centres for the mentality of 'Woke'. James McConnell, professor of psychology at the University of Michigan, wrote in *Psychology Today* in 1970:

> The day has come when we can combine sensory deprivation with drugs, hypnosis, and astute manipulation of reward and punishment, to gain almost absolute control over an individual's behaviour. It should then be possible to achieve a very rapid and highly effective type of brainwashing that would allow us to make dramatic changes in a person's behaviour and personality ...
>
> … We should reshape society so that we all would be trained from birth to want to do what society wants us to do. We have the techniques to do it... no-one owns his own personality you acquired, and there's no reason to believe you should have the right to refuse to acquire a new personality if your old one is anti-social.

This was the potential for mass brainwashing in 1970 and the mentality there displayed captures the arrogant psychopathy that drives it forward. I emphasise that not all young people have succumbed to Woke programming and those that haven't are incredibly impressive people given that today's young are the most perceptually-targeted generations in history with all the technology now involved. Vast swathes of the young generations, however, have fallen into the spell – and that's what it is – of Woke. The Woke mentality and perceptual program is founded on *inversion* and you will appreciate later why that is so significant. Everything with Woke is inverted and the opposite of what it is claimed to be. Woke was a term used in African-American culture from the 1900s and referred to an awareness of social and racial justice. This is not the meaning of the modern version or 'New Woke' as I call it in *The Answer*. Oh, no, Woke today means something very different no matter how much Wokers may seek to hide that and insist Old Woke and New

Woke are the same. See if you find any ‘awareness of social justice’ here in the modern variety:

- Woke demands ‘inclusivity’ while excluding anyone with a different opinion and calls for mass censorship to silence other views.
- Woke claims to stand against oppression when imposing oppression is the foundation of all that it does. It is the driver of political correctness which is nothing more than a Cult invention to manipulate the population to silence itself.
- Woke believes itself to be ‘liberal’ while pursuing a global society that can only be described as fascist (see ‘anti-fascist’ fascist Antifa).
- Woke calls for ‘social justice’ while spreading injustice wherever it goes against the common ‘enemy’ which can be easily identified as a differing view.
- Woke is supposed to be a metaphor for ‘awake’ when it is solid-gold asleep and deep in a Cult-induced coma that meets the criteria for ‘off with the fairies’.

I state these points as obvious facts if people only care to look. I don’t do this with a sense of condemnation. We need to appreciate that the onslaught of perceptual programming on the young has been incessant and merciless. I can understand why so many have been reframed, or, given their youth, framed from the start to see the world as the Cult demands. The Cult has had access to their minds day after day in its ‘education’ system for their entire formative years. Perception is formed from information received and the Cult-created system is a life-long download of information delivered to elicit a particular perception, thus behaviour. The more this has expanded into still new extremes in recent decades and ever-increasing censorship has deleted other opinions and information why wouldn’t that lead to a perceptual reframing on a mass scale? I have described already cradle-to-grave programming and in more recent times the targeting of young minds from birth to adulthood has entered the stratosphere. This has taken the form of skewing what is ‘taught’ to fit the Cult agenda and the omnipresent techniques of group-think to isolate non-believers and pressure them into line. There has always been a tendency to follow the herd, but we really are in a new world now in relation to that. We have parents who can see the ‘Covid’ hoax told by their children not to stop them wearing masks at school, being ‘Covid’ tested or having the ‘vaccine’ in fear of the peer-pressure consequences of being different. What is ‘peer-pressure’ if not pressure to conform to group-think? Renegade Minds never group-think and always retain a set of perceptions that are unique to them. Group-think is always underpinned by consequences for not group-thinking. Abuse now aimed at those refusing DNA-manipulating ‘Covid vaccines’ are a potent example of this. The biggest pressure to conform comes from the very group which is itself being manipulated. ‘I am programmed to be part of a hive mind and so you must be.’

Woke control structures in ‘education’ now apply to every mainstream organisation. Those at the top of the ‘education’ hierarchy (the Cult) decide the policy. This is imposed on governments through the Cult network; governments impose it on schools, colleges and universities; their leadership impose the policy on teachers and academics and they impose it on children and students. At any level where there is resistance, perhaps from a teacher or university lecturer, they are targeted by the authorities and often fired. Students themselves regularly demand the dismissal of academics (increasingly few) at odds with the narrative that the students have been programmed to believe in. It is quite a thought that students who are being targeted by the Cult become so consumed by programmed group-think that they launch protests and demand the removal of those who are trying to push back against those targeting the students. Such is the scale of perceptual inversion. We see this with ‘Covid’ programming as the Cult imposes the rules via psycho-psychologists and governments on shops, transport companies and businesses which impose them on their staff who impose them on their customers who pressure Pushbackers to conform to the will of the Cult which is in the process of destroying them and their families. Scan all aspects of society and you will see the same sequence every time.

Fact free Woke and hijacking the 'left'

There is no more potent example of this than 'Woke', a mentality only made possible by the deletion of factual evidence by an 'education' system seeking to produce an ever more uniform society. Why would you bother with facts when you don't know any? Deletion of credible history both in volume and type is highly relevant. Orwell said: 'Who controls the past controls the future: who controls the present controls the past.' They who control the perception of the past control the perception of the future and they who control the present control the perception of the past through the writing and deleting of history. Why would you oppose the imposition of Marxism in the name of Wokeism when you don't know that Marxism cost at least 100 million lives in the 20th century alone? Watch videos and read reports in which Woker generations are asked basic historical questions – it's mind-blowing. A survey of 2,000 people found that six percent of millennials (born approximately early1980s to early 2000s) believed the Second World War (1939-1945) broke out with the assassination of President Kennedy (in 1963) and one in ten thought Margaret Thatcher was British Prime Minister at the time. She was in office between 1979 and 1990. We are in a post-fact society. Provable facts are no defence against the fascism of political correctness or Silicon Valley censorship. Facts don't matter anymore as we have witnessed with the 'Covid' hoax. Sacrificing uniqueness to the Woke group-think religion is all you are required to do and that means thinking for yourself is the biggest Woke no, no. All religions are an expression of group-think and censorship and Woke is just another religion with an orthodoxy defended by group-think and censorship. Burned at the stake becomes burned on Twitter which leads back eventually to burned at the stake as Woke humanity regresses to ages past.

The biggest Woke inversion of all is its creators and funders. I grew up in a traditional left of centre political household on a council estate in Leicester in the 1950s and 60s – you know, the left that challenged the power of wealth-hoarding elites and threats to freedom of speech and opinion. In those days students went on marches defending freedom of speech while today's Wokers march for its deletion. What on earth could have happened? Those very elites (collectively the Cult) that we opposed in my youth and early life have funded into existence the antithesis of that former left and hijacked the 'brand' while inverting everything it ever stood for. We have a mentality that calls itself 'liberal' and 'progressive' while acting like fascists. Cult billionaires and their corporations have funded themselves into control of 'education' to ensure that Woke programming is unceasing throughout the formative years of children and young people and that non-Wokers are isolated (that word again) whether they be students, teachers or college professors. The Cult has funded into existence the now colossal global network of Woke organisations that have spawned and promoted all the 'causes' on the Cult wish-list for global transformation and turned Wokers into demanders of them. Does anyone really think it's a coincidence that the Cult agenda for humanity is a carbon (sorry) copy of the societal transformations desired by Woke?? These are only some of them:

Political correctness: The means by which the Cult deletes all public debates that it knows it cannot win if we had the free-flow of information and evidence.

Human-caused 'climate change': The means by which the Cult seeks to transform society into a globally-controlled dictatorship imposing its will over the fine detail of everyone's lives 'to save the planet' which doesn't actually need saving.

Transgender obsession: Preparing collective perception to accept the ‘new human’ which would not have genders because it would be created technologically and not through procreation. I’ll have much more on this in Human 2.0.

Race obsession: The means by which the Cult seeks to divide and rule the population by triggering racial division through the perception that society is more racist than ever when the opposite is the case. Is it perfect in that regard? No. But to compare today with the racism of apartheid and segregation brought to an end by the civil rights movement in the 1960s is to insult the memory of that movement and inspirations like Martin Luther King. Why is the ‘anti-racism’ industry (which it is) so dominated by privileged white people?

White supremacy: This is a label used by privileged white people to demonise poor and deprived white people pushing back on tyranny to marginalise and destroy them. White people are being especially targeted as the dominant race by number within Western society which the Cult seeks to transform in its image. If you want to change a society you must weaken and undermine its biggest group and once you have done that by using the other groups you next turn on them to do the same … ‘Then they came for the Jews and I was not a Jew so I did nothing.’

Mass migration: The mass movement of people from the Middle East, Africa and Asia into Europe, from the south into the United States and from Asia into Australia are another way the Cult seeks to dilute the racial, cultural and political influence of white people on Western society. White people ask why their governments appear to be working against them while being politically and culturally biased towards incoming cultures. Well, here’s your answer. In the same way sexually ‘straight’ people, men and women, ask why the authorities are biased against them in favour of other sexualities. The answer is the same – that's the way the Cult wants it to be for very sinister motives.

These are all central parts of the Cult agenda and central parts of the Woke agenda and Woke was created and continues to be funded to an immense degree by Cult billionaires and corporations. If anyone begins to say 'coincidence' the syllables should stick in their throat.

Billionaire 'social justice warriors'

Joe Biden is a 100 percent-owned asset of the Cult and the Wokers' man in the White House whenever he can remember his name and for however long he lasts with his rapidly diminishing cognitive function. Even walking up the steps of an aircraft without falling on his arse would appear to be a challenge. He's not an empty-shell puppet or anything. From the minute Biden took office (or the Cult did) he began his executive orders promoting the Woke wish-list. You will see the Woke agenda imposed ever more severely because it's really the *Cult* agenda. Woke organisations and activist networks spawned by the Cult are funded to the extreme so long as they promote what the Cult wants to happen. Woke is funded to promote 'social justice' by billionaires who become billionaires by destroying social justice. The social justice mantra is only a cover for dismantling social justice and funded by billionaires that couldn't give a damn about social justice. Everything makes sense when you see that. One of Woke's premier funders is Cult billionaire financier George Soros who said: 'I am basically there to make money, I cannot and do not look at the social consequences of what I do.' This is the same Soros who has given more than $32 billion to his Open Society Foundations global Woke network and funded Black Lives Matter, mass immigration into Europe and the United States, transgender activism, climate change activism, political correctness and groups targeting 'white supremacy' in the form of privileged white thugs that dominate Antifa. What a scam it all is and when you are dealing with the unquestioning fact-free zone of Woke scamming them is child's play. All you need to pull it off in all these organisations are a few in-the-know agents of the Cult and an army of naïve, reframed, uninformed, narcissistic, know-nothings convinced of their own self-righteousness, self-purity and virtue.

Soros and fellow billionaires and billionaire corporations have poured hundreds of millions into Black Lives Matter and connected groups and promoted them to a global audience. None of this is motivated by caring about black people. These are the billionaires that have controlled and exploited a system that leaves millions of black people in abject poverty and deprivation which they do absolutely nothing to address. The same Cult networks funding BLM were behind the *slave trade!* Black Lives Matter hijacked a phrase that few would challenge and they have turned this laudable concept into a political weapon to divide society. You know that BLM is a fraud when it claims that *All* Lives Matter, the most inclusive statement of all, is 'racist'. BLM and its Cult masters don't want to end racism. To them it's a means to an end to control all of humanity never mind the colour, creed, culture or background. What has destroying the nuclear family got to do with ending racism? Nothing – but that is one of the goals of BLM and also happens to be a goal of the Cult as I have been exposing in my books for decades. Stealing children from loving parents and giving schools ever more power to override parents is part of that same agenda. BLM is a Marxist organisation and why would that not be the case when the Cult created Marxism *and* BLM? Patrisse Cullors, a BLM co-founder, said in a 2015 video that she and her fellow organisers, including co-founder Alicia Garza, are 'trained Marxists'. The lady known after marriage as Patrisse Khan-Cullors bought a $1.4 million home in 2021 in one of the whitest areas of California with a black population of just 1.6 per cent and has so far bought *four* high-end homes for a total of $3.2 million. How very Marxist. There must be a bit of spare in the BLM coffers, however, when Cult corporations and billionaires have handed over the best part of $100 million. Many black people can see that Black Lives Matter is not working for them, but against them, and this is still more confirmation. Black journalist Jason Whitlock, who had his account suspended by Twitter for simply linking to the story about the 'Marxist's' home buying spree, said that BLM leaders are 'making millions of dollars off the backs of these dead black men who they wouldn't spit on if they were on fire and alive'.

Black Lies Matter

Cult assets and agencies came together to promote BLM in the wake of the death of career criminal George Floyd who had been jailed a number of times including for forcing his way into the home of a black woman with others in a raid in which a gun was pointed at her stomach. Floyd was filmed being held in a Minneapolis street in 2020 with the knee of a police officer on his neck and he subsequently died. It was an appalling thing for the officer to do, but the same technique has been used by police on peaceful protestors of lockdown without any outcry from the Woke brigade. As unquestioning supporters of the Cult agenda Wokers have supported lockdown and all the 'Covid' claptrap while attacking anyone standing up to the tyranny imposed in its name. Court documents would later include details of an autopsy on Floyd by County Medical Examiner Dr Andrew Baker who concluded that Floyd had taken a fatal level of the drug fentanyl. None of this mattered to fact-free, question-free, Woke. Floyd's death was followed by worldwide protests against police brutality amid calls to defund the police. Throwing babies out with the bathwater is a Woke speciality. In the wake of the murder of British woman Sarah Everard a Green Party member of the House of Lords, Baroness Jones of Moulescoomb (Nincompoopia would have been better), called for a 6pm curfew for all men. This would be in breach of the Geneva Conventions on war crimes which ban collective punishment, but that would never have crossed the black and white Woke mind of Baroness Nincompoopia who would have been far too convinced of her own self-righteousness to compute such details. Many American cities did defund the police in the face of Floyd riots and after $15 million was deleted from the police budget in Washington DC under useless Woke mayor Muriel Bowser car-jacking alone rose by 300 percent and within six months the US capital recorded its highest murder rate in 15 years. The same happened in Chicago and other cities in line with the Cult/Soros plan to bring fear to streets and neighbourhoods by reducing the police, releasing violent criminals and not prosecuting crime. This is the mob-rule agenda that I have warned in the books was coming for so long. Shootings in the area of Minneapolis where Floyd was arrested increased by 2,500 percent compared with the year before. Defunding the police over George Floyd has led to a big increase in dead people with many of them black. Police protection for politicians making these decisions stayed the same or increased as you would expect from professional hypocrites. The Cult doesn't actually want to abolish the police. It wants to abolish local control over the police and hand it to federal government as the psychopaths advance the Hunger Games Society. Many George Floyd protests turned into violent riots with black stores and businesses destroyed by fire and looting across America fuelled by Black Lives Matter. Woke doesn't do irony. If you want civil rights you must loot the liquor store and the supermarket and make off with a smart TV. It's the only way.

It's not a race war – it's a class war

Black people are patronised by privileged blacks and whites alike and told they are victims of white supremacy. I find it extraordinary to watch privileged blacks supporting the very system and bloodline networks behind the slave trade and parroting the same Cult-serving manipulative crap of their privileged white, often billionaire, associates. It is indeed not a race war but a class war and colour is just a diversion. Black Senator Cory Booker and black Congresswoman Maxine Waters, more residents of Nincompoopia, personify this. Once you tell people they are victims of someone else you devalue both their own responsibility for their plight and the power they have to impact on their reality and experience. Instead we have: 'You are only in your situation because of whitey – turn on them and everything will change.' It won't change. Nothing changes in our lives unless *we* change it. Crucial to that is never seeing yourself as a victim and always as the creator of your reality. Life is a simple sequence of choice and consequence. Make different choices and you create different consequences. *You* have to make those choices – not Black Lives Matter, the Woke Mafia and anyone else that seeks to dictate your life. Who are they these Wokers, an emotional and psychological road traffic accident, to tell you what to do? Personal empowerment is the last thing the Cult and its Black Lives Matter want black people or anyone else to have. They claim to be defending the underdog while *creating* and perpetuating the underdog. The Cult's worst nightmare is human unity and if they are going to keep blacks, whites and every other race under economic servitude and control then the focus must be diverted from what they have in common to what they can be manipulated to believe divides them. Blacks have to be told that their poverty and plight is the fault of the white bloke living on the street in the same poverty and with the same plight they are experiencing. The difference is that your plight black people is due to him, a white supremacist with 'white privilege' living on the street. Don't unite as one human family against your mutual oppressors and suppressors – fight the oppressor with the white face who is as financially deprived as you are. The Cult knows that as its 'Covid' agenda moves into still new levels of extremism people are going to respond and it has been spreading the seeds of disunity everywhere to stop a united response to the evil that targets *all of us*.

Racist attacks on 'whiteness' are getting ever more outrageous and especially through the American Democratic Party which has an appalling history for anti-black racism. Barack Obama, Joe Biden, Hillary Clinton and Nancy Pelosi all eulogised about Senator Robert Byrd at his funeral in 2010 after a nearly 60-year career in Congress. Byrd was a brutal Ku Klux Klan racist and a violent abuser of Cathy O'Brien in MKUltra. He said he would never fight in the military 'with a negro by my side' and 'rather I should die a thousand times, and see Old Glory trampled in the dirt never to rise again, than to see this beloved land of ours become degraded by race mongrels, a throwback to the blackest specimen from the wilds'. Biden called Byrd a 'very close friend and mentor'. These 'Woke' hypocrites are not anti-racist they are anti-poor and anti-people not of their perceived class. Here is an illustration of the scale of anti-white racism to which we have now descended. Seriously Woke and moronic *New York Times* contributor Damon Young described whiteness as a 'virus' that 'like other viruses will not die until there are no bodies left for it to infect'. He went on: '… the only way to stop it is to locate it, isolate it, extract it, and kill it.' Young can say that as a black man with no consequences when a white man saying the same in reverse would be facing a jail sentence. *That's* racism. We had super-Woke numbskull senators Tammy Duckworth and Mazie Hirono saying they would object to future Biden Cabinet appointments if he did not nominate more Asian Americans and Pacific Islanders. Never mind the ability of the candidate what do they look like? Duckworth said: 'I will vote for racial minorities and I will vote for LGBTQ, but anyone else I'm not voting for.' Appointing people on the grounds of race is illegal, but that was not a problem for this ludicrous pair. They were on-message and that's a free pass in any situation.

Critical race racism

White children are told at school they are intrinsically racist as they are taught the divisive 'critical race theory'. This claims that the law and legal institutions are inherently racist and that race is a socially constructed concept used by white people to further their economic and political interests at the expense of people of colour. White is a 'virus' as we've seen. Racial inequality results from 'social, economic, and legal differences that white people create between races to maintain white interests which leads to poverty and criminality in minority communities'. I must tell that to the white guy sleeping on the street. The principal of East Side Community School in New York sent white parents a manifesto that called on them to become ‘white traitors’ and advocate for full ‘white abolition’. These people are teaching your kids when they urgently need a psychiatrist. The ‘school’ included a chart with ‘eight white identities’ that ranged from ‘white supremacist’ to ‘white abolition’ and defined the behaviour white people must follow to end ‘the regime of whiteness’. Woke blacks and their privileged white associates are acting exactly like the slave owners of old and Ku Klux Klan racists like Robert Byrd. They are too full of their own self-purity to see that, but it’s true. Racism is not a body type; it’s a state of mind that can manifest through any colour, creed or culture.

Another racial fraud is *‘equity’*. Not equality of treatment and opportunity – equity. It’s a term spun as equality when it means something very different. Equality in its true sense is a raising up while ‘equity’ is a race to the bottom. Everyone in the same level of poverty is ‘equity’. Keep everyone down – that’s equity. The Cult doesn’t want anyone in the human family to be empowered and BLM leaders, like all these ‘anti-racist’ organisations, continue their privileged, pampered existence by perpetuating the perception of gathering racism. When is the last time you heard an ‘anti-racist’ or ‘anti-Semitism’ organisation say that acts of racism and discrimination have *fallen?* It’s not in the interests of their fund-raising and power to influence and the same goes for the professional soccer anti-racism operation, Kick It Out. Two things confirmed that the Black Lives Matter riots in the summer of 2020 were Cult creations. One was that while anti-lockdown protests were condemned in this same period for ‘transmitting ‘Covid’ the authorities supported mass gatherings of Black Lives Matter supporters. I even saw self-deluding people claiming to be doctors say the two types of protest were not the same. No – the non-existent ‘Covid’ was in favour of lockdowns and attacked those that protested against them while ‘Covid’ supported Black Lives Matter and kept well away from its protests. The whole thing was a joke and as lockdown protestors were arrested, often brutally, by reframed Face-Nappies we had the grotesque sight of police officers taking the knee to Black Lives Matter, a Cult-funded Marxist organisation that supports violent riots and wants to destroy the nuclear family and white people.

He's not white? Shucks!

Woke obsession with race was on display again when ten people were shot dead in Boulder, Colorado, in March, 2021. Cult-owned Woke TV channels like CNN said the shooter appeared to be a white man and Wokers were on Twitter condemning 'violent white men' with the usual mantras. Then the shooter's name was released as Ahmad Al Aliwi Alissa, an anti-Trump Arab-American, and the sigh of disappointment could be heard five miles away. Never mind that ten people were dead and what that meant for their families. Race baiting was all that mattered to these sick Cult-serving people like Barack Obama who exploited the deaths to further divide America on racial grounds which is his job for the Cult. This is the man that 'racist' white Americans made the first black president of the United States and then gave him a second term. Not-very-bright Obama has become filthy rich on the back of that and today appears to have a big influence on the Biden administration. Even so he's still a downtrodden black man and a victim of white supremacy. This disingenuous fraud reveals the contempt he has for black people when he puts on a Deep South Alabama accent whenever he talks to them, no, *at* them.

Another BLM red flag was how the now fully-Woke (fully-Cult) and fully-virtue-signalled professional soccer authorities had their teams taking the knee before every match in support of Marxist Black Lives Matter. Soccer authorities and clubs displayed 'Black Lives Matter' on the players' shirts and flashed the name on electronic billboards around the pitch. Any fans that condemned what is a Freemasonic taking-the-knee ritual were widely condemned as you would expect from the Woke virtue-signallers of professional sport and the now fully-Woke media. We have reverse racism in which you are banned from criticising any race or culture except for white people for whom anything goes – say what you like, no problem. What has this got to do with racial harmony and equality? We've had black supremacists from Black Lives Matter telling white people to fall to their knees in the street and apologise for their white supremacy. Black supremacists acting like white supremacist slave owners of the past couldn't breach their self-obsessed, race-obsessed sense of self-purity. Joe Biden appointed a race-obsessed black supremacist Kristen Clarke to head the Justice Department Civil Rights Division. Clarke claimed that blacks are endowed with 'greater mental, physical and spiritual abilities' than whites. If anyone reversed that statement they would be vilified. Clarke is on-message so no problem. She's never seen a black-white situation in which the black figure is anything but a virtuous victim and she heads the Civil Rights Division which should treat everyone the same or it isn't civil rights. Another perception of the Renegade Mind: If something or someone is part of the Cult agenda they will be supported by Woke governments and media no matter what. If they're not, they will be condemned and censored. It really is that simple and so racist Clarke prospers despite (make that because of) her racism.

The end of culture

Biden's administration is full of such racial, cultural and economic bias as the Cult requires the human family to be divided into warring factions. We are now seeing racially-segregated graduations and everything, but everything, is defined through the lens of perceived 'racism. We have 'racist' mathematics, 'racist' food and even 'racist' *plants*. World famous Kew Gardens in London said it was changing labels on plants and flowers to tell its pre-'Covid' more than two million visitors a year how racist they are. Kew director Richard Deverell said this was part of an effort to 'move quickly to decolonise collections' after they were approached by one Ajay Chhabra 'an actor with an insight into how sugar cane was linked to slavery'. They are *plants* you idiots. 'Decolonisation' in the Woke manual really means colonisation of society with its mentality and by extension colonisation by the Cult. We are witnessing a new Chinese-style 'Cultural Revolution' so essential to the success of all

Marxist takeovers. Our cultural past and traditions have to be swept away to allow a new culture to be built-back-better. Woke targeting of long-standing Western cultural pillars including historical monuments and cancelling of historical figures is what happened in the Mao revolution in China which 'purged remnants of capitalist and traditional elements from Chinese society' and installed Maoism as the dominant ideology'. For China see the Western world today and for 'dominant ideology' see Woke. Better still see Marxism or Maoism. The 'Covid' hoax has specifically sought to destroy the arts and all elements of Western culture from people meeting in a pub or restaurant to closing theatres, music venues, sports stadiums, places of worship and even banning *singing*. Destruction of Western society is also why criticism of any religion is banned except for Christianity which again is the dominant religion as white is the numerically-dominant race. Christianity may be fading rapidly, but its history and traditions are weaved through the fabric of Western society. Delete the pillars and other structures will follow until the whole thing collapses. I am not a Christian defending that religion when I say that. I have no religion. It's just a fact. To this end Christianity has itself been turned Woke to usher its own downfall and its ranks are awash with 'change agents' – knowing and unknowing – at every level including Pope Francis (*definitely* knowing) and the clueless Archbishop of Canterbury Justin Welby (possibly not, but who can be sure?). Woke seeks to coordinate attacks on Western culture, traditions, and ways of life through 'intersectionality' defined as 'the complex, cumulative way in which the effects of multiple forms of discrimination (such as racism, sexism, and classism) combine, overlap, or intersect especially in the experiences of marginalised individuals or groups'. Wade through the Orwellian Woke-speak and this means coordinating disparate groups in a common cause to overthrow freedom and liberal values.

The entire structure of public institutions has been infested with Woke – government at all levels, political parties, police, military, schools, universities, advertising, media and trade unions. This abomination has been achieved through the Cult web by appointing

Wokers to positions of power and battering non-Wokers into line through intimidation, isolation and threats to their job. Many have been fired in the wake of the empathy-deleted, vicious hostility of 'social justice' Wokers and the desire of gutless, spineless employers to virtue-signal their Wokeness. Corporations are filled with Wokers today, most notably those in Silicon Valley. Ironically at the top they are not Woke at all. They are only exploiting the mentality their Cult masters have created and funded to censor and enslave while the Wokers cheer them on until it's their turn. Thus the Woke 'liberal left' is an inversion of the traditional liberal left. Campaigning for justice on the grounds of power and wealth distribution has been replaced by campaigning for identity politics. The genuine traditional left would never have taken money from today's billionaire abusers of fairness and justice and nor would the billionaires have wanted to fund that genuine left. It would not have been in their interests to do so. The division of opinion in those days was between the haves and have nots. This all changed with Cult manipulated and funded identity politics. The division of opinion today is between Wokers and non-Wokers and not income brackets. Cult corporations and their billionaires may have taken wealth disparity to cataclysmic levels of injustice, but as long as they speak the language of Woke, hand out the dosh to the Woke network and censor the enemy they are 'one of us'. Billionaires who don't give a damn about injustice are laughing at them till their bellies hurt. Wokers are not even close to self-aware enough to see that. The transformed 'left' dynamic means that Wokers who drone on about 'social justice' are funded by billionaires that have destroyed social justice the world over. It's *why* they are billionaires.

The climate con

Nothing encapsulates what I have said more comprehensively than the hoax of human-caused global warming. I have detailed in my books over the years how Cult operatives and organisations were the pump-primers from the start of the climate con. A purpose-built vehicle for this is the Club of Rome established by the Cult in 1968 with the Rockefellers and Rothschilds centrally involved all along. Their gofer frontman Maurice Strong, a Canadian oil millionaire, hosted the Earth Summit in Rio de Janeiro, Brazil, in 1992 where the global 'green movement' really expanded in earnest under the guiding hand of the Cult. The Earth Summit established Agenda 21 through the Cult-created-and-owned United Nations to use the illusion of human-caused climate change to justify the transformation of global society to save the world from climate disaster. It is a No-Problem-Reaction-Solution sold through governments, media, schools and universities as whole generations have been terrified into believing that the world was going to end in their lifetimes unless what old people had inflicted upon them was stopped by a complete restructuring of how everything is done. Chill, kids, it's all a hoax. Such restructuring is precisely what the Cult agenda demands (purely by coincidence of course). Today this has been given the codename of the Great Reset which is only an updated term for Agenda 21 and its associated Agenda 2030. The latter, too, is administered through the UN and was voted into being by the General Assembly in 2015. Both 21 and 2030 seek centralised control of all resources and food right down to the raindrops falling on your own land. These are some of the demands of Agenda 21 established in 1992. See if you recognise this society emerging today:

- End national sovereignty
- State planning and management of all land resources, ecosystems, deserts, forests, mountains, oceans and fresh water; agriculture; rural development; biotechnology; and ensuring *'equity'*
- The state to 'define the role' of business and financial resources
- Abolition of private property
- 'Restructuring' the family unit (see BLM)
- Children raised by the state
- People told what their job will be
- Major restrictions on movement
- Creation of 'human settlement zones'

- Mass resettlement as people are forced to vacate land where they live
- Dumbing down education
- Mass global depopulation in pursuit of all the above

The United Nations was created as a Trojan horse for world government. With the climate con of critical importance to promoting that outcome you would expect the UN to be involved. Oh, it's involved all right. The UN is promoting Agenda 21 and Agenda 2030 justified by 'climate change' while also driving the climate hoax through its Intergovernmental Panel on Climate Change (IPCC), one of the world's most corrupt organisations. The IPCC has been lying ferociously and constantly since the day it opened its doors with the global media hanging unquestioningly on its every mendacious word. The Green movement is entirely Woke and has long lost its original environmental focus since it was co-opted by the Cult. An obsession with 'global warming' has deleted its values and scrambled its head. I experienced a small example of what I mean on a beautiful country walk that I have enjoyed several times a week for many years. The path merged into the fields and forests and you felt at one with the natural world. Then a 'Green' organisation, the Hampshire and Isle of Wight Wildlife Trust, took over part of the land and proceeded to cut down a large number of trees, including mature ones, to install a horrible big, bright steel 'this-is-ours-stay-out' fence that destroyed the whole atmosphere of this beautiful place. No one with a feel for nature would do that. Day after day I walked to the sound of chainsaws and a magnificent mature weeping willow tree that I so admired was cut down at the base of the trunk. When I challenged a Woke young girl in a green shirt (of course) about this vandalism she replied: 'It's a weeping willow – it will grow back.' This is what people are paying for when they donate to the Hampshire and Isle of Wight Wildlife Trust and many other 'green' organisations today. It is not the environmental movement that I knew and instead has become a support-system – as with Extinction Rebellion – for a very dark agenda.

Private jets for climate justice

The Cult-owned, Gates-funded, World Economic Forum and its founder Klaus Schwab were behind the emergence of Greta Thunberg to harness the young behind the climate agenda and she was invited to speak to the world at … the UN. Schwab published a book, *Covid-19: The Great Reset* in 2020 in which he used the 'Covid' hoax and the climate hoax to lay out a new society straight out of Agenda 21 and Agenda 2030. Bill Gates followed in early 2021 when he took time out from destroying the world to produce a book in his name about the way to save it. Gates flies across the world in private jets and admitted that 'I probably have one of the highest greenhouse gas footprints of anyone on the planet … my personal flying alone is gigantic.' He has also bid for the planet's biggest private jet operator. Other climate change saviours who fly in private jets include John Kerry, the US Special Presidential Envoy for Climate, and actor Leonardo DiCaprio, a 'UN Messenger of Peace with special focus on climate change'. These people are so full of bullshit they could corner the market in manure. We mustn't be sceptical, though, because the Gates book, *How to Avoid a Climate Disaster: The Solutions We Have and the Breakthroughs We Need*, is a genuine attempt to protect the world and not an obvious pile of excrement attributed to a mega-psychopath aimed at selling his masters' plans for humanity. The Gates book and the other shite-pile by Klaus Schwab could have been written by the same person and may well have been. Both use 'climate change' and 'Covid' as the excuses for their new society and by coincidence the Cult's World Economic Forum and Bill and Melinda Gates Foundation promote the climate hoax and hosted Event 201 which pre-empted with a 'simulation' the very 'coronavirus' hoax that would be simulated for real on humanity within weeks. The British 'royal' family is promoting the 'Reset' as you would expect through Prince 'climate change caused the war in Syria' Charles and his hapless son Prince William who said that we must 'reset our relationship with nature and our trajectory as a species' to avoid a climate disaster. Amazing how many promotors of the 'Covid' and 'climate change' control systems are connected to Gates and the World Economic Forum. A ‘study’ in early 2021 claimed that carbon dioxide emissions must fall by the equivalent of a global lockdown roughly every two years for the next decade to save the planet. The ‘study’ appeared in the same period that the Schwab mob claimed in a video that lockdowns destroying the lives of billions are good because they make the earth ‘quieter’ with less ‘ambient noise’. They took down the video amid a public backlash for such arrogant, empathy-deleted stupidity You see, however, where they are going with this. Corinne Le Quéré, a professor at the Tyndall Centre for Climate Change Research, University of East Anglia, was lead author of the climate lockdown study, and she writes for … the World Economic Forum. Gates calls in ‘his’ book for changing ‘every aspect of the economy’ (long-time Cult agenda) and for humans to eat synthetic ‘meat’ (predicted in my books) while cows and other farm animals are eliminated. Australian TV host and commentator Alan Jones described what carbon emission targets would mean for farm animals in Australia alone if emissions were reduced as demanded by 35 percent by 2030 and zero by 2050:

> Well, let’s take agriculture, the total emissions from agriculture are about 75 million tonnes of carbon dioxide, equivalent. Now reduce that by 35 percent and you have to come down to 50 million tonnes, I’ve done the maths. So if you take for example 1.5 million cows, you’re going to have to reduce the herd by 525,000 [by] 2030, nine years, that’s 58,000 cows a year. The beef herd’s 30 million, reduce that by 35 percent, that’s 10.5 million, which means 1.2 million cattle have to go every year between now and 2030. This is insanity!
>
> There are 75 million sheep. Reduce that by 35 percent, that’s 26 million sheep, that’s almost 3 million a year. So under the Paris Agreement over 30 million beasts. dairy cows, cattle, pigs and sheep would go. More than 8,000 every minute of every hour for the next decade, do these people know what they’re talking about?

Clearly they don’t at the level of campaigners, politicians and administrators. The Cult *does* know; that’s the outcome it wants. We are faced with not just a war on humanity. Animals and the natural world are being targeted and I have been saying since the ‘Covid’ hoax began that the plan eventually was to claim that the ‘deadly virus’ is able to jump from animals, including farm animals and domestic pets, to humans. Just before this book went into production came this story: 'Russia registers world's first Covid-19 vaccine for cats & dogs as makers of Sputnik V warn pets & farm animals could spread virus'. The report said 'top scientists warned that the deadly pathogen could soon begin spreading through homes and farms' and 'the next stage is the infection of farm and domestic animals'. Know the outcome and you'll see the journey. Think what that would mean for animals and keep your eye on a term called zoonosis or zoonotic diseases which transmit between animals and humans. The Cult wants to break the connection between animals and people as it does between people and people. Farm animals fit with the Cult agenda to transform food from natural to synthetic.

The gas of life is killing us

There can be few greater examples of Cult inversion than the condemnation of carbon dioxide as a dangerous pollutant when it is the gas of life. Without it the natural world would be dead and so we would all be dead. We breathe in oxygen and breathe out carbon dioxide while plants produce oxygen and absorb carbon dioxide. It is a perfect symbiotic relationship that the Cult wants to dismantle for reasons I will come to in the final two chapters. Gates, Schwab, other Cult operatives and mindless repeaters, want the world to be 'carbon neutral' by at least 2050 and the earlier the better. 'Zero carbon' is the cry echoed by lunatics calling for 'Zero Covid' when we already have it. These carbon emission targets will deindustrialise the world in accordance with Cult plans – the post-industrial, post-democratic society – and with so-called renewables like solar and wind not coming even close to meeting human energy needs blackouts and cold are inevitable. Texans got the picture in the winter of 2021 when a snow storm stopped wind turbines and solar panels from working and the lights went down along with water which relies on electricity for its supply system. Gates wants everything to be powered by electricity to ensure that his masters have the kill switch to stop all human activity, movement, cooking, water and warmth any time they like. The climate lie is so stupendously inverted that it claims we must urgently reduce carbon dioxide when we *don't have enough*.

Co2 in the atmosphere is a little above 400 parts per million when the optimum for plant growth is 2,000 ppm and when it falls anywhere near 150 ppm the natural world starts to die and so do we. It fell to as low as 280 ppm in an 1880 measurement in Hawaii and rose to 413 ppm in 2019 with industrialisation which is why the planet has become *greener* in the industrial period. How insane then that psychopathic madman Gates is not satisfied only with blocking the rise of Co2. He's funding technology to suck it out of the atmosphere. The reason why will become clear. The industrial era is not destroying the world through Co2 and has instead turned around a potentially disastrous ongoing fall in Co2. Greenpeace co-founder and scientist Patrick Moore walked away from Greenpeace in 1986 and has exposed the green movement for fear-mongering and lies. He said that 500 million years ago there was *17 times* more Co2 in the atmosphere than we have today and levels have been falling for hundreds of millions of years. In the last 150 million years Co2 levels in Earth's atmosphere had reduced by *90 percent*. Moore said that by the time humanity began to unlock carbon dioxide from fossil fuels we were at '38 seconds to midnight' and in that sense: 'Humans are [the Earth's] salvation.' Moore made the point that only half the Co2 emitted by fossil fuels stays in the atmosphere and we should remember that all pollution pouring from chimneys that we are told is carbon dioxide is in fact nothing of the kind. It's pollution. Carbon dioxide is an invisible gas.

William Happer, Professor of Physics at Princeton University and long-time government adviser on climate, has emphasised the Co2 deficiency for maximum growth and food production. Greenhouse growers don't add carbon dioxide for a bit of fun. He said that most of the warming in the last 100 years, after the earth emerged from the super-cold period of the 'Little Ice Age' into a natural warming cycle, was over by 1940. Happer said that a peak year for warming in 1988 can be explained by a 'monster El Nino' which is a natural and cyclical warming of the Pacific that has nothing to do with 'climate change'. He said the effect of Co2 could be compared to painting a wall with red paint in that once two or three coats have been applied it didn't matter how much more you slapped on because the wall will not get much redder. Almost all the effect of the rise in Co2 has already happened, he said, and the volume in the atmosphere would now have to *double* to increase temperature by a single degree. Climate hoaxers know this and they have invented the most ridiculously complicated series of 'feedback' loops to try to overcome this rather devastating fact. You hear puppet Greta going on cluelessly about feedback loops and this is why.

The Sun affects temperature? No you *climate denier*

Some other nonsense to contemplate: Climate graphs show that rises in temperature do not follow rises in Co2 – *it's the other way round* with a lag between the two of some 800 years. If we go back 800 years from present time we hit the Medieval Warm Period when temperatures were higher than now without any industrialisation and this was followed by the Little Ice Age when temperatures plummeted. The world was still emerging from these centuries of serious cold when many climate records began which makes the ever-repeated line of the 'hottest year since records began' meaningless when you are not comparing like with like. The coldest period of the Little Ice Age corresponded with the lowest period of sunspot activity when the Sun was at its least active. Proper scientists will not be at all surprised by this when it confirms the obvious fact that earth temperature is affected by the scale of Sun activity and the energetic power that it subsequently emits; but when is the last time you heard a climate hoaxer talking about the Sun as a source of earth temperature?? Everything has to be focussed on Co2 which makes up just 0.117 percent of so-called greenhouse gases and only a fraction of even that is generated by human activity. The rest is natural. More than *90 percent* of those greenhouse gases are water vapour and clouds (Fig 9). Ban moisture I say. Have you noticed that the climate hoaxers no longer use the polar bear as their promotion image? That's because far from becoming extinct polar bear communities are stable or thriving. Joe Bastardi, American meteorologist, weather forecaster and outspoken critic of the climate lie, documents in his book *The Climate Chronicles* how weather patterns and events claimed to be evidence of climate change have been happening since long before industrialisation: 'What happened before naturally is happening again, as is to be expected given the cyclical nature of the climate due to the design of the planet.' If you read the detailed background to the climate hoax in my other books you will shake your head and wonder how anyone could believe the crap which has spawned a multi-trillion dollar industry based on absolute garbage (see HIV causes AIDs and Sars-Cov-2 causes 'Covid-19'). Climate and 'Covid' have much in common given they have the same source. They both have the contradictory *everything* factor in which everything is explained by reference to them. It's hot – 'it's climate change'. It's cold – 'it's climate change'. I got a sniffle – 'it's Covid'. I haven't got a sniffle – 'it's Covid'. Not having a sniffle has to be a symptom of 'Covid'. Everything is and not having a sniffle is especially dangerous if you are a slow walker. For sheer audacity I offer you a Cambridge University 'study' that actually linked 'Covid' to 'climate change'. It had to happen eventually. They concluded that climate change played a role in 'Covid-19' spreading from animals to humans because … wait for it … I kid you not … *the two groups were forced closer together as populations grow*. Er, that's it. The whole foundation on which this depended was that 'Bats are the likely zoonotic origin of SARS-CoV-1 and SARS-CoV-2'. Well, they are not. They are nothing to do with it. Apart from bats not being the origin and therefore 'climate change' effects on bats being irrelevant I am in awe of their academic insight. Where would we be without them? Not where we are that's for sure.

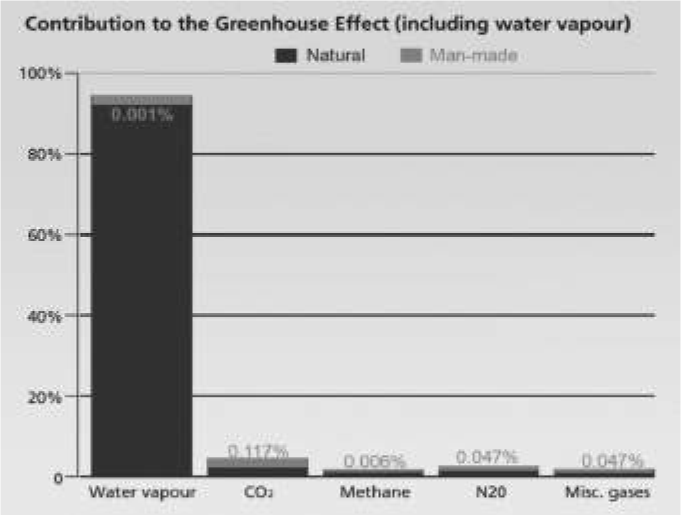


Figure 9: The idea that the gas of life is disastrously changing the climate is an insult to brain cell activity.

One other point about the weather is that climate modification is now well advanced and not every major weather event is natural – or earthquake come to that. I cover this subject at some length in other books. China is openly planning a rapid expansion of its weather modification programme which includes changing the climate in an area more than one and a half times the size of India. China used weather manipulation to ensure clear skies during the 2008 Olympics in Beijing. I have quoted from US military documents detailing how to employ weather manipulation as a weapon of war and they did that in the 1960s and 70s during the conflict in Vietnam with Operation Popeye manipulating monsoon rains for military purposes. Why would there be international treaties on weather modification if it wasn't possible? Of course it is. Weather is energetic information and it can be changed.

How was the climate hoax pulled off? See 'Covid'

If you can get billions to believe in a 'virus' that doesn't exist you can get them to believe in human-caused climate change that doesn't exist. Both are being used by the Cult to transform global society in the way it has long planned. Both hoaxes have been achieved in pretty much the same way. First you declare a lie is a fact. There's a ‘virus’ you call SARS-Cov-2 or humans are warming the planet with their behaviour. Next this becomes, via Cult networks, the foundation of government, academic and science policy and belief. Those who parrot the mantra are given big grants to produce research that confirms the narrative is true and ever more ‘symptoms’ are added to make the ‘virus’/’climate change’ sound even more scary. Scientists and researchers who challenge the narrative have their grants withdrawn and their careers destroyed. The media promote the lie as the unquestionable truth and censor those with an alternative view or evidence. A great percentage of the population believe what they are told as the lie becomes an everybody-knows-that and the believing-masses turn on those with a mind of their own. The technique has been used endlessly throughout human history. Wokers are the biggest promotors of the climate lie *and* ‘Covid’ fascism because their minds are owned by the Cult; their sense of self-righteous self-purity knows no bounds; and they exist in a bubble of reality in which facts are irrelevant and only get in the way of looking without seeing.

Running through all of this like veins in a blue cheese is control of information, which means control of perception, which means control of behaviour, which collectively means control of human society. The Cult owns the global media and Silicon Valley fascists for the simple reason that it *has* to. Without control of information it can’t control perception and through that human society. Examine every facet of the Cult agenda and you will see that anything supporting its introduction is never censored while anything pushing back is always censored. I say again: Psychopaths that know why they are doing this must go before Nuremberg trials and those that follow their orders must trot along behind them into the same dock. ‘I was just following orders’ didn’t work the first time and it must not work now. Nuremberg trials must be held all over the world before public juries for politicians, government officials, police, compliant doctors, scientists and virologists, and all Cult operatives such as Gates, Tedros, Fauci, Vallance, Whitty, Ferguson, Zuckerberg, Wojcicki, Brin, Page, Dorsey, the whole damn lot of them – including, no *especially*, the psychopath psychologists. Without them and the brainless, gutless excuses for journalists that have repeated their lies, none of this could be happening. Nobody can be allowed to escape justice for the psychological and economic Armageddon they are all responsible for visiting upon the human race.

As for the compliant, unquestioning, swathes of humanity, and the self-obsessed, all-knowing ignorance of the Wokers … don’t start me. God help their kids. God help their grandkids. God *help them*.

CHAPTER NINE

We must have it? So what is it?

Well I won't back down. No, I won't back down. You can stand me up at the Gates of Hell. But I won't back down
Tom Petty

I will now focus on the genetically-manipulating 'Covid vaccines' which do not meet this official definition of a vaccine by the US Centers for Disease Control (CDC): 'A product that stimulates a person's immune system to produce immunity to a specific disease, protecting the person from that disease.' On that basis 'Covid vaccines' are not a vaccine in that the makers don't even claim they stop infection or transmission.

They are instead part of a multi-levelled conspiracy to change the nature of the human body and what it means to be 'human' and to depopulate an enormous swathe of humanity. What I shall call Human 1.0 is on the cusp of becoming Human 2.0 and for very sinister reasons. Before I get to the 'Covid vaccine' in detail here's some background to vaccines in general. Government regulators do not test vaccines – the makers do – and the makers control which data is revealed and which isn't. Children in America are given 50 vaccine doses by age six and 69 by age 19 and the effect of the whole combined schedule has never been tested. Autoimmune diseases when the immune system attacks its own body have soared in the mass vaccine era and so has disease in general in children and the young. Why wouldn't this be the case when vaccines target the *immune system?* The US government gave Big Pharma drug companies immunity from prosecution for vaccine death and injury in the 1986 National Childhood Vaccine Injury Act (NCVIA) and since then the government (taxpayer) has been funding compensation for the consequences of Big Pharma vaccines. The criminal and satanic drug giants can't lose and the vaccine schedule has increased dramatically since 1986 for this reason. There is no incentive to make vaccines safe and a big incentive to make money by introducing ever more. Even against a ridiculously high bar to prove vaccine liability, and with the government controlling the hearing in which it is being challenged for compensation, the vaccine court has so far paid out more than $4 billion. These are the vaccines we are told are safe and psychopaths like Zuckerberg censor posts saying otherwise. The immunity law was even justified by a ruling that vaccines by their nature were 'unavoidably unsafe'.

Check out the ingredients of vaccines and you will be shocked if you are new to this. *They put that in children's bodies?? What??* Try aluminium, a brain toxin connected to dementia, aborted foetal tissue and formaldehyde which is used to embalm corpses. World-renowned aluminium expert Christopher Exley had his research into the health effect of aluminium in vaccines shut down by Keele University in the UK when it began taking funding from the Bill and Melinda Gates Foundation. Research when diseases 'eradicated' by vaccines began to decline and you will find the fall began long *before* the vaccine was introduced. Sometimes the fall even plateaued after the vaccine. Diseases like scarlet fever for which there was no vaccine declined in the same way because of environmental and other factors. A perfect case in point is the polio vaccine. Polio began when lead arsenate was first sprayed as an insecticide and residues remained in food products. Spraying started in 1892 and the first US polio epidemic came in Vermont in 1894. The simple answer was to stop spraying, but Rockefeller-created Big Pharma had a better idea. Polio was decreed to be caused by the *poliovirus* which 'spreads from person to person and can infect a person's spinal cord'. Lead arsenate was replaced by the lethal DDT which had the same effect of causing paralysis by damaging the brain and central nervous system. Polio plummeted when DDT was reduced and then banned, but the vaccine is still given the credit for something it didn't do. Today by far the biggest cause of polio is the vaccines promoted by Bill Gates. Vaccine justice campaigner Robert Kennedy Jr, son of assassinated (by the Cult) US Attorney General Robert Kennedy, wrote:

> In 2017, the World Health Organization (WHO) reluctantly admitted that the global explosion in polio is predominantly vaccine strain. The most frightening epidemics in Congo, Afghanistan, and the Philippines, are all linked to vaccines. In fact, by 2018, 70% of global polio cases were vaccine strain.

Vaccines make fortunes for Cult-owned Gates and Big Pharma while undermining the health and immune systems of the population. We had a glimpse of the mentality behind the Big Pharma cartel with a report on WION (World is One News), an international English language TV station based in India, which exposed the extraordinary behaviour of US drug company Pfizer over its 'Covid vaccine'. The WION report told how Pfizer had made fantastic demands of Argentina, Brazil and other countries in return for its 'vaccine'. These included immunity from prosecution, even for Pfizer negligence, government insurance to protect Pfizer from law suits and handing over as collateral sovereign assets of the country to include Argentina's bank reserves, military bases and embassy buildings. Pfizer demanded the same of Brazil in the form of waiving sovereignty of its assets abroad; exempting Pfizer from Brazilian laws; and giving Pfizer immunity from all civil liability. This is a 'vaccine' developed with government funding. Big Pharma is evil incarnate as a creation of the Cult and all must be handed tickets to Nuremberg.

Phantom 'vaccine' for a phantom 'disease'

I'll expose the 'Covid vaccine' fraud and then go on to the wider background of why the Cult has set out to 'vaccinate' every man, woman and child on the planet for an alleged 'new disease' with a survival rate of 99.77 percent (or more) even by the grotesquelymanipulated figures of the World Health Organization and Johns Hopkins University. The 'infection' to 'death' ratio is 0.23 to 0.15 percent according to Stanford epidemiologist Dr John Ioannidis and while estimates vary the danger remains tiny. I say that if the truth be told the fake infection to fake death ratio is zero. Never mind all the evidence I have presented here and in *The Answer* that there is no 'virus' let us just focus for a moment on that death-rate figure of say 0.23 percent. The figure includes all those worldwide who have tested positive with a test not testing for the 'virus' and then died within 28 days or even longer of any other cause – *any other cause*. Now subtract all those illusory 'Covid' deaths on the global data sheets from the 0.23 percent. What do you think you would be left with? *Zero*. A vaccination has never been successfully developed for a so-called coronavirus. They have all failed at the animal testing stage when they caused hypersensitivity to what they were claiming to protect against and made the impact of a disease far worse. Cult-owned vaccine corporations got around that problem this time by bypassing animal trials, going straight to humans and making the length of the 'trials' before the public rollout as short as they could get away with. Normally it takes five to ten years or more to develop vaccines that still cause demonstrable harm to many people and that's without including the long-term effects that are never officially connected to the vaccination. 'Covid' non-vaccines have been officially produced and approved in a matter of months from a standing start and part of the reason is that (a) they were developed before the 'Covid' hoax began and (b) they are based on computer programs and not natural sources. Official non-trials were so short that government agencies gave *emergency,* not full, approval. 'Trials' were not even completed and full approval cannot be secured until they are. Public 'Covid vaccination' is actually a *continuation of the trial*. Drug company 'trials' are not scheduled to end until 2023 by which time a lot of people are going to be dead. Data on which government agencies gave this emergency approval was supplied by the Big Pharma corporations themselves in the form of Pfizer/BioNTech, AstraZeneca, Moderna, Johnson & Johnson, and others, and this is the case with all vaccines. By its very nature *emergency* approval means drug companies do not have to prove that the 'vaccine' is 'safe and effective'. How could they with trials way short of complete? Government regulators only have to *believe* that they *could* be safe and effective. It is criminal manipulation to get products in circulation with no testing worth the name. Agencies giving that approval are infested with Big Pharma-connected place-people and they act in the interests of Big Pharma (the Cult) and not the public about whom they do not give a damn.

More human lab rats

'Covid vaccines' produced in record time by Pfizer/BioNTech and Moderna employ a technique *never approved before for use on humans*. They are known as mRNA 'vaccines' and inject a synthetic version of 'viral' mRNA or 'messenger RNA'. The key is in the term 'messenger'. The body works, or doesn't, on the basis of information messaging. Communications are constantly passing between and within the genetic system and the brain. Change those messages and you change the state of the body and even its very nature and you can change psychology and behaviour by the way the brain processes information. I think you are going to see significant changes in personality and perception of many people who have had the 'Covid vaccine' synthetic potions. Insider Aldous Huxley predicted the following in 1961 and mRNA 'vaccines' can be included in the term 'pharmacological methods':

> There will be, in the next generation or so, a pharmacological method of making people love their servitude, and producing dictatorship without tears, so to speak, producing a kind of painless concentration camp for entire societies, so that people will in fact have their own liberties taken away from them, but rather enjoy it, because they will be distracted from any desire to rebel by propaganda or brainwashing, or brainwashing enhanced by pharmacological methods. And this seems to be the final revolution.

Apologists claim that mRNA synthetic 'vaccines' don't change the DNA genetic blueprint because RNA does not affect DNA only the other way round. This is so disingenuous. A process called 'reverse transcription' can convert RNA into DNA and be integrated into DNA in the cell nucleus. This was highlighted in December, 2020, by scientists at Harvard and Massachusetts Institute of Technology (MIT). Geneticists report that more than 40 percent of mammalian genomes results from reverse transcription. On the most basic level if messaging changes then that sequence must lead to changes in DNA which is receiving and transmitting those communications. How can introducing synthetic material into cells not change the cells where DNA is located? The process is known as transfection which is defined as 'a technique to insert foreign nucleic acid (DNA or RNA) into a cell, typically with the intention of altering the properties of the cell'. Researchers at the Sloan Kettering Institute in New York found that changes in messenger RNA can deactivate tumour-suppressing proteins and thereby promote cancer. This is what happens when you mess with messaging. 'Covid vaccine' maker Moderna was founded in 2010 by Canadian stem cell biologist Derrick J. Rossi after his breakthrough discovery in the field of transforming and reprogramming stem cells. These are neutral cells that can be programmed to become any cell including sperm cells. Moderna was therefore founded on the principle of genetic manipulation and has never produced any vaccine or drug before its genetically-manipulating synthetic 'Covid' shite. Look at the name – Mode-RNA or Modify-RNA. Another important point is that the US Supreme Court has ruled that genetically-modified DNA, or complementary DNA (cDNA) synthesized in the laboratory from messenger RNA, can be patented and owned. These psychopaths are doing this to the human body.

Cells replicate synthetic mRNA in the 'Covid vaccines' and in theory the body is tricked into making antigens which trigger antibodies to target the 'virus spike proteins' which as Dr Tom Cowan said have *never been seen*. Cut the crap and these 'vaccines' deliver *self-replicating* synthetic material to the cells with the effect of changing human DNA. The more of them you have the more that process is compounded while synthetic material is all the time self-replicating. 'Vaccine'-maker Moderna describes mRNA as 'like software for the cell' and so they are messing with the body's software. What happens when you change the software in a computer? Everything changes. For this reason the Cult is preparing a production line of mRNA 'Covid vaccines' and a long list of excuses to use them as with all the 'variants' of a 'virus' never shown to exist. The plan is further to transfer the mRNA technique to other vaccines mostly given to children and young people. The cumulative consequences will be a transformation of human DNA through a constant infusion of synthetic genetic material which will kill many and change the rest. Now consider that governments that have given emergency approval for a vaccine that's not a vaccine; never been approved for humans before; had no testing worth the name; and the makers have been given immunity from prosecution for any deaths or adverse effects suffered by the public. The UK government awarded *permanent legal indemnity* to itself and its employees for harm done when a patient is being treated for 'Covid-19' or 'suspected Covid-19'. That is quite a thought when these are possible 'side-effects' from the 'vaccine' (they are not 'side', they are effects) listed by the US Food and Drug Administration:

> Guillain-Barre syndrome; acute disseminated encephalomyelitis; transverse myelitis; encephalitis; myelitis; encephalomyelitis; meningoencephalitis; meningitis; encephalopathy; convulsions; seizures; stroke; narcolepsy; cataplexy; anaphylaxis; acute myocardial infarction (heart attack); myocarditis; pericarditis; autoimmune disease; death; implications for pregnancy, and birth outcomes; other acute demyelinating diseases; non anaphylactic allergy reactions; thrombocytopenia ; disseminated intravascular coagulation; venous thromboembolism; arthritis; arthralgia; joint pain; Kawasaki disease; multisystem inflammatory syndrome in children; vaccine enhanced disease. The latter is the way the 'vaccine' has the potential to make diseases far worse than they would otherwise be.

UK doctor and freedom campaigner Vernon Coleman described the conditions in this list as ‘all unpleasant, most of them very serious, and you can’t get more serious than death’. The thought that anyone at all has had the ‘vaccine’ in these circumstances is testament to the potential that humanity has for clueless, unquestioning, stupidity and for many that programmed stupidity has already been terminal.

An insider speaks

Dr Michael Yeadon is a former Vice President, head of research and Chief Scientific Adviser at vaccine giant Pfizer. Yeadon worked on the inside of Big Pharma, but that did not stop him becoming a vocal critic of ‘Covid vaccines’ and their potential for multiple harms, including infertility in women. By the spring of 2021 he went much further and even used the no, no, term ‘conspiracy’. When you begin to see what is going on it is impossible not to do so. Yeadon spoke out in an interview with freedom campaigner James Delingpole and I mentioned earlier how he said that no one had samples of ‘the virus’. He explained that the mRNA technique originated in the anti-cancer field and ways to turn on and off certain genes which could be advantageous if you wanted to stop cancer growing out of control. ‘That’s the origin of them. They are a very unusual application, really.’ Yeadon said that treating a cancer patient with an aggressive procedure might be understandable if the alternative was dying, but it was quite another thing to use the same technique as a public health measure. Most people involved wouldn’t catch the infectious agent you were vaccinating against and if they did they probably wouldn’t die:

> If you are really using it as a public health measure you really want to as close as you can get to zero sides-effects … I find it odd that they chose techniques that were really cutting their teeth in the field of oncology and I’m worried that in using gene-based vaccines that have to be injected in the body and spread around the body, get taken up into some cells, and the regulators haven’t quite told us which cells they get taken up into … you are going to be generating a wide range of responses … with multiple steps each of which could go well or badly.

I doubt the Cult intends it to go well. Yeadon said that you can put any gene you like into the body through the 'vaccine'. 'You can certainly give them a gene that would do them some harm if you wanted.' I was intrigued when he said that when used in the cancer field the technique could turn genes on and off. I explore this process in *The Answer* and with different genes having different functions you could create mayhem – physically and psychologically – if you turned the wrong ones on and the right ones off. I read reports of an experiment by researchers at the University of Washington's school of computer science and engineering in which they encoded DNA to infect computers. The body is itself a biological computer and if human DNA can inflict damage on a computer why can't the computer via synthetic material mess with the human body? It can. The Washington research team said it was possible to insert malicious malware into 'physical DNA strands' and corrupt the computer system of a gene sequencing machine as it 'reads gene letters and stores them as binary digits 0 and 1'. They concluded that hackers could one day use blood or spit samples to access computer systems and obtain sensitive data from police forensics labs or infect genome files. It is at this level of digital interaction that synthetic 'vaccines' need to be seen to get the full picture and that will become very clear later on. Michael Yeadon said it made no sense to give the 'vaccine' to younger people who were in no danger from the 'virus'. What was the benefit? It was all downside with potential effects:

> The fact that my government in what I thought was a civilised, rational country, is raining [the 'vaccine'] on people in their 30s and 40s, even my children in their 20s, they're getting letters and phone calls, I know this is not right and any of you doctors who are vaccinating you know it's not right, too. They are not at risk. They are not at risk from the disease, so you are now hoping that the side-effects are so rare that you get away with it. You don't give new technology … that you don't understand to 100 percent of the population.

Blood clot problems with the AstraZeneca 'vaccine' have been affecting younger people to emphasise the downside risks with no benefit. AstraZeneca's version, produced with Oxford University, does not use mRNA, but still gets its toxic cocktail inside cells where it targets DNA. The Johnson & Johnson 'vaccine' which uses a similar technique has also produced blood clot effects to such an extent that the United States paused its use at one point. They are all 'gene therapy' (cell modification) procedures and not 'vaccines'. The truth is that once the content of these injections enter cells we have no idea what the effect will be. People can speculate and some can give very educated opinions and that's good. In the end, though, only the makers know what their potions are designed to do and even they won't know every last consequence. Michael Yeadon was scathing about doctors doing what they knew to be wrong. 'Everyone's mute', he said. Doctors in the NHS must know this was not right, coming into work and injecting people. 'I don't know how they sleep at night. I know I couldn't do it. I know that if I were in that position I'd have to quit.' He said he knew enough about toxicology to know this was not a good risk-benefit. Yeadon had spoken to seven or eight university professors and all except two would not speak out publicly. Their universities had a policy that no one said anything that countered the government and its medical advisors. They were afraid of losing their government grants. This is how intimidation has been used to silence the truth at every level of the system. I say silence, but these people could still speak out if they made that choice. Yeadon called them 'moral cowards' – 'This is about your children and grandchildren's lives and you have just buggered off and left it.'

'Variant' nonsense

Some of his most powerful comments related to the alleged 'variants' being used to instil more fear, justify more lockdowns, and introduce more 'vaccines'. He said government claims about 'variants' were nonsense. He had checked the alleged variant 'codes' and they were 99.7 percent identical to the 'original'. This was the human identity difference equivalent to putting a baseball cap on and off or wearing it the other way round. A 0.3 percent difference would make it impossible for that 'variant' to escape immunity from the 'original'. This made no sense of having new 'vaccines' for ‘variants’. He said there would have to be at least a *30 percent* difference for that to be justified and even then he believed the immune system would still recognise what it was. Gates-funded ‘variant modeller’ and ‘vaccine’-pusher John Edmunds might care to comment. Yeadon said drug companies were making new versions of the ‘vaccine’ as a ‘top up’ for ‘variants’. Worse than that, he said, the ‘regulators’ around the world like the MHRA in the UK had got together and agreed that because ‘vaccines’ for ‘variants’ were so similar to the first ‘vaccines’ *they did not have to do safety studies*. How transparently sinister that is. This is when Yeadon said: ‘There is a conspiracy here.’ There was no need for another vaccine for ‘variants’ and yet we were told that there was and the country had shut its borders because of them. ‘They are going into hundreds of millions of arms without passing ‘go’ or any regulator. Why did they do that? Why did they pick this method of making the vaccine?’

The reason had to be something bigger than that it seemed and ‘it’s not protection against the virus’. It’s was a far bigger project that meant politicians and advisers were willing to do things and not do things that knowingly resulted in avoidable deaths – ‘that’s already happened when you think about lockdown and deprivation of health care for a year.’ He spoke of people prepared to do something that results in the avoidable death of their fellow human beings and it not bother them. This is the penny-drop I have been working to get across for more than 30 years – the level of pure evil we are dealing with. Yeadon said his friends and associates could not believe there could be that much evil, but he reminded them of Stalin, Pol Pot and Hitler and of what Stalin had said: ‘One death is a tragedy. A million? A statistic.’ He could not think of a benign explanation for why you need top-up vaccines ‘which I’m sure you don’t’ and for the regulators ‘to just get out of the way and wave them through’. Why would the regulators do that when they were still wrestling with the dangers of the ‘parent’ vaccine? He was clearly shocked by what he had seen since the ‘Covid’ hoax began and now he was thinking the previously unthinkable:

> If you wanted to depopulate a significant proportion of the world and to do it in a way that doesn't involve destruction of the environment with nuclear weapons, poisoning everyone with anthrax or something like that, and you wanted plausible deniability while you had a multi-year infectious disease crisis, I actually don't think you could come up with a better plan of work than seems to be in front of me. I can't say that's what they are going to do, but I can't think of a benign explanation why they are doing it.

He said he never thought that they would get rid of 99 percent of humans, but now he wondered. 'If you wanted to that this would be a hell of a way to do it – it would be unstoppable folks.' Yeadon had concluded that those who submitted to the 'vaccine' would be allowed to have some kind of normal life (but for how long?) while screws were tightened to coerce and mandate the last few percent. 'I think they'll put the rest of them in a prison camp. I wish I was wrong, but I don't think I am.' Other points he made included: There were no coronavirus vaccines then suddenly they all come along at the same time; we have no idea of the long term affect with trials so short; coercing or forcing people to have medical procedures is against the Nuremberg Code instigated when the Nazis did just that; people should at least delay having the 'vaccine'; a quick Internet search confirms that masks don't reduce respiratory viral transmission and 'the government knows that'; they have smashed civil society and they know that, too; two dozen peer-reviewed studies show no connection between lockdown and reducing deaths; he knew from personal friends the elite were still flying around and going on holiday while the public were locked down; the elite were not having the 'vaccines'. He was also asked if 'vaccines' could be made to target difference races. He said he didn't know, but the document by the Project for the New American Century in September, 2000, said developing 'advanced forms of biological warfare that can target *specific genotypes* may transform biological warfare from the realm of terror to a politically useful tool.' Oh, they're evil all right. Of that we can be *absolutely* sure.

Another cull of old people

We have seen from the CDC definition that the mRNA 'Covid vaccine' is not a vaccine and nor are the others that *claim* to reduce 'severity of symptoms' in *some* people, but not protect from infection or transmission. What about all the lies about returning to 'normal' if people were 'vaccinated'? If they are not claimed to stop infection and transmission of the alleged 'virus', how does anything change? This was all lies to manipulate people to take the jabs and we are seeing that now with masks and distancing still required for the 'vaccinated'. How did they think that elderly people with fragile health and immune responses were going to be affected by infusing their cells with synthetic material and other toxic substances? They *knew* that in the short and long term it would be devastating and fatal as the culling of the old that began with the first lockdowns was continued with the 'vaccine'. Death rates in care homes soared immediately residents began to be 'vaccinated' – infused with synthetic material. Brave and committed whistleblower nurses put their careers at risk by exposing this truth while the rest kept their heads down and their mouths shut to put their careers before those they are supposed to care for. A long-time American Certified Nursing Assistant who gave his name as James posted a video in which he described emotionally what happened in his care home when vaccination began. He said that during 2020 very few residents were sick with 'Covid' and no one died during the entire year; but shortly after the Pfizer mRNA injections 14 people died within two weeks and many others were near death. 'They're dropping like flies', he said. Residents who walked on their own before the shot could no longer and they had lost their ability to conduct an intelligent conversation. The home's management said the sudden deaths were caused by a 'super-spreader' of 'Covid-19'. Then how come, James asked, that residents who refused to take the injections were not sick? It was a case of inject the elderly with mRNA synthetic potions and blame their illness and death that followed on the 'virus'. James described what was happening in care homes as 'the greatest crime of genocide this country has ever seen'. Remember the NHS staff nurse from earlier who used the same word ‘genocide’ for what was happening with the ‘vaccines’ and that it was an ‘act of human annihilation’. A UK care home whistleblower told a similar story to James about the effect of the ‘vaccine’ in deaths and ‘outbreaks’ of illness dubbed ‘Covid’ after getting the jab. She told how her care home management and staff had zealously imposed government regulations and no one was allowed to even question the official narrative let alone speak out against it. She said the NHS was even worse. Again we see the results of reframing. A worker at a local care home where I live said they had not had a single case of ‘Covid’ there for almost a year and when the residents were ‘vaccinated’ they had 19 positive cases in two weeks with eight dying.

It’s not the ‘vaccine’ – honest

The obvious cause and effect was being ignored by the media and most of the public. Australia’s health minister Greg Hunt (a former head of strategy at the World Economic Forum) was admitted to hospital after he had the ‘vaccine’. He was suffering according to reports from the skin infection ‘cellulitis’ and it must have been a severe case to have warranted days in hospital. Immediately the authorities said this was nothing to do with the ‘vaccine’ when an effect of some vaccines is a ‘cellulitis-like reaction’. We had families of perfectly healthy old people who died after the ‘vaccine’ saying that if only they had been given the ‘vaccine’ earlier they would still be alive. As a numbskull rating that is off the chart. A father of four ‘died of Covid’ at aged 48 when he was taken ill two days after having the ‘vaccine’. The man, a health administrator, had been ‘shielding during the pandemic’ and had ‘not really left the house’ until he went for the ‘vaccine’. Having the ‘vaccine’ and then falling ill and dying does not seem to have qualified as a possible cause and effect and ‘Covid-19’ went on his death certificate. His family said they had no idea how he ‘caught the virus’. A family member said: ‘Tragically, it could be that going for a vaccination ultimately led to him catching Covid …The sad truth is that they are never going to know where it came from.’ The family warned people to remember that the virus still existed and was ‘very real’. So was their stupidity. Nurses and doctors who had the first round of the ‘vaccine’ were collapsing, dying and ending up in a hospital bed while they or their grieving relatives were saying they’d still have the ‘vaccine’ again despite what happened. I kid you not. You mean if your husband returned from the dead he’d have the same ‘vaccine’ again that killed him??

Doctors at the VCU Medical Center in Richmond, Virginia, said the Johnson & Johnson ‘vaccine’ was to blame for a man’s skin peeling off. Patient Richard Terrell said: ‘It all just happened so fast. My skin peeled off. It’s still coming off on my hands now.’ He said it was stinging, burning and itching and when he bent his arms and legs it was very painful with ‘the skin swollen and rubbing against itself’. Pfizer/BioNTech and Moderna vaccines use mRNA to change the cell while the Johnson & Johnson version uses DNA in a process similar to AstraZeneca’s technique. Johnson & Johnson and AstraZeneca have both had their ‘vaccines’ paused by many countries after causing serious blood problems. Terrell’s doctor Fnu Nutan said he could have died if he hadn’t got medical attention. It sounds terrible so what did Nutan and Terrell say about the ‘vaccine’ now? Oh, they still recommend that people have it. A nurse in a hospital bed 40 minutes after the vaccination and unable to swallow due to throat swelling was told by a doctor that he lost mobility in his arm for 36 hours following the vaccination. What did he say to the ailing nurse? ‘Good for you for getting the vaccination.’ We are dealing with a serious form of cognitive dissonance madness in both public and medical staff. There is a remarkable correlation between those having the ‘vaccine’ and trumpeting the fact and suffering bad happenings shortly afterwards. Witold Rogiewicz, a Polish doctor, made a video of his ‘vaccination’ and ridiculed those who were questioning its safety and the intentions of Bill Gates: ‘Vaccinate yourself to protect yourself, your loved ones, friends and also patients. And to mention quickly I have info for anti-vaxxers and anti-Coviders if you want to contact Bill Gates you can do this through me.’ He further ridiculed the dangers of 5G. Days later he was dead, but naturally the vaccination wasn't mentioned in the verdict of 'heart attack'.

Lies, lies and more lies

So many members of the human race have slipped into extreme states of insanity and unfortunately they include reframed doctors and nursing staff. Having a 'vaccine' and dying within minutes or hours is not considered a valid connection while death from any cause within 28 days or longer of a positive test with a test not testing for the 'virus' means 'Covid-19' goes on the death certificate. How could that 'vaccine'-death connection not have been made except by calculated deceit? US figures in the initial rollout period to February 12th, 2020, revealed that a third of the deaths reported to the CDC after 'Covid vaccines' happened within 48 hours. Five men in the UK suffered an 'extremely rare' blood clot problem after having the AstraZeneca 'vaccine', but no causal link was established said the Gates-funded Medicines and Healthcare products Regulatory Agency (MHRA) which had given the 'vaccine' emergency approval to be used. Former Pfizer executive Dr Michael Yeadon explained in his interview how the procedures could cause blood coagulation and clots. People who should have been at no risk were dying from blood clots in the brain and he said he had heard from medical doctor friends that people were suffering from skin bleeding and massive headaches. The AstraZeneca 'shot' was stopped by some 20 countries over the blood clotting issue and still the corrupt MHRA, the European Medicines Agency (EMA) and the World Health Organization said that it should continue to be given even though the EMA admitted that it 'still cannot rule out definitively' a link between blood clotting and the 'vaccine'. Later Marco Cavaleri, head of EMA vaccine strategy, said there was indeed a clear link between the 'vaccine' and thrombosis, but they didn't know why. So much for the trials showing the 'vaccine' is safe. Blood clots were affecting younger people who would be under virtually no danger from 'Covid' even if it existed which makes it all the more stupid and sinister.

The British government responded to public alarm by wheeling out June Raine, the terrifyingly weak infant school headmistress sound-alike who heads the UK MHRA drug ‘regulator’. The idea that she would stand up to Big Pharma and government pressure is laughable and she told us that all was well in the same way that she did when allowing untested, never-used-on-humans-before, genetically-manipulating ‘vaccines’ to be exposed to the public in the first place. Mass lying is the new normal of the ‘Covid’ era. The MHRA later said 30 cases of rare blood clots had by then been connected with the AstraZeneca ‘vaccine’ (that means a lot more in reality) while stressing that the benefits of the jab in preventing ‘Covid-19’ outweighed any risks. A more ridiculous and disingenuous statement with callous disregard for human health it is hard to contemplate. Immediately after the mendacious ‘all-clears’ two hospital workers in Denmark experienced blood clots and cerebral haemorrhaging following the AstraZeneca jab and one died. Top Norwegian health official Pål Andre Holme said the ‘vaccine’ was the only common factor: ‘There is nothing in the patient history of these individuals that can give such a powerful immune response … I am confident that the antibodies that we have found are the cause, and I see no other explanation than it being the vaccine which triggers it.’ Strokes, a clot or bleed in the brain, were clearly associated with the ‘vaccine’ from word of mouth and whistleblower reports. Similar consequences followed with all these ‘vaccines’ that we were told were so safe and as the numbers grew by the day it was clear we were witnessing human carnage.

Learning the hard way

A woman interviewed by UKColumn told how her husband suffered dramatic health effects after the vaccine when he’d been in good health all his life. He went from being a little unwell to losing all feeling in his legs and experiencing ‘excruciating pain’. Misdiagnosis followed twice at Accident and Emergency (an ‘allergy’ and ‘sciatica’) before he was admitted to a neurology ward where doctors said his serious condition had been caused by the ‘vaccine’. Another seven ‘vaccinated’ people were apparently being treated on the same ward for similar symptoms. The woman said he had the ‘vaccine’ because they believed media claims that it was safe. ‘I didn’t think the government would give out a vaccine that does this to somebody; I believed they would be bringing out a vaccination that would be safe.’ What a tragic way to learn that lesson. Another woman posted that her husband was transporting stroke patients to hospital on almost every shift and when he asked them if they had been ‘vaccinated’ for ‘Covid’ they all replied ‘yes’. One had a ‘massive brain bleed’ the day after his second dose. She said her husband reported the ‘just been vaccinated’ information every time to doctors in A and E only for them to ignore it, make no notes and appear annoyed that it was even mentioned. This particular report cannot be verified, but it expresses a common theme that confirms the monumental underreporting of ‘vaccine’ consequences. Interestingly as the ‘vaccines’ and their brain blood clot/stroke consequences began to emerge the UK National Health Service began a publicity campaign telling the public what to do in the event of a stroke. A Scottish NHS staff nurse who quit in disgust in March, 2021, said:

> I have seen traumatic injuries from the vaccine, they’re not getting reported to the yellow card [adverse reaction] scheme, they’re treating the symptoms, not asking why, why it’s happening. It’s just treating the symptoms and when you speak about it you’re dismissed like you’re crazy, I’m not crazy, I’m not crazy because every other colleague I’ve spoken to is terrified to speak out, they’ve had enough.

Videos appeared on the Internet of people uncontrollably shaking after the ‘vaccine’ with no control over muscles, limbs and even their face. A Scottish mother broke out in a severe rash all over her body almost immediately after she was given the AstraZeneca ‘vaccine’. The pictures were horrific. Leigh King, a 41-year-old hairdresser from Lanarkshire said: ‘Never in my life was I prepared for what I was about to experience … My skin was so sore and constantly hot … I have never felt pain like this …’ But don’t you worry, the ‘vaccine’ is perfectly safe. Then there has been the effect on medical staff who have been pressured to have the 'vaccine' by psychopathic 'health' authorities and government. A London hospital consultant who gave the name K. Polyakova wrote this to the *British Medical Journal* or *BMJ*:

> I am currently struggling with … the failure to report the reality of the morbidity caused by our current vaccination program within the health service and staff population. The levels of sickness after vaccination is unprecedented and staff are getting very sick and some with neurological symptoms which is having a huge impact on the health service function. Even the young and healthy are off for days, some for weeks, and some requiring medical treatment. Whole teams are being taken out as they went to get vaccinated together.
>
> Mandatory vaccination in this instance is stupid, unethical and irresponsible when it comes to protecting our staff and public health. We are in the voluntary phase of vaccination, and encouraging staff to take an unlicensed product that is impacting on their immediate health … it is clearly stated that these vaccine products do not offer immunity or stop transmission. In which case why are we doing it?

Not to protect health that's for sure. Medical workers are lauded by governments for agenda reasons when they couldn't give a toss about them any more than they can for the population in general. Schools across America faced the same situation as they closed due to the high number of teachers and other staff with bad reactions to the Pfizer/BioNTech, Moderna, and Johnson & Johnson 'Covid vaccines' all of which were linked to death and serious adverse effects. The *BMJ* took down the consultant's comments pretty quickly on the grounds that they were being used to spread 'disinformation'. They were exposing the truth about the 'vaccine' was the real reason. The cover-up is breathtaking.

Hiding the evidence

The scale of the 'vaccine' death cover-up worldwide can be confirmed by comparing official figures with the personal experience of the public. I heard of many people in my community who died immediately or soon after the vaccine that would never appear in the media or even likely on the official totals of 'vaccine' fatalities and adverse reactions when only about ten percent are estimated to be reported and I have seen some estimates as low as one percent in a Harvard study. In the UK alone by April 29th, 2021, some 757,654 adverse reactions had been officially reported from the Pfizer/BioNTech, Oxford/AstraZeneca and Moderna 'vaccines' with more than a thousand deaths linked to jabs and that means an estimated ten times this number in reality from a ten percent reporting rate percentage. That's seven million adverse reactions and 10,000 potential deaths and a one percent reporting rate would be ten times *those* figures. In 1976 the US government pulled the swine flu vaccine after *53* deaths. The UK data included a combined 10,000 eye disorders from the 'Covid vaccines' with more than 750 suffering visual impairment or blindness and again multiply by the estimated reporting percentages. As 'Covid cases' officially fell hospitals virtually empty during the 'Covid crisis' began to fill up with a range of other problems in the wake of the 'vaccine' rollout. The numbers across America have also been catastrophic. Deaths linked to *all* types of vaccine increased by *6,000 percent* in the first quarter of 2021 compared with 2020. A 39-year-old woman from Ogden, Utah, died four days after receiving a second dose of Moderna's 'Covid vaccine' when her liver, heart and kidneys all failed despite the fact that she had no known medical issues or conditions. Her family sought an autopsy, but Dr Erik Christensen, Utah's chief medical examiner, said proving vaccine injury as a cause of death almost never happened. He could think of only one instance where an autopsy would name a vaccine as the official cause of death and that would be anaphylaxis where someone received a vaccine and died almost instantaneously. 'Short of that, it would be difficult for us to definitively say this is the vaccine,' Christensen said. If that is true this must be added to the estimated ten percent (or far less) reporting rate of vaccine deaths and serious reactions and the conclusion can only be that vaccine deaths and serious reactions – including these 'Covid' potions' – are phenomenally understated in official figures. The same story can be found everywhere. Endless accounts of deaths and serious reactions among the public, medical and care home staff while official figures did not even begin to reflect this.

Professional script-reader Dr David Williams, a 'top public-health official' in Ontario, Canada, insulted our intelligence by claiming only four serious adverse reactions and no deaths from the more than 380,000 vaccine doses then given. This bore no resemblance to what people knew had happened in their owns circles and we had Dirk Huyer in charge of getting millions vaccinated in Ontario while at the same time he was Chief Coroner for the province investigating causes of death including possible death from the vaccine. An aide said he had stepped back from investigating deaths, but evidence indicated otherwise. Rosemary Frei, who secured a Master of Science degree in molecular biology at the Faculty of Medicine at Canada's University of Calgary before turning to investigative journalism, was one who could see that official figures for 'vaccine' deaths and reactions made no sense. She said that doctors seldom reported adverse events and when people got really sick or died after getting a vaccination they would attribute that to anything except the vaccines. It had been that way for years and anyone who wondered aloud whether the 'Covid vaccines' or other shots cause harm is immediately branded as 'anti-vax' and 'anti-science'. This was 'career-threatening' for health professionals. Then there was the huge pressure to support the push to 'vaccinate' billions in the quickest time possible. Frei said:

> So that's where we're at today. More than half a million vaccine doses have been given to people in Ontario alone. The rush is on to vaccinate all 15 million of us in the province by September. And the mainstream media are screaming for this to be sped up even more. That all adds up to only a very slim likelihood that we're going to be told the truth by officials about how many people are getting sick or dying from the vaccines.

What is true of Ontario is true of everywhere.

They KNEW – and still did it

The authorities knew what was going to happen with multiple deaths and adverse reactions. The UK government's Gates-funded and Big Pharma-dominated Medicines and Healthcare products Regulatory Agency (MHRA) hired a company to employ AI in compiling the projected reactions to the 'vaccine' that would otherwise be uncountable. The request for applications said: 'The MHRA urgently seeks an Artificial Intelligence (AI) software tool to process the expected high volume of Covid-19 vaccine Adverse Drug Reaction …' This was from the agency, headed by the disingenuous June Raine, that gave the 'vaccines' emergency approval and the company was hired before the first shot was given. 'We are going to kill and maim you – is that okay?' 'Oh, yes, perfectly fine – I'm very grateful, thank you, doctor.' The range of 'Covid vaccine' adverse reactions goes on for page after page in the MHRA criminally underreported 'Yellow Card' system and includes affects to eyes, ears, skin, digestion, blood and so on. Raine's MHRA amazingly claimed that the 'overall safety experience … is so far as expected from the clinical trials'. The death, serious adverse effects, deafness and blindness were *expected?* When did they ever mention that? If these human tragedies were expected then those that gave approval for the use of these 'vaccines' must be guilty of crimes against humanity including murder – a definition of which is 'killing a person with malice aforethought or with recklessness manifesting extreme indifference to the value of human life.' People involved at the MHRA, the CDC in America and their equivalent around the world must go before Nuremberg trials to answer for their callous inhumanity. We are only talking here about the immediate effects of the 'vaccine'. The longer-term impact of the DNA synthetic manipulation is the main reason they are so hysterically desperate to inoculate the entire global population in the shortest possible time.

Africa and the developing world are a major focus for the 'vaccine' depopulation agenda and a mass vaccination sales-pitch is underway thanks to caring people like the Rockefellers and other Cult assets. The Rockefeller Foundation, which pre-empted the 'Covid pandemic' in a document published in 2010 that 'predicted' what happened a decade later, announced an initial $34.95 million grant in February, 2021, 'to ensure more equitable access to Covid-19 testing and vaccines' among other things in Africa in collaboration with '24 organizations, businesses, and government agencies'. The pan-Africa initiative would focus on 10 countries: Burkina Faso, Ethiopia, Ghana, Kenya, Nigeria, Rwanda, South Africa, Tanzania, Uganda, and Zambia'. Rajiv Shah, President of the Rockefeller Foundation and former administrator of CIA-controlled USAID, said that if Africa was not mass-vaccinated (to change the DNA of its people) it was a 'threat to all of humanity' and not fair on Africans. When someone from the Rockefeller Foundation says they want to do something to help poor and deprived people and countries it is time for a belly-laugh. They are doing this out of the goodness of their 'heart' because 'vaccinating' the entire global population is what the 'Covid' hoax set out to achieve. Official 'decolonisation' of Africa by the Cult was merely a prelude to financial colonisation on the road to a return to physical colonisation. The 'vaccine' is vital to that and the sudden and convenient death of the 'Covid' sceptic president of Tanzania can be seen in its true light. A lot of people in Africa are aware that this is another form of colonisation and exploitation and they need to stand their ground.

The 'vaccine is working' scam

A potential problem for the Cult was that the 'vaccine' is meant to change human DNA and body messaging and not to protect anyone from a 'virus' never shown to exist. The vaccine couldn't work because it was not designed to work and how could they make it *appear* to be working so that more people would have it? This was overcome by lowering the amplification rate of the PCR test to produce fewer 'cases' and therefore fewer 'deaths'. Some of us had been pointing out since March, 2020, that the amplification rate of the test not testing for the 'virus' had been made artificially high to generate positive tests which they could call 'cases' to justify lockdowns. The World Health Organization recommended an absurdly high 45 amplification cycles to ensure the high positives required by the Cult and then remained silent on the issue until January 20th, 2021 – Biden's Inauguration Day. This was when the 'vaccinations' were seriously underway and on that day the WHO recommended after discussions with America's CDC that laboratories *lowered their testing amplification*. Dr David Samadi, a certified urologist and health writer, said the WHO was encouraging all labs to reduce their cycle count for PCR tests. He said the current cycle was much too high and was 'resulting in any particle being declared a positive case'. Even one mainstream news report I saw said this meant the number of 'Covid' infections may have been 'dramatically inflated'. Oh, just a little bit. The CDC in America issued new guidance to laboratories in April, 2021, to use 28 cycles *but only for 'vaccinated' people*. The timing of the CDC/WHO interventions were cynically designed to make it appear the 'vaccines' were responsible for falling cases and deaths when the real reason can be seen in the following examples. New York's state lab, the Wadsworth Center, identified 872 positive tests in July, 2020, based on a threshold of 40 cycles. When the figure was lowered to 35 cycles *43 percent* of the 872 were no longer 'positives'. At 30 cycles the figure was 63 percent. A Massachusetts lab found that between *85 to 90 percent* of people who tested positive in July with a cycle threshold of 40 would be negative at 30 cycles, Ashish Jha, MD, director of the Harvard Global Health Institute, said: 'I'm really shocked that it could be that high … Boy, does it really change the way we need to be thinking about testing.' I'm shocked that I could see the obvious in the spring of 2020, with no medical background, and most medical professionals still haven't worked it out. No, that's not shocking – it's terrifying.

Three weeks after the WHO directive to lower PCR cycles the London *Daily Mail* ran this headline: 'Why ARE Covid cases plummeting? New infections have fallen 45% in the US and 30% globally in the past 3 weeks but experts say vaccine is NOT the main driver because only 8% of Americans and 13% of people worldwide have received their first dose.' They acknowledged that the drop could not be attributed to the 'vaccine', but soon this morphed throughout the media into the 'vaccine' has caused cases and deaths to fall when it was the PCR threshold. In December, 2020, there was chaos at English Channel ports with truck drivers needing negative 'Covid' tests before they could board a ferry home for Christmas. The government wanted to remove the backlog as fast as possible and they brought in troops to do the 'testing'. Out of 1,600 drivers just *36* tested positive and the rest were given the all clear to cross the Channel. I guess the authorities thought that 36 was the least they could get away with without the unquestioning catching on. The amplification trick which most people believed in the absence of information in the mainstream applied more pressure on those refusing the 'vaccine' to succumb when it 'obviously worked'. The truth was the exact opposite with deaths in care homes soaring with the 'vaccine' and in Israel the term used was 'skyrocket'. A re-analysis of published data from the Israeli Health Ministry led by Dr Hervé Seligmann at the Medicine Emerging Infectious and Tropical Diseases at Aix-Marseille University found that Pfizer's 'Covid vaccine' killed 'about 40 times more [elderly] people than the disease itself would have killed' during a five-week vaccination period and *260 times* more younger people than would have died from the 'virus' even according to the manipulated 'virus' figures. Dr Seligmann and his co-study author, Haim Yativ, declared after reviewing the Israeli 'vaccine' death data: 'This is a new Holocaust.'

Then, in mid-April, 2021, after vast numbers of people worldwide had been 'vaccinated', the story changed with clear coordination. The UK government began to prepare the ground for more future lockdowns when Nuremberg-destined Boris Johnson told yet another whopper. He said that cases had fallen because of *lockdowns* not 'vaccines'. Lockdowns are irrelevant when *there is no 'virus'* and the test and fraudulent death certificates are deciding the number of 'cases' and 'deaths'. Study after study has shown that lockdowns don't work and instead kill and psychologically destroy people. Meanwhile in the United States Anthony Fauci and Rochelle Walensky, the ultra-Zionist head of the CDC, peddled the same line. More lockdown was the answer and not the 'vaccine', a line repeated on cue by the moron that is Canadian Prime Minister Justin Trudeau. Why all the hysteria to get everyone 'vaccinated' if lockdowns and not ‘vaccines’ made the difference? None of it makes sense on the face of it. Oh, but it does. The Cult wants lockdowns *and* the ‘vaccine’ and if the ‘vaccine’ is allowed to be seen as the total answer lockdowns would no longer be justified when there are still livelihoods to destroy. ‘Variants’ and renewed upward manipulation of PCR amplification are planned to instigate never-ending lockdown *and* more ‘vaccines’.

You *must* have it – we’re desperate

Israel, where the Jewish and Arab population are ruled by the Sabbatian Cult, was the front-runner in imposing the DNA-manipulating ‘vaccine’ on its people to such an extent that Jewish refusers began to liken what was happening to the early years of Nazi Germany. This would seem to be a fantastic claim. Why would a government of Jewish people be acting like the Nazis did? If you realise that the Sabbatian Cult was behind the Nazis and that Sabbatians hate Jews the pieces start to fit and the question of why a ‘Jewish’ government would treat Jews with such callous disregard for their lives and freedom finds an answer. Those controlling the government of Israel *aren’t Jewish* – they’re Sabbatian. Israeli lawyer Tamir Turgal was one who made the Nazi comparison in comments to German lawyer Reiner Fuellmich who is leading a class action lawsuit against the psychopaths for crimes against humanity. Turgal described how the Israeli government was vaccinating children and pregnant women on the basis that there was no evidence that this was dangerous when they had no evidence that it *wasn’t* dangerous either. They just had no evidence. This was medical experimentation and Turgal said this breached the Nuremberg Code about medical experimentation and procedures requiring informed consent and choice. Think about that. A Nuremberg Code developed because of Nazi experimentation on Jews and others in concentration camps by people like the evil-beyond-belief Josef Mengele is being breached by the *Israeli* government; but when you know that it’s a *Sabbatian* government along with its intelligence and military agencies like Mossad, Shin Bet and the Israeli Defense Forces, and that Sabbatians were the force behind the Nazis, the kaleidoscope comes into focus. What have we come to when Israeli Jews are suing their government for violating the Nuremberg Code by essentially making Israelis subject to a medical experiment using the controversial 'vaccines'? It's a shocker that this has to be done in the light of what happened in Nazi Germany. The Anshe Ha-Emet, or 'People of the Truth', made up of Israeli doctors, lawyers, campaigners and public, have launched a lawsuit with the International Criminal Court. It says:

> When the heads of the Ministry of Health as well as the prime minister presented the vaccine in Israel and began the vaccination of Israeli residents, the vaccinated were not advised, that, in practice, they are taking part in a medical experiment and that their consent is required for this under the Nuremberg Code.

The irony is unbelievable, but easily explained in one word: Sabbatians. The foundation of Israeli 'Covid' apartheid is the 'green pass' or 'green passport' which allows Jews and Arabs who have had the DNA-manipulating 'vaccine' to go about their lives – to work, fly, travel in general, go to shopping malls, bars, restaurants, hotels, concerts, gyms, swimming pools, theatres and sports venues, while non-'vaccinated' are banned from all those places and activities. Israelis have likened the 'green pass' to the yellow stars that Jews in Nazi Germany were forced to wear – the same as the yellow stickers that a branch of UK supermarket chain Morrisons told exempt mask-wears they had to display when shopping. How very sensitive. The Israeli system is blatant South African-style apartheid on the basis of compliance or non-compliance to fascism rather than colour of the skin. How appropriate that the Sabbatian Israeli government was so close to the pre-Mandela apartheid regime in Pretoria. The Sabbatian-instigated 'vaccine passport' in Israel is planned for everywhere. Sabbatians struck a deal with Pfizer that allowed them to lead the way in the percentage of a national population infused with synthetic material and the result was catastrophic. Israeli freedom activist Shai Dannon told me how chairs were appearing on beaches that said 'vaccinated only'. Health Minister Yuli Edelstein said that anyone unwilling or unable to get the jabs that 'confer immunity' will be 'left behind'. The man's a liar. Not even the makers claim the 'vaccines' confer immunity. When you see those figures of 'vaccine' deaths these psychopaths were saying that you must take the chance the 'vaccine' will kill you or maim you while knowing it will change your DNA or lockdown for you will be permanent. That's fascism. The Israeli parliament passed a law to allow personal information of the non-vaccinated to be shared with local and national authorities for three months. This was claimed by its supporters to be a way to 'encourage' people to be vaccinated. Hadas Ziv from Physicians for Human Rights described this as a 'draconian law which crushed medical ethics and the patient rights'. But that's the idea, the Sabbatians would reply.

Your papers, please

Sabbatian Israel was leading what has been planned all along to be a global 'vaccine pass' called a 'green passport' without which you would remain in permanent lockdown restriction and unable to do anything. This is how badly – *desperately* – the Cult is to get everyone 'vaccinated'. The term and colour 'green' was not by chance and related to the psychology of fusing the perception of the green climate hoax with the 'Covid' hoax and how the 'solution' to both is the same Great Reset. Lying politicians, health officials and psychologists denied there were any plans for mandatory vaccinations or restrictions based on vaccinations, but they knew that was exactly what was meant to happen with governments of all countries reaching agreements to enforce a global system. 'Free' Denmark and 'free' Sweden unveiled digital vaccine certification. Cyprus, Czech Republic, Estonia, Greece, Hungary, Iceland, Italy, Poland, Portugal, Slovakia, and Spain have all committed to a vaccine passport system and the rest including the whole of the EU would follow. The satanic UK government will certainly go this way despite mendacious denials and at the time of writing it is trying to manipulate the public into having the 'vaccine' so they could go abroad on a summer holiday. How would that work without something to prove you had the synthetic toxicity injected into you?

Documents show that the EU's European Commission was moving towards 'vaccine certificates' in 2018 and 2019 before the 'Covid' hoax began. They knew what was coming. Abracadabra – Ursula von der Leyen, the German President of the Commission, announced in March, 2021, an EU 'Digital Green Certificate' – green again – to track the public's 'Covid status'. The passport sting is worldwide and the Far East followed the same pattern with South Korea ruling that only those with 'vaccination' passports – again the *green* pass – would be able to 'return to their daily lives'.

Bill Gates has been preparing for this 'passport' with other Cult operatives for years and beyond the paper version is a Gates-funded 'digital tattoo' to identify who has been vaccinated and who hasn't. The 'tattoo' is reported to include a substance which is externally readable to confirm who has been vaccinated. This is a bio-luminous light-generating enzyme (think fireflies) called … *Luciferase*. Yes, named after the Cult 'god' Lucifer the 'light bringer' of whom more to come. Gates said he funded the readable tattoo to ensure children in the developing world were vaccinated and no one was missed out. He cares so much about poor kids as we know. This was just the cover story to develop a vaccine tagging system for everyone on the planet. Gates has been funding the ID2020 'alliance' to do just that in league with other lovely people at Microsoft, GAVI, the Rockefeller Foundation, Accenture and IDEO.org. He said in interviews in March, 2020, before any 'vaccine' publicly existed, that the world must have a globalised digital certificate to track the 'virus' and who had been vaccinated. Gates knew from the start that the mRNA vaccines were coming and when they would come and that the plan was to tag the 'vaccinated' to marginalise the intelligent and stop them doing anything including travel. Evil just doesn't suffice. Gates was exposed for offering a $10 million bribe to the Nigerian House of Representatives to invoke compulsory 'Covid' vaccination of all Nigerians. Sara Cunial, a member of the Italian Parliament, called Gates a 'vaccine criminal'. She urged the Italian President to hand him over to the International Criminal Court for crimes against humanity and condemned his plans to ‘chip the human race’ through ID2020.

You know it’s a long-planned agenda when war criminal and Cult gofer Tony Blair is on the case. With the scale of arrogance only someone as dark as Blair can muster he said: ‘Vaccination in the end is going to be your route to liberty.’ Blair is a disgusting piece of work and he confirms that again. The media has given a lot of coverage to a bloke called Charlie Mullins, founder of London’s biggest independent plumbing company, Pimlico Plumbers, who has said he won’t employ anyone who has not been vaccinated or have them go to any home where people are not vaccinated. He said that if he had his way no one would be allowed to walk the streets if they have not been vaccinated. Gates was cheering at the time while I was alerting the white coats. The plan is that people will qualify for ‘passports’ for having the first two doses and then to keep it they will have to have all the follow ups and new ones for invented ‘variants’ until human genetics is transformed and many are dead who can’t adjust to the changes. Hollywood celebrities – the usual propaganda stunt – are promoting something called the WELL Health-Safety Rating to verify that a building or space has ‘taken the necessary steps to prioritize the health and safety of their staff, visitors and other stakeholders’. They included Lady Gaga, Jennifer Lopez, Michael B. Jordan, Robert DeNiro, Venus Williams, Wolfgang Puck, Deepak Chopra and 17th Surgeon General Richard Carmona. Yawn. WELL Health-Safety has big connections with China. Parent company Delos is headed by former Goldman Sachs partner Paul Scialla. This is another example – and we will see so many others – of using the excuse of ‘health’ to dictate the lives and activities of the population. I guess one confirmation of the ‘safety’ of buildings is that only ‘vaccinated’ people can go in, right?

Electronic concentration camps

I wrote decades ago about the plans to restrict travel and here we are for those who refuse to bow to tyranny. This can be achieved in one go with air travel if the aviation industry makes a blanket decree.

The ‘vaccine’ and guaranteed income are designed to be part of a global version of China’s social credit system which tracks behaviour 24/7 and awards or deletes ‘credits’ based on whether your behaviour is supported by the state or not. I mean your entire lifestyle – what you do, eat, say, everything. Once your credit score falls below a certain level consequences kick in. In China tens of millions have been denied travel by air and train because of this. All the locations and activities denied to refusers by the ‘vaccine’ passports will be included in one big mass ban on doing almost anything for those that don’t bow their head to government. It’s beyond fascist and a new term is required to describe its extremes – I guess fascist technocracy will have to do. The way the Chinese system of technological – technocratic – control is sweeping the West can be seen in the Los Angeles school system and is planned to be expanded worldwide. Every child is required to have a ‘Covid’-tracking app scanned daily before they can enter the classroom. The so-called Daily Pass tracking system is produced by Gates’ Microsoft which I’m sure will shock you rigid. The pass will be scanned using a barcode (one step from an inside-the-body barcode) and the information will include health checks, ‘Covid’ tests and vaccinations. Entry codes are for one specific building only and access will only be allowed if a student or teacher has a negative test with a test not testing for the ‘virus’, has no symptoms of anything alleged to be related to ‘Covid’ (symptoms from a range of other illness), and has a temperature under 100 degrees. No barcode, no entry, is planned to be the case for everywhere and not only schools.

Kids are being psychologically prepared to accept this as ‘normal’ their whole life which is why what they can impose in schools is so important to the Cult and its gofers. Long-time American freedom campaigner John Whitehead of the Rutherford Institute was not exaggerating when he said: ‘Databit by databit, we are building our own electronic concentration camps.’ Canada under its Cult gofer prime minister Justin Trudeau has taken a major step towards the real thing with people interned against their will if they test positive with a test not testing for the ‘virus’ when they arrive at a Canadian airport. They are jailed in internment hotels often without food or water for long periods and with many doors failing to lock there have been sexual assaults. The interned are being charged sometimes $2,000 for the privilege of being abused in this way. Trudeau is fully on board with the Cult and says the 'Covid pandemic' has provided an opportunity for a global 'reset' to permanently change Western civilisation. His number two, Deputy Prime Minister Chrystia Freeland, is a trustee of the World Economic Forum and a Rhodes Scholar. The Trudeau family have long been servants of the Cult. See *The Biggest Secret* and Cathy O'Brien's book *Trance-Formation of America* for the horrific background to Trudeau's father Pierre Trudeau another Canadian prime minister. Hide your fascism behind the façade of a heart-on-the-sleeve liberal. It's a well-honed Cult technique.

What can the 'vaccine' *really* do?

We have a 'virus' never shown to exist and 'variants' of the 'virus' that have also never been shown to exist except, like the 'original', as computer-generated fictions. Even if you believe there's a 'virus' the 'case' to 'death' rate is in the region of 0.23 to 0.15 percent and those 'deaths' are concentrated among the very old around the same average age that people die anyway. In response to this lack of threat (in truth none) psychopaths and idiots, knowingly and unknowingly answering to Gates and the Cult, are seeking to 'vaccinate' every man, woman and child on Planet Earth. Clearly the 'vaccine' is not about 'Covid' – none of this ever has been. So what is it all about *really*? Why the desperation to infuse genetically-manipulating synthetic material into everyone through mRNA fraudulent 'vaccines' with the intent of doing this over and over with the excuses of 'variants' and other 'virus' inventions? Dr Sherri Tenpenny, an osteopathic medical doctor in the United States, has made herself an expert on vaccines and their effects as a vehement campaigner against their use. Tenpenny was board certified in emergency medicine, the director of a level two trauma centre for 12 years, and moved to Cleveland in 1996 to start an integrative medicine practice which has treated patients from all 50 states and some 17 other countries. Weaning people off pharmaceutical drugs is a speciality.

She became interested in the consequences of vaccines after attending a meeting at the National Vaccine Information Center in Washington DC in 2000 where she ‘sat through four days of listening to medical doctors and scientists and lawyers and parents of vaccine injured kids’ and asked: ‘What’s going on?’ She had never been vaccinated and never got ill while her father was given a list of vaccines to be in the military and was ‘sick his entire life’. The experience added to her questions and she began to examine vaccine documents from the Centers for Disease Control (CDC). After reading the first one, the 1998 version of *The General Recommendations of Vaccination*, she thought: ‘This is it?’ The document was poorly written and bad science and Tenpenny began 20 years of research into vaccines that continues to this day. She began her research into ‘Covid vaccines’ in March, 2020, and she describes them as ‘deadly’. For many, as we have seen, they already have been. Tenpenny said that in the first 30 days of the ‘vaccine’ rollout in the United States there had been more than 40,000 adverse events reported to the vaccine adverse event database. A document had been delivered to her the day before that was 172 pages long. ‘We have over 40,000 adverse events; we have over 3,100 cases of [potentially deadly] anaphylactic shock; we have over 5,000 neurological reactions.’ Effects ranged from headaches to numbness, dizziness and vertigo, to losing feeling in hands or feet and paraesthesia which is when limbs ‘fall asleep’ and people have the sensation of insects crawling underneath their skin. All this happened in the first 30 days and remember that only about *ten percent* (or far less) of adverse reactions and vaccine-related deaths are estimated to be officially reported. Tenpenny said:

> So can you think of one single product in any industry, any industry, for as long as products have been made on the planet that within 30 days we have 40,000 people complaining of side effects that not only is still on the market but … we’ve got paid actors telling us how great

> they are for getting their vaccine. We're offering people $500 if they will just get their vaccine and we've got nurses and doctors going; 'I got the vaccine, I got the vaccine'.

Tenpenny said they were not going to be 'happy dancing folks' when they began to suffer Bell's palsy (facial paralysis), neuropathies, cardiac arrhythmias and autoimmune reactions that kill through a blood disorder. 'They're not going to be so happy, happy then, but we're never going to see pictures of those people' she said. Tenpenny described the 'vaccine' as 'a well-designed killing tool'.

No off-switch

Bad as the initial consequences had been Tenpenny said it would be maybe 14 months before we began to see the 'full ravage' of what is going to happen to the 'Covid vaccinated' with full-out consequences taking anything between two years and 20 years to show. You can understand why when you consider that variations of the 'Covid vaccine' use mRNA (messenger RNA) to in theory activate the immune system to produce protective antibodies without using the actual 'virus'. How can they when it's a computer program and they've never isolated what they claim is the 'real thing'? Instead they use *synthetic* mRNA. They are inoculating synthetic material into the body which through a technique known as the Trojan horse is absorbed into cells to change the nature of DNA. Human DNA is changed by an infusion of messenger RNA and with each new 'vaccine' of this type it is changed even more. Say so and you are banned by Cult Internet platforms. The contempt the contemptuous Mark Zuckerberg has for the truth and human health can be seen in an internal Facebook video leaked to the Project Veritas investigative team in which he said of the 'Covid vaccines': '… I share some caution on this because we just don't know the long term side-effects of basically modifying people's DNA and RNA.' At the same time this disgusting man's Facebook was censoring and banning anyone saying exactly the same. He must go before a Nuremberg trial for crimes against humanity when he *knows* that he is censoring legitimate concerns and denying the right of informed consent on behalf of the Cult that owns him. People have been killed and damaged by the very 'vaccination' technique he cast doubt on himself when they may not have had the 'vaccine' with access to information that he denied them. The plan is to have at least annual 'Covid vaccinations', add others to deal with invented 'variants', and change all other vaccines into the mRNA system. Pfizer executives told shareholders at a virtual Barclays Global Healthcare Conference in March, 2021, that the public may need a third dose of 'Covid vaccine', plus regular yearly boosters and the company planned to hike prices to milk the profits in a 'significant opportunity for our vaccine'. These are the professional liars, cheats and opportunists who are telling you their 'vaccine' is safe. Given this volume of mRNA planned to be infused into the human body and its ability to then replicate we will have a transformation of human genetics from biological to synthetic biological – exactly the long-time Cult plan for reasons we'll see – and many will die. Sherri Tenpenny said of this replication:

> It's like having an on-button but no off-button and that whole mechanism … they actually give it a name and they call it the Trojan horse mechanism, because it allows that [synthetic] virus and that piece of that [synthetic] virus to get inside of your cells, start to replicate and even get inserted into other parts of your DNA as a Trojan-horse.

Ask the overwhelming majority of people who have the 'vaccine' what they know about the contents and what they do and they would reply: 'The government says it will stop me getting the virus.' Governments give that false impression on purpose to increase take-up. You can read Sherri Tenpenny's detailed analysis of the health consequences in her blog at Vaxxter.com, but in summary these are some of them. She highlights the statement by Bill Gates about how human beings can become their own 'vaccine manufacturing machine'. The man is insane. ['Vaccine'-generated] 'antibodies' carry synthetic messenger RNA into the cells and the damage starts, Tenpenny contends, and she says that lungs can be adversely affected through varying degrees of pus and bleeding which obviously affects breathing and would be dubbed ‘Covid-19’. Even more sinister was the impact of ‘antibodies’ on macrophages, a white blood cell of the immune system. They consist of Type 1 and Type 2 which have very different functions. She said Type 1 are ‘hyper-vigilant’ white blood cells which ‘gobble up’ bacteria etc. However, in doing so, this could cause inflammation and in extreme circumstances be fatal. She says these affects are mitigated by Type 2 macrophages which kick in to calm down the system and stop it going rogue. They clear up dead tissue debris and reduce inflammation that the Type 1 ‘fire crews’ have caused. Type 1 kills the infection and Type 2 heals the damage, she says. This is her punchline with regard to ‘Covid vaccinations’: She says that mRNA ‘antibodies’ block Type 2 macrophages by attaching to them and deactivating them. This meant that when the Type 1 response was triggered by infection there was nothing to stop that getting out of hand by calming everything down. There’s an on-switch, but no off-switch, she says. What follows can be ‘over and out, see you when I see you’.

Genetic suicide

Tenpenny also highlights the potential for autoimmune disease – the body attacking itself – which has been associated with vaccines since they first appeared. Infusing a synthetic foreign substance into cells could cause the immune system to react in a panic believing that the body is being overwhelmed by an invader (it is) and the consequences can again be fatal. There is an autoimmune response known as a ‘cytokine storm’ which I have likened to a homeowner panicked by an intruder and picking up a gun to shoot randomly in all directions before turning the fire on himself. The immune system unleashes a storm of inflammatory response called cytokines to a threat and the body commits hara-kiri. The lesson is that you mess with the body’s immune response at your peril and these ‘vaccines’ seriously – fundamentally – mess with immune response. Tenpenny refers to a consequence called anaphylactic shock which is a severe and highly dangerous allergic reaction when the immune system floods the body with chemicals. She gives the example of having a bee sting which primes the immune system and makes it sensitive to those chemicals. When people are stung again maybe years later the immune response can be so powerful that it leads to anaphylactic shock. Tenpenny relates this 'shock' with regard to the 'Covid vaccine' to something called polyethylene glycol or PEG. Enormous numbers of people have become sensitive to this over decades of use in a whole range of products and processes including food, drink, skin creams and 'medicine'. Studies have claimed that some 72 percent of people have antibodies triggered by PEG compared with two percent in the 1960s and allergic hypersensitive reactions to this become a gathering cause for concern. Tenpenny points out that the 'mRNA vaccine' is coated in a 'bubble' of polyethylene glycol which has the potential to cause anaphylactic shock through immune sensitivity. Many reports have appeared of people reacting this way after having the 'Covid vaccine'. What do we think is going to happen as humanity has more and more of these 'vaccines'? Tenpenny said: 'All these pictures we have seen with people with these rashes … these weepy rashes, big reactions on their arms and things like that – it's an acute allergic reaction most likely to the polyethylene glycol that you've been previously primed and sensitised to.'

Those who have not studied the conspiracy and its perpetrators at length might think that making the population sensitive to PEG and then putting it in these 'vaccines' is just a coincidence. It is not. It is instead testament to how carefully and coldly-planned current events have been and the scale of the conspiracy we are dealing with. Tenpenny further explains that the 'vaccine' mRNA procedure can breach the blood-brain barrier which protects the brain from toxins and other crap that will cause malfunction. In this case they could make two proteins corrupt brain function to cause Amyotrophic lateral sclerosis (ALS) , a progressive nervous system disease leading to loss of muscle control, and frontal lobe degeneration – Alzheimer's and dementia. Immunologist J. Bart Classon published a paper connecting mRNA 'vaccines' to prion disease which can lead to Alzheimer's and other forms of neurogenerative disease while others have pointed out the potential to affect the placenta in ways that make women infertile. This will become highly significant in the next chapter when I will discuss other aspects of this non-vaccine that relate to its nanotechnology and transmission from the injected to the uninjected.

Qualified in idiocy

Tenpenny describes how research has confirmed that these 'vaccine'-generated antibodies can interact with a range of other tissues in the body and attack many other organs including the lungs. 'This means that if you have a hundred people standing in front of you that all got this shot they could have a hundred different symptoms.' Anyone really think that Cult gofers like the Queen, Tony Blair, Christopher Whitty, Anthony Fauci, and all the other psychopaths have really had this 'vaccine' in the pictures we've seen? Not a bloody chance. Why don't doctors all tell us about all these dangers and consequences of the 'Covid vaccine'? Why instead do they encourage and pressure patients to have the shot? Don't let's think for a moment that doctors and medical staff can't be stupid, lazy, and psychopathic and that's without the financial incentives to give the jab. Tenpenny again:

> Some people are going to die from the vaccine directly but a large number of people are going to start to get horribly sick and get all kinds of autoimmune diseases 42 days to maybe a year out. What are they going to do, these stupid doctors who say; 'Good for you for getting that vaccine.' What are they going to say; 'Oh, it must be a mutant, we need to give an extra dose of that vaccine.'
>
> Because now the vaccine, instead of one dose or two doses we need three or four because the stupid physicians aren't taking the time to learn anything about it. If I can learn this sitting in my living room reading a 19 page paper and several others so can they. There's nothing special about me, I just take the time to do it.

Remember how Sara Kayat, the NHS and TV doctor, said that the 'Covid vaccine' would '100 percent prevent hospitalisation and death'. Doctors can be idiots like every other profession and they should not be worshipped as infallible. They are not and far from it. Behind many medical and scientific 'experts' lies an uninformed prat trying to hide themselves from you although in the 'Covid' era many have failed to do so as with UK narrative-repeating 'TV doctor' Hilary Jones. Pushing back against the minority of proper doctors and scientists speaking out against the 'vaccine' has been the entire edifice of the Cult global state in the form of governments, medical systems, corporations, mainstream media, Silicon Valley, and an army of compliant doctors, medical staff and scientists willing to say anything for money and to enhance their careers by promoting the party line. If you do that you are an 'expert' and if you won't you are an 'anti-vaxxer' and 'Covidiot'. The pressure to be 'vaccinated' is incessant. We have even had reports claiming that the 'vaccine' can help cure cancer and Alzheimer's and make the lame walk. I am waiting for the announcement that it can bring you coffee in the morning and cook your tea. Just as the symptoms of 'Covid' seem to increase by the week so have the miracles of the 'vaccine'. American supermarket giant Kroger Co. offered nearly 500,000 employees in 35 states a $100 bonus for having the 'vaccine' while donut chain Krispy Kreme promised 'vaccinated' customers a free glazed donut every day for the rest of 2021. Have your DNA changed and you will get a doughnut although we might not have to give you them for long. Such offers and incentives confirm the desperation.

Perhaps the worse vaccine-stunt of them all was UK 'Health' Secretary Matt-the-prat Hancock on live TV after watching a clip of someone being 'vaccinated' when the roll-out began. Hancock faked tears so badly it was embarrassing. Brain-of-Britain Piers Morgan, the lockdown-supporting, 'vaccine' supporting, 'vaccine' passport-supporting, TV host played along with Hancock – 'You're quite emotional about that' he said in response to acting so atrocious it would have been called out at a school nativity which will presumably today include Mary and Jesus in masks, wise men keeping their camels six feet apart, and shepherds under tent arrest. System-serving Morgan tweeted this: 'Love the idea of covid vaccine passports for everywhere: flights, restaurants, clubs, football, gyms, shops etc. It’s time covid-denying, anti-vaxxer loonies had their bullsh*t bluff called & bar themselves from going anywhere that responsible citizens go.’ If only I could aspire to his genius. To think that Morgan, who specialises in shouting over anyone he disagrees with, was lauded as a free speech hero when he lost his job after storming off the set of his live show like a child throwing his dolly out of the pram. If he is a free speech hero we are in real trouble. I have no idea what ‘bullsh*t’ means, by the way, the * throws me completely.

The Cult is desperate to infuse its synthetic DNA-changing concoction into everyone and has been using every lie, trick and intimidation to do so. The question of ‘*Why*?’ we shall now address.

CHAPTER TEN

Human 2.0

***I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted* –**
Alan Turing (1912-1954), the 'Father of artificial intelligence'

I have been exposing for decades the plan to transform the human body from a biological to a synthetic-biological state. The new human that I will call Human 2.0 is planned to be connected to artificial intelligence and a global AI 'Smart Grid' that would operate as one global system in which AI would control everything from your fridge to your heating system to your car to your mind. Humans would no longer be 'human', but post-human and sub-human, with their thinking and emotional processes replaced by AI.

What I said sounded crazy and beyond science fiction and I could understand that. To any balanced, rational, mind it *is* crazy. Today, however, that world is becoming reality and it puts the 'Covid vaccine' into its true context. Ray Kurzweil is the ultra-Zionist 'computer scientist, inventor and futurist' and co-founder of the Singularity University. Singularity refers to the merging of humans with machines or 'transhumanism'. Kurzweil has said humanity would be connected to the cyber 'cloud' in the period of the ever-recurring year of 2030:

> Our thinking … will be a hybrid of biological and non-biological thinking … humans will be able to extend their limitations and 'think in the cloud' … We're going to put gateways to the

> cloud in our brains ... We're going to gradually merge and enhance ourselves ... In my view, that's the nature of being human – we transcend our limitations. As the technology becomes vastly superior to what we are then the small proportion that is still human gets smaller and smaller and smaller until it's just utterly negligible.

They are trying to sell this end-of-humanity-as-we-know-it as the next stage of 'evolution' when we become super-human and 'like the gods'. They are lying to you. Shocked, eh? The population, and again especially the young, have been manipulated into addiction to technologies designed to enslave them for life. First they induced an addiction to smartphones (holdables); next they moved to technology on the body (wearables); and then began the invasion of the body (implantables). I warned way back about the plan for microchipped people and we are now entering that era. We should not be diverted into thinking that this refers only to chips we can see. Most important are the nanochips known as smart dust, neural dust and nanobots which are far too small to be seen by the human eye. Nanotechnology is everywhere, increasingly in food products, and released into the atmosphere by the geoengineering of the skies funded by Bill Gates to 'shut out the Sun' and 'save the planet from global warming'. Gates has been funding a project to spray millions of tonnes of chalk (calcium carbonate) into the stratosphere over Sweden to 'dim the Sun' and cool the Earth. Scientists warned the move could be disastrous for weather systems in ways no one can predict and opposition led to the Swedish space agency announcing that the 'experiment' would not be happening as planned in the summer of 2021; but it shows where the Cult is going with dimming the impact of the Sun and there's an associated plan to change the planet's atmosphere. Who gives psychopath Gates the right to dictate to the entire human race and dismantle planetary systems? The world will not be safe while this man is at large.

The global warming hoax has made the Sun, like the gas of life, something to fear when both are essential to good health and human survival (more inversion). The body transforms sunlight into vital vitamin D through a process involving … *cholesterol*. This is the cholesterol we are also told to fear. We are urged to take Big Pharma statin drugs to reduce cholesterol and it's all systematic. Reducing cholesterol means reducing vitamin D uptake with all the multiple health problems that will cause. At least if you take statins long term it saves the government from having to pay you a pension. The delivery system to block sunlight is widely referred to as chemtrails although these have a much deeper agenda, too. They appear at first to be contrails or condensation trails streaming from aircraft into cold air at high altitudes. Contrails disperse very quickly while chemtrails do not and spread out across the sky before eventually their content falls to earth. Many times I have watched aircraft cross-cross a clear blue sky releasing chemtrails until it looks like a cloudy day. Chemtrails contain many things harmful to humans and the natural world including toxic heavy metals, aluminium (see Alzheimer's) and nanotechnology. Ray Kurzweil reveals the reason without actually saying so: 'Nanobots will infuse all the matter around us with information. Rocks, trees, everything will become these intelligent creatures.' How do you deliver that? *From the sky*. Self-replicating nanobots would connect everything to the Smart Grid. The phenomenon of Morgellons disease began in the chemtrail era and the correlation has led to it being dubbed the 'chemtrail disease'. Self-replicating fibres appear in the body that can be pulled out through the skin. Morgellons fibres continue to grow outside the body and have a form of artificial intelligence. I cover this at greater length in *Phantom Self*.

'Vaccine' operating system

'Covid vaccines' with their self-replicating synthetic material are also designed to make the connection between humanity and Kurzweil's 'cloud'. American doctor and dedicated campaigner for truth, Carrie Madej, an Internal Medicine Specialist in Georgia with more than 20 years medical experience, has highlighted the nanotechnology aspect of the fake 'vaccines'. She explains how one of the components in at least the Moderna and Pfizer synthetic potions are 'lipid nanoparticles' which are 'like little tiny computer bits' – a 'sci-fi substance' known as nanobots and hydrogel which can be 'triggered at any moment to deliver its payload' and act as 'biosensors'. The synthetic substance had 'the ability to accumulate data from your body like your breathing, your respiration, thoughts and emotions, all kind of things' and each syringe could carry a *million* nanobots:

> This substance because it's like little bits of computers in your body, crazy, but it's true, it can do that, [and] obviously has the ability to act through Wi-Fi. It can receive and transmit energy, messages, frequencies or impulses. That issue has never been addressed by these companies. What does that do to the human?
>
> Just imagine getting this substance in you and it can react to things all around you, the 5G, your smart device, your phones, what is happening with that? What if something is triggering it, too, like an impulse, a frequency? We have something completely foreign in the human body.

Madej said her research revealed that electromagnetic (EMF) frequencies emitted by phones and other devices had increased dramatically in the same period of the 'vaccine' rollout and she was seeing more people with radiation problems as 5G and other electromagnetic technology was expanded and introduced to schools and hospitals. She said she was 'floored with the EMF coming off' the devices she checked. All this makes total sense and syncs with my own work of decades when you think that Moderna refers in documents to its mRNA 'vaccine' as an 'operating system':

> Recognizing the broad potential of mRNA science, we set out to create an mRNA technology platform that functions very much like an operating system on a computer. It is designed so that it can plug and play interchangeably with different programs. In our case, the 'program' or 'app' is our mRNA drug – the unique mRNA sequence that codes for a protein …
>
> … Our MRNA Medicines – 'The 'Software Of Life': When we have a concept for a new mRNA medicine and begin research, fundamental components are already in place. Generally, the only thing that changes from one potential mRNA medicine to another is the coding region – the actual genetic code that instructs ribosomes to make protein. Utilizing these instruction sets gives our investigational mRNA medicines a software-like quality. We also have the ability to combine different mRNA sequences encoding for different proteins in a single mRNA investigational medicine.

Who needs a real 'virus' when you can create a computer version to justify infusing your operating system into the entire human race on the road to making living, breathing people into cyborgs? What is missed with the 'vaccines' is the *digital* connection between synthetic material and the body that I highlighted earlier with the study that hacked a computer with human DNA. On one level the body is digital, based on mathematical codes, and I'll have more about that in the next chapter. Those who ridiculously claim that mRNA 'vaccines' are not designed to change human genetics should explain the words of Dr Tal Zaks, chief medical officer at Moderna, in a 2017 TED talk. He said that over the last 30 years 'we've been living this phenomenal digital scientific revolution, and I'm here today to tell you, that we are actually *hacking the software of life*, and that it's changing the way we think about prevention and treatment of disease':

> In every cell there's this thing called messenger RNA, or mRNA for short, that transmits the critical information from the DNA in our genes to the protein, which is really the stuff we're all made out of. This is the critical information that determines what the cell will do. So we think about it as an operating system. So if you could change that, if you could introduce a line of code, or change a line of code, it turns out, that has profound implications for everything, from the flu to cancer.

Zaks should more accurately have said that this has profound implications for the human genetic code and the nature of DNA. Communications within the body go both ways and not only one. But, hey, no, the 'Covid vaccine' will not affect your genetics. Cult fact-checkers say so even though the man who helped to develop the mRNA technique says that it does. Zaks said in 2017:

> If you think about what it is we're trying to do. We've taken information and our understanding of that information and how that information is transmitted in a cell, and we've taken our understanding of medicine and how to make drugs, and we're fusing the two. We think of it as information therapy.

I have been writing for decades that the body is an information field communicating with itself and the wider world. This is why radiation which is information can change the information field of body and mind through phenomena like 5G and change their nature and function. ‘Information therapy’ means to change the body’s information field and change the way it operates. DNA is a receiver-transmitter of information and can be mutated by information like mRNA synthetic messaging. Technology to do this has been ready and waiting in the underground bases and other secret projects to be rolled out when the ‘Covid’ hoax was played. ‘Trials’ of such short and irrelevant duration were only for public consumption. When they say the ‘vaccine’ is ‘experimental’ that is not true. It may appear to be ‘experimental’ to those who don’t know what’s going on, but the trials have already been done to ensure the Cult gets the result it desires. Zaks said that it took decades to sequence the human genome, completed in 2003, but now they could do it in a week. By ‘they’ he means scientists operating in the public domain. In the secret projects they were sequencing the genome in a week long before even 2003.

Deluge of mRNA

Highly significantly the Moderna document says the guiding premise is that if using mRNA as a medicine works for one disease then it should work for many diseases. They were leveraging the flexibility afforded by their platform and the fundamental role mRNA plays in protein synthesis to pursue mRNA medicines for a broad spectrum of diseases. Moderna is confirming what I was saying through 2020 that multiple ‘vaccines’ were planned for ‘Covid’ (and later invented ‘variants’) and that previous vaccines would be converted to the mRNA system to infuse the body with massive amounts of genetically-manipulating synthetic material to secure a transformation to a synthetic-biological state. The ‘vaccines’ are designed to kill stunning numbers as part of the long-exposed Cult depopulation agenda and transform the rest. Given this is the goal you can appreciate why there is such hysterical demand for every human to be ‘vaccinated’ for an alleged ‘disease’ that has an estimated ‘infection’ to ‘death’ ratio of 0.23-0.15 percent. As I write children are being given the ‘vaccine’ in trials (their parents are a disgrace) and ever-younger people are being offered the vaccine for a ‘virus’ that even if you believe it exists has virtually zero chance of harming them. Horrific effects of the ‘trials’ on a 12-year-old girl were revealed by a family member to be serious brain and gastric problems that included a bowel obstruction and the inability to swallow liquids or solids. She was unable to eat or drink without throwing up, had extreme pain in her back, neck and abdomen, and was paralysed from the waist down which stopped her urinating unaided. When the girl was first taken to hospital doctors said it was all in her mind. She was signed up for the ‘trial’ by her parents for whom no words suffice. None of this ‘Covid vaccine’ insanity makes any sense unless you see what the ‘vaccine’ really is – a body-changer. Synthetic biology or ‘SynBio’ is a fast-emerging and expanding scientific discipline which includes everything from genetic and molecular engineering to electrical and computer engineering. Synthetic biology is defined in these ways:

- A multidisciplinary area of research that seeks to create new biological parts, devices, and systems, or to redesign systems that are already found in nature.

- The use of a mixture of physical engineering and genetic engineering to create new (and therefore synthetic) life forms.

- An emerging field of research that aims to combine the knowledge and methods of biology, engineering and related disciplines in the design of chemically-synthesized DNA to create organisms with novel or enhanced characteristics and traits (synthetic organisms including humans).

We now have synthetic blood, skin, organs and limbs being developed along with synthetic body parts produced by 3D printers. These are all elements of the synthetic human programme and this comment by Kurzweil’s co-founder of the Singularity University, Peter Diamandis, can be seen in a whole new light with the 'Covid' hoax and the sanctions against those that refuse the 'vaccine':

> Anybody who is going to be resisting the progress forward [to transhumanism] is going to be resisting evolution and, fundamentally, they will die out. It's not a matter of whether it's good or bad. It's going to happen.

'Resisting evolution'? What absolute bollocks. The arrogance of these people is without limit. His 'it's going to happen' mantra is another way of saying 'resistance is futile' to break the spirit of those pushing back and we must not fall for it. Getting this genetically-transforming 'vaccine' into everyone is crucial to the Cult plan for total control and the desperation to achieve that is clear for anyone to see. Vaccine passports are a major factor in this and they, too, are a form of resistance is futile. It's NOT. The paper funded by the Rockefeller Foundation for the 2013 'health conference' in China said:

> We will interact more with artificial intelligence. The use of robotics, bio-engineering to augment human functioning is already well underway and will advance. Re-engineering of humans into potentially separate and unequal forms through genetic engineering or mixed human-robots raises debates on ethics and equality.
>
> A new demography is projected to emerge after 2030 [that year again] of technologies (robotics, genetic engineering, nanotechnology) producing robots, engineered organisms, 'nanobots' and artificial intelligence (AI) that can self-replicate. Debates will grow on the implications of an impending reality of human designed life.

What is happening today is so long planned. The world army enforcing the will of the world government is intended to be a robot army, not a human one. Today's military and its technologically 'enhanced' troops, pilotless planes and driverless vehicles are just stepping stones to that end. Human soldiers are used as Cult fodder and its time they woke up to that and worked for the freedom of the population instead of their own destruction and their family's destruction – the same with the police. Join us and let's sort this out. The phenomenon of enforce my own destruction is widespread in the 'Covid' era with Woker 'luvvies' in the acting and entertainment industries supporting ‘Covid’ rules which have destroyed their profession and the same with those among the public who put signs on the doors of their businesses ‘closed due to Covid – stay safe’ when many will never reopen. It’s a form of masochism and most certainly insanity.

Transgender = transhumanism

When something explodes out of nowhere and is suddenly everywhere it is always the Cult agenda and so it is with the tidal wave of claims and demands that have infiltrated every aspect of society under the heading of ‘transgenderism’. The term ‘trans’ is so ‘in’ and this is the dictionary definition:

> A prefix meaning ‘across’, ’through’, occurring … in loanwords from Latin, used in particular for denoting movement or conveyance from place to place (transfer; transmit; transplant) or complete change (transform; transmute), or to form adjectives meaning ’crossing’, ‘on the other side of’, or ‘going beyond’ the place named (transmontane; transnational; trans-Siberian).

Transgender means to go beyond gender and transhuman means to go beyond human. Both are aspects of the Cult plan to transform the human body to a synthetic state with *no gender*. Human 2.0 is not designed to procreate and would be produced technologically with no need for parents. The new human would mean the end of parents and so men, and increasingly women, are being targeted for the deletion of their rights and status. Parental rights are disappearing at an ever-quickening speed for the same reason. The new human would have no need for men or women when there is no procreation and no gender. Perhaps the transgender movement that appears to be in a permanent state of frenzy might now contemplate on how it is being used. This was never about transgender rights which are only the interim excuse for confusing gender, particularly in the young, on the road to *fusing* gender. Transgender activism is not an end; it is a *means* to an end. We see again the technique of creative destruction in which you destroy the status quo to ‘build back better’ in the form that you want. The gender status quo had to be destroyed by persuading the Cult-created Woke mentality to believe that you can have 100 genders or more. A programme for 9 to 12 year olds produced by the Cult-owned BBC promoted the 100 genders narrative. The very idea may be the most monumental nonsense, but it is not what is true that counts, only what you can make people *believe* is true. Once the gender of 2 + 2 = 4 has been dismantled through indoctrination, intimidation and 2 + 2 = 5 then the new no-gender normal can take its place with Human 2.0. Aldous Huxley revealed the plan in his prophetic *Brave New World* in 1932:

> Natural reproduction has been done away with and children are created, decanted', and raised in 'hatcheries and conditioning centres'. From birth, people are genetically designed to fit into one of five castes, which are further split into 'Plus' and 'Minus' members and designed to fulfil predetermined positions within the social and economic strata of the World State.

How could Huxley know this in 1932? For the same reason George Orwell knew about the Big Brother state in 1948, Cult insiders I have quoted knew about it in 1969, and I have known about it since the early 1990s. If you are connected to the Cult or you work your balls off to uncover the plan you can predict the future. The process is simple. If there is a plan for the world and nothing intervenes to stop it then it will happen. Thus if you communicate the plan ahead of time you are perceived to have predicted the future, but you haven't. You have revealed the plan which without intervention will become the human future. The whole reason I have done what I have is to alert enough people to inspire an intervention and maybe at last that time has come with the Cult and its intentions now so obvious to anyone with a brain in working order.

The future is here

Technological wombs that Huxley described to replace parent procreation are already being developed and they are only the projects we know about in the public arena. Israeli scientists told *The Times of Israel* in March, 2021, that they have grown 250-cell embryos into mouse foetuses with fully formed organs using artificial wombs in a development they say could pave the way for gestating humans outside the womb. Professor Jacob Hanna of the Weizmann Institute of Science said:

> We took mouse embryos from the mother at day five of development, when they are just of 250 cells, and had them in the incubator from day five until day 11, by which point they had grown all their organs.
>
> By day 11 they make their own blood and have a beating heart, a fully developed brain. Anybody would look at them and say, 'this is clearly a mouse foetus with all the characteristics of a mouse.' It's gone from being a ball of cells to being an advanced foetus.

A special liquid is used to nourish embryo cells in a laboratory dish and they float on the liquid to duplicate the first stage of embryonic development. The incubator creates all the right conditions for its development, Hanna said. The liquid gives the embryo 'all the nutrients, hormones and sugars they need' along with a custom-made electronic incubator which controls gas concentration, pressure and temperature. The cutting-edge in the underground bases and other secret locations will be light years ahead of that, however, and this was reported by the London *Guardian* in 2017:

> We are approaching a biotechnological breakthrough. Ectogenesis, the invention of a complete external womb, could completely change the nature of human reproduction. In April this year, researchers at the Children's Hospital of Philadelphia announced their development of an artificial womb.

The article was headed 'Artificial wombs could soon be a reality. What will this mean for women?' What would it mean for children is an even bigger question. No mother to bond with only a machine in preparation for a life of soulless interaction and control in a world governed by machines (see the *Matrix* movies). Now observe the calculated manipulations of the 'Covid' hoax as human interaction and warmth has been curtailed by distancing, isolation and fear with people communicating via machines on a scale never seen before.

These are all dots in the same picture as are all the personal assistants, gadgets and children's toys through which kids and adults communicate with AI as if it is human. The AI 'voice' on Sat-Nav should be included. All these things are psychological preparation for the Cult endgame. Before you can make a physical connection with AI you have to make a psychological connection and that is what people are being conditioned to do with this ever gathering human-AI interaction. Movies and TV programmes depicting the transhuman, robot dystopia relate to a phenomenon known as 'pre-emptive programming' in which the world that is planned is portrayed everywhere in movies, TV and advertising. This is conditioning the conscious and subconscious mind to become familiar with the planned reality to dilute resistance when it happens for real. What would have been a shock such is the change is made less so. We have young children put on the road to transgender transition surgery with puberty blocking drugs at an age when they could never be able to make those life-changing decisions.

Rachel Levine, a professor of paediatrics and psychiatry who believes in treating children this way, became America's highest-ranked openly-transgender official when she was confirmed as US Assistant Secretary at the Department of Health and Human Services after being nominated by Joe Biden (the Cult). Activists and governments press for laws to deny parents a say in their children's transition process so the kids can be isolated and manipulated into agreeing to irreversible medical procedures. A Canadian father Robert Hoogland was denied bail by the Vancouver Supreme Court in 2021 and remained in jail for breaching a court order that he stay silent over his young teenage daughter, a minor, who was being offered life-changing hormone therapy without parental consent. At the age of 12 the girl's 'school counsellor' said she may be transgender, referred her to a doctor and told the school to treat her like a boy. This is another example of state-serving schools imposing ever more control over children's lives while parents have ever less.

Contemptible and extreme child abuse is happening all over the world as the Cult gender-fusion operation goes into warp-speed.

Why the war on men – and now women?

The question about what artificial wombs mean for women should rightly be asked. The answer can be seen in the deletion of women's rights involving sport, changing rooms, toilets and status in favour of people in male bodies claiming to identify as women. I can identify as a mountain climber, but it doesn't mean I can climb a mountain any more than a biological man can be a biological woman. To believe so is a triumph of belief over factual reality which is the very perceptual basis of everything Woke. Women's sport is being destroyed by allowing those with male bodies who say they identify as female to 'compete' with girls and women. Male body 'women' dominate 'women's' competition with their greater muscle mass, bone density, strength and speed. With that disadvantage sport for women loses all meaning. To put this in perspective nearly 300 American high school boys can run faster than the quickest woman sprinter in the world. Women are seeing their previously protected spaces invaded by male bodies simply because they claim to identify as women. That's all they need to do to access all women's spaces and activities under the Biden 'Equality Act' that destroys equality for women with the usual Orwellian Woke inversion. Male sex offenders have already committed rapes in women's prisons after claiming to identify as women to get them transferred. Does this not matter to the Woke 'equality' hypocrites? Not in the least. What matters to Cult manipulators and funders behind transgender activists is to advance gender fusion on the way to the no-gender 'human'. When you are seeking to impose transparent nonsense like this, or the 'Covid' hoax, the only way the nonsense can prevail is through censorship and intimidation of dissenters, deletion of factual information, and programming of the unquestioning, bewildered and naive. You don't have to scan the world for long to see that all these things are happening.

Many women's rights organisations have realised that rights and status which took such a long time to secure are being eroded and that it is systematic. Kara Dansky of the global Women's Human Rights Campaign said that Biden's transgender executive order immediately he took office, subsequent orders, and Equality Act legislation that followed 'seek to erase women and girls in the law as a category'. *Exactly*. I said during the long ago-started war on men (in which many women play a crucial part) that this was going to turn into a war on them. The Cult is phasing out *both* male and female genders. To get away with that they are brought into conflict so they are busy fighting each other while the Cult completes the job with no unity of response. Unity, people, *unity*. We need unity everywhere. Transgender is the only show in town as the big step towards the no-gender human. It's not about rights for transgender people and never has been. Woke political correctness is deleting words relating to genders to the same end. Wokers believe this is to be 'inclusive' when the opposite is true. They are deleting words describing gender because gender *itself* is being deleted by Human 2.0. Terms like 'man', 'woman', 'mother' and 'father' are being deleted in the universities and other institutions to be replaced by the *no*-gender, not trans-gender, 'individuals' and 'guardians'. Women's rights campaigner Maria Keffler of Partners for Ethical Care said: 'Children are being taught from kindergarten upward that some boys have a vagina, some girls have a penis, and that kids can be any gender they want to be.' Do we really believe that suddenly countries all over the world at the same time had the idea of having drag queens go into schools or read transgender stories to very young children in the local library? It's coldly-calculated confusion of gender on the way to the fusion of gender. Suzanne Vierling, a psychologist from Southern California, made another important point:

> Yesterday's slave woman who endured gynecological medical experiments is today's girl-child being butchered in a booming gender-transitioning sector. Ovaries removed, pushing her into menopause and osteoporosis, uncharted territory, and parents' rights and authority decimated.

The erosion of parental rights is a common theme in line with the Cult plans to erase the very concept of parents and 'ovaries removed, pushing her into menopause' means what? Those born female lose the ability to have children – another way to discontinue humanity as we know it.

Eliminating Human 1.0 (before our very eyes)

To pave the way for Human 2.0 you must phase out Human 1.0. This is happening through plummeting sperm counts and making women infertile through an onslaught of chemicals, radiation (including smartphones in pockets of men) and mRNA 'vaccines'. Common agriculture pesticides are also having a devastating impact on human fertility. I have been tracking collapsing sperm counts in the books for a long time and in 2021 came a book by fertility scientist and reproductive epidemiologist Shanna Swan, *Count Down: How Our Modern World Is Threatening Sperm Counts, Altering Male and Female Reproductive Development and Imperiling the Future of the Human Race*. She reports how the global fertility rate dropped by *half* between 1960 and 2016 with America's birth rate 16 percent below where it needs to be to sustain the population. Women are experiencing declining egg quality, more miscarriages, and more couples suffer from infertility. Other findings were an increase in erectile dysfunction, infant boys developing more genital abnormalities, male problems with conception, and plunging levels of the male hormone testosterone which would explain why so many men have lost their backbone and masculinity. This has been very evident during the 'Covid' hoax when women have been prominent among the Pushbackers and big strapping blokes have bowed their heads, covered their faces with a nappy and quietly submitted. Mind control expert Cathy O'Brien also points to how global education introduced the concept of 'we're all winners' in sport and classrooms: 'Competition was defused, and it in turn defused a sense of fighting back.' This is another version of the 'equity' doctrine in which you drive down rather than raise up. What a contrast in Cult-controlled China with its global ambitions where the government published plans in January, 2021, to 'cultivate masculinity' in boys from kindergarten through to high school in the face of a 'masculinity crisis'. A government adviser said boys would be soon become 'delicate, timid and effeminate' unless action was taken. Don't expect any similar policy in the targeted West. A 2006 study showed that a 65-year-old man in 2002 had testosterone levels *15 percent* lower than a 65-year-old man in 1987 while a 2020 study found a similar story with young adults and adolescents. Men are getting prescriptions for testosterone replacement therapy which causes an even greater drop in sperm count with up to 99 percent seeing sperm counts drop to zero during the treatment. More sperm is defective and malfunctioning with some having two heads or not pursuing an egg.

A class of *synthetic* chemicals known as phthalates are being blamed for the decline. These are found everywhere in plastics, shampoos, cosmetics, furniture, flame retardants, personal care products, pesticides, canned foods and even receipts. Why till receipts? Everyone touches them. Let no one delude themselves that all this is not systematic to advance the long-time agenda for human body transformation. Phthalates mimic hormones and disrupt the hormone balance causing testosterone to fall and genital birth defects in male infants. Animals and fish have been affected in the same way due to phthalates and other toxins in rivers. When fish turn gay or change sex through chemicals in rivers and streams it is a pointer to why there has been such an increase in gay people and the sexually confused. It doesn't matter to me what sexuality people choose to be, but if it's being affected by chemical pollution and consumption then we need to know. Does anyone really think that this is not connected to the transgender agenda, the war on men and the condemnation of male 'toxic masculinity'? You watch this being followed by 'toxic femininity'. It's already happening. When breastfeeding becomes 'chest-feeding', pregnant women become pregnant people along with all the other Woke claptrap you know that the world is going insane and there's a Cult scam in progress. Transgender activists are promoting the Cult agenda while Cult billionaires support and fund the insanity as they laugh themselves to sleep at the sheer stupidity for which humans must be infamous in galaxies far, far away.

'Covid vaccines' and female infertility

We can now see why the 'vaccine' has been connected to potential infertility in women. Dr Michael Yeadon, former Vice President and Chief Scientific Advisor at Pfizer, and Dr Wolfgang Wodarg in Germany, filed a petition with the European Medicines Agency in December, 2020, urging them to stop trials for the Pfizer/BioNTech shot and all other mRNA trials until further studies had been done. They were particularly concerned about possible effects on fertility with 'vaccine'-produced antibodies attacking the protein Syncytin-1 which is responsible for developing the placenta. The result would be infertility 'of indefinite duration' in women who have the 'vaccine' with the placenta failing to form. Section 10.4.2 of the Pfizer/BioNTech trial protocol says that pregnant women or those who might become so should not have mRNA shots. Section 10.4 warns men taking mRNA shots to 'be abstinent from heterosexual intercourse' and not to donate sperm. The UK government said that it *did not know* if the mRNA procedure had an effect on fertility. *Did not know?* These people have to go to jail. UK government advice did not recommend at the start that pregnant women had the shot and said they should avoid pregnancy for at least two months after 'vaccination'. The 'advice' was later updated to pregnant women should only have the 'vaccine' if the benefits outweighed the risks to mother and foetus. What the hell is that supposed to mean? Then 'spontaneous abortions' began to appear and rapidly increase on the adverse reaction reporting schemes which include only a fraction of adverse reactions. Thousands and ever-growing numbers of 'vaccinated' women are describing changes to their menstrual cycle with heavier blood flow, irregular periods and menstruating again after going through the menopause – all links to reproduction effects. Women are passing blood clots and the lining of their uterus while men report erectile dysfunction and blood effects. Most significantly of all *un*vaccinated women began to report similar menstrual changes after interaction with *'vaccinated'* people and men and children were also affected with bleeding noses, blood clots and other conditions. 'Shedding' is when vaccinated people can emit the content of a vaccine to affect the unvaccinated, but this is different. 'Vaccinated' people were not shedding a 'live virus' allegedly in 'vaccines' as before because the fake 'Covid vaccines' involve synthetic material and other toxicity. Doctors exposing what is happening prefer the term 'transmission' to shedding. Somehow those that have had the shots are transmitting effects to those that haven't. Dr Carrie Madej said the nano-content of the 'vaccines' can 'act like an antenna' to others around them which fits perfectly with my own conclusions. This 'vaccine' transmission phenomenon was becoming known as the book went into production and I deal with this further in the Postscript.

Vaccine effects on sterility are well known. The World Health Organization was accused in 2014 of sterilising millions of women in Kenya with the evidence confirmed by the content of the vaccines involved. The same WHO behind the 'Covid' hoax admitted its involvement for more than ten years with the vaccine programme. Other countries made similar claims. Charges were lodged by Tanzania, Nicaragua, Mexico, and the Philippines. The Gardasil vaccine claimed to protect against a genital 'virus' known as HPV has also been linked to infertility. Big Pharma and the WHO (same thing) are criminal and satanic entities. Then there's the Bill Gates Foundation which is connected through funding and shared interests with 20 pharmaceutical giants and laboratories. He stands accused of directing the policy of United Nations Children's Fund (UNICEF), vaccine alliance GAVI, and other groupings, to advance the vaccine agenda and silence opposition at great cost to women and children. At the same time Gates wants to reduce the global population. Coincidence?

Great Reset = Smart Grid = new human

The Cult agenda I have been exposing for 30 years is now being openly promoted by Cult assets like Gates and Klaus Schwab of the World Economic Forum under code-terms like the 'Great Reset', 'Build Back Better' and 'a rare but narrow window of opportunity to reflect, reimagine, and reset our world'. What provided this 'rare but narrow window of opportunity'? The 'Covid' hoax did. Who created that? *They* did. My books from not that long ago warned about the planned 'Internet of Things' (IoT) and its implications for human freedom. This was the plan to connect all technology to the Internet and artificial intelligence and today we are way down that road with an estimated 36 billion devices connected to the World Wide Web and that figure is projected to be 76 billion by 2025. I further warned that the Cult planned to go beyond that to the Internet of *Everything* when the human brain was connected via AI to the Internet and Kurzweil's 'cloud'. Now we have Cult operatives like Schwab calling for precisely that under the term 'Internet of Bodies', a fusion of the physical, digital and biological into one centrally-controlled Smart Grid system which the Cult refers to as the 'Fourth Industrial Revolution'. They talk about the 'biological', but they really mean the synthetic-biological which is required to fully integrate the human body and brain into the Smart Grid and artificial intelligence planned to replace the human mind. We have everything being synthetically manipulated including the natural world through GMO and smart dust, the food we eat and the human body itself with synthetic 'vaccines'. I said in *The Answer* that we would see the Cult push for synthetic meat to replace animals and in February, 2021, the so predictable psychopath Bill Gates called for the introduction of synthetic meat to save us all from 'climate change'. The climate hoax just keeps on giving like the 'Covid' hoax. The war on meat by vegan activists is a carbon (oops, sorry) copy of the manipulation of transgender activists. They have no idea (except their inner core) that they are being used to promote and impose the agenda of the Cult or that they are only the *vehicle* and not the *reason*. This is not to say those who choose not to eat meat shouldn't be respected and supported in that right, but there are ulterior motives for those in power. A *Forbes* article in December, 2019, highlighted the plan so beloved of Schwab and the Cult under the heading: 'What Is The Internet of Bodies? And How Is It Changing Our World?' The article said the human body is the latest data platform (remember 'our vaccine is an operating system'). *Forbes* described the plan very accurately and the words could have come straight out of my books from long before:

> The Internet of Bodies (IoB) is an extension of the IoT and basically connects the human body to a network through devices that are ingested, implanted, or connected to the body in some way. Once connected, data can be exchanged, and the body and device can be remotely monitored and controlled.

They were really describing a human hive mind with human perception centrally-dictated via an AI connection as well as allowing people to be 'remotely monitored and controlled'. Everything from a fridge to a human mind could be directed from a central point by these insane psychopaths and 'Covid vaccines' are crucial to this. Forbes explained the process I mentioned earlier of holdable and wearable technology followed by implantable. The article said there were three generations of the Internet of Bodies that include:

- Body external: These are wearable devices such as Apple Watches or Fitbits that can monitor our health.
- Body internal: These include pacemakers, cochlear implants, and digital pills that go inside our bodies to monitor or control various aspects of health.
- Body embedded: The third generation of the Internet of Bodies is embedded technology where technology and the human body are melded together and have a real-time connection to a remote machine.

Forbes noted the development of the Brain Computer Interface (BCI) which merges the brain with an external device for monitoring and controlling in real-time. 'The ultimate goal is to help restore function to individuals with disabilities by using brain signals rather than conventional neuromuscular pathways.' Oh, do fuck off. The goal of brain interface technology is controlling human thought and emotion from the central point in a hive mind serving its masters wishes. Many people are now agreeing to be chipped to open doors without a key. You can recognise them because they'll be wearing a mask, social distancing and lining up for the 'vaccine'. The Cult plans a Great Reset money system after they have completed the demolition of the global economy in which 'money' will be exchanged through communication with body operating systems. Rand Corporation, a Cult-owned think tank, said of the Internet of Bodies or IoB:

> Internet of Bodies technologies fall under the broader IoT umbrella. But as the name suggests, IoB devices introduce an even more intimate interplay between humans and gadgets. IoB devices monitor the human body, collect health metrics and other personal information, and transmit those data over the Internet. Many devices, such as fitness trackers, are already in use … IoB devices … and those in development can track, record, and store users' whereabouts, bodily functions, and what they see, hear, and even think.

Schwab's World Economic Forum, a long-winded way of saying 'fascism' or 'the Cult', has gone full-on with the Internet of Bodies in the 'Covid' era. 'We're entering the era of the Internet of Bodies', it declared, 'collecting our physical data via a range of devices that can be implanted, swallowed or worn'. The result would be a huge amount of health-related data that could improve human wellbeing around the world, and prove crucial in fighting the 'Covid-19 pandemic'. Does anyone think these clowns care about 'human wellbeing' after the death and devastation their pandemic hoax has purposely caused? Schwab and co say we should move forward with the Internet of Bodies because 'Keeping track of symptoms could help us stop the spread of infection, and quickly detect new cases'. How wonderful, but keeping track' is all they are really bothered about. Researchers were investigating if data gathered from smartwatches and similar devices could be used as viral infection alerts by tracking the user's heart rate and breathing. Schwab said in his 2018 book *Shaping the Future of the Fourth Industrial Revolution*:

> The lines between technologies and beings are becoming blurred and not just by the ability to create lifelike robots or synthetics. Instead it is about the ability of new technologies to literally become part of us. Technologies already influence how we understand ourselves, how we think about each other, and how we determine our realities. As the technologies … give us deeper access to parts of ourselves, we may begin to integrate digital technologies into our bodies.

You can see what the game is. Twenty-four hour control and people – if you could still call them that – would never know when something would go ping and take them out of circulation. It's the most obvious rush to a global fascist dictatorship and the complete submission of humanity and yet still so many are locked away in their Cult-induced perceptual coma and can't see it.

Smart Grid control centres

The human body is being transformed by the 'vaccines' and in other ways into a synthetic cyborg that can be attached to the global Smart Grid which would be controlled from a central point and other sub-locations of Grid manipulation. Where are these planned to be? Well, China for a start which is one of the Cult's biggest centres of operation. The technological control system and technocratic rule was incubated here to be unleashed across the world after the 'Covid' hoax came out of China in 2020. Another Smart Grid location that will surprise people new to this is Israel. I have exposed in *The Trigger* how Sabbatian technocrats, intelligence and military operatives were behind the horrors of 9/11 and not '19 Arab hijackers' who somehow manifested the ability to pilot big passenger airliners when instructors at puddle-jumping flying schools described some of them as a joke. The 9/11 attacks were made possible through control of civilian and military air computer systems and those of the White House, Pentagon and connected agencies. See *The Trigger* – it will blow your mind. The controlling and coordinating force were the Sabbatian networks in Israel and the United States which by then had infiltrated the entire US government, military and intelligence system. The real name of the American Deep State is 'Sabbatian State'. Israel is a tiny country of only nine million people, but it is one of the global centres of cyber operations and fast catching Silicon Valley in importance to the Cult. Israel is known as the 'start-up nation' for all the cyber companies spawned there with the Sabbatian specialisation of 'cyber security' that I mentioned earlier which gives those companies access to computer systems of their clients in real time through 'backdoors' written into the coding when security software is downloaded. The Sabbatian centre of cyber operations outside Silicon Valley is the Israeli military Cyber Intelligence Unit, the biggest infrastructure project in Israel's history, headquartered in the desert-city of Beersheba and involving some 20,000 'cyber soldiers'. Here are located a literal army of Internet trolls scanning social media, forums and comment lists for anyone challenging the Cult agenda. The UK military has something similar with its 77th Brigade and associated operations. The Beersheba complex includes research and development centres for other Cult operations such as Intel, Microsoft, IBM, Google, Apple, Hewlett-Packard, Cisco Systems, Facebook and Motorola. Techcrunch.com ran an article about the Beersheba global Internet technology centre headlined 'Israel's desert city of Beersheba is turning into a cybertech oasis':

> The military's massive relocation of its prestigious technology units, the presence of multinational and local companies, a close proximity to Ben Gurion University and generous government subsidies are turning Beersheba into a major global cybertech hub. Beersheba has all of the ingredients of a vibrant security technology ecosystem, including Ben Gurion University with its graduate program in cybersecurity and Cyber Security Research Center, and the presence of companies such as EMC, Deutsche Telekom, PayPal, Oracle, IBM, and Lockheed Martin. It's also the future home of the INCB (Israeli National Cyber Bureau); offers a special income tax incentive for cyber security companies, and was the site for the relocation of the army's intelligence corps units.

Sabbatians have taken over the cyber world through the following process: They scan the schools for likely cyber talent and develop them at Ben Gurion University and their period of conscription in the Israeli Defense Forces when they are stationed at the Beersheba complex. When the cyber talented officially leave the army they are funded to start cyber companies with technology developed by themselves or given to them by the state. Much of this is stolen through backdoors of computer systems around the world with America top of the list. Others are sent off to Silicon Valley to start companies or join the major ones and so we have many major positions filled by apparently 'Jewish' but really Sabbatian operatives. Google, YouTube and Facebook are all run by 'Jewish' CEOs while Twitter is all but run by ultra-Zionist hedge-fund shark Paul Singer. At the centre of the Sabbatian global cyber web is the Israeli army's Unit 8200 which specialises in hacking into computer systems of other countries, inserting viruses, gathering information, instigating malfunction, and even taking control of them from a distance. A long list of Sabbatians involved with 9/11, Silicon Valley and Israeli cyber security companies are operatives of Unit 8200. This is not about Israel. It's about the Cult. Israel is planned to be a Smart Grid hub as with China and what is happening at Beersheba is not for the benefit of Jewish people who are treated disgustingly by the Sabbatian elite that control the country. A glance at the Nuremberg Codes will tell you that.

The story is much bigger than 'Covid', important as that is to where we are being taken. Now, though, it's time to really strap in. There's more … much more …

CHAPTER ELEVEN

Who controls the Cult?

Awake, arise or be forever fall'n
John Milton, *Paradise Lost*

I have exposed this far the level of the Cult conspiracy that operates in the world of the seen and within the global secret society and satanic network which operates in the shadows one step back from the seen. The story, however, goes much deeper than that.

The 'Covid' hoax is major part of the Cult agenda, but only part, and to grasp the biggest picture we have to expand our attention beyond the realm of human sight and into the infinity of possibility that we cannot see. It is from here, ultimately, that humanity is being manipulated into a state of total control by the force which dictates the actions of the Cult. How much of reality can we see? Next to damn all is the answer. We may appear to see all there is to see in the 'space' our eyes survey and observe, but little could be further from the truth. The human 'world' is only a tiny band of frequency that the body's visual and perceptual systems can decode into *perception* of a 'world'. According to mainstream science the electromagnetic spectrum is 0.005 percent of what exists in the Universe (Fig 10). The maximum estimate I have seen is 0.5 percent and either way it's miniscule. I say it is far, far, smaller even than 0.005 percent when you compare reality we see with the totality of reality that we don't. Now get this if you are new to such information: Visible light, the only band of frequency that we can see, is a *fraction* of the 0.005

percent (Fig 11 overleaf). Take this further and realise that our universe is one of infinite universes and that universes are only a fragment of overall reality – *infinite* reality. Then compare that with the almost infinitesimal frequency band of visible light or human sight. You see that humans are as near blind as it is possible to be without actually being so. Artist and filmmaker, Sergio Toporek, said:

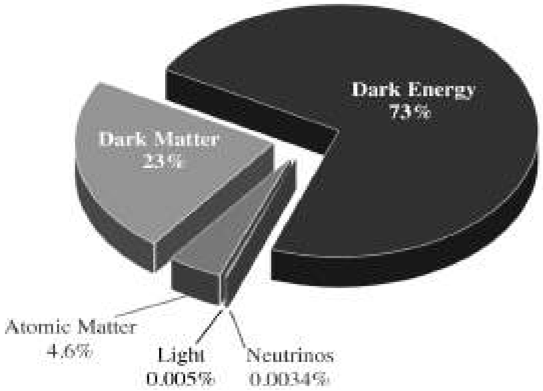


Figure 10: Humans can perceive such a tiny band of visual reality it's laughable.

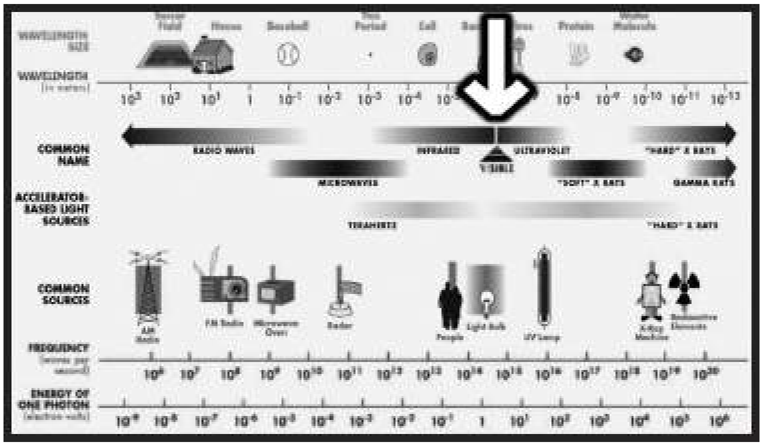

WAVELENGTH SIZE
Soccer Field
House
Baseball
This Period
Cell
Protein
Water Molecule
WAVELENGTH (in meters)
COMMON NAME
RADIO WAVES
INFRARED
VISIBLE
ULTRAVIOLET
"HARD" X RAYS
MICROWAVES
"SOFT" X RAYS
GAMMA RAYS
ACCELERATOR-BASED LIGHT SOURCES
TERAHERTZ
"HARD" X RAYS
COMMON SOURCES
AM Radio
FM Radio
Microwave Oven
Radar
People
Light Bulb
UV Lamp
X-Ray Machine
Radioactive Elements
FREQUENCY (waves per second)
ENERGY OF ONE PHOTON (electron volts)

Figure 11: We can see a smear of the 0.005 percent electromagnetic spectrum, but we still know it all. Yep, makes sense.

Consider that you can see less than 1% of the electromagnetic spectrum and hear less than 1% of the acoustic spectrum. 90% of the cells in your body carry their own microbial DNA and are not 'you'. The atoms in your body are 99.9999999999999999% empty space and none of them are the ones you were born with ... Human beings have 46 chromosomes, two less than a potato.

The existence of the rainbow depends on the conical photoreceptors in your eyes; to animals without cones, the rainbow does not exist. So you don't just look at a rainbow, you create it. This is pretty amazing, especially considering that all the beautiful colours you see represent less than 1% of the electromagnetic spectrum.

Suddenly the 'world' of humans looks a very different place. Take into account, too, that Planet Earth when compared with the projected size of this single universe is the equivalent of a billionth of a pinhead. Imagine the ratio that would be when compared to infinite reality. To think that Christianity once insisted that Earth and humanity were the centre of everything. This background is vital if we are going to appreciate the nature of 'human' and how we can be manipulated by an unseen force. To human visual reality virtually *everything* is unseen and yet the prevailing perception within the institutions and so much of the public is that if we can't see it, touch it, hear it, taste it and smell it then it cannot exist. Such perception is indoctrinated and encouraged by the Cult and its agents because it isolates believers in the strictly limited, village-idiot, realm of the five senses where perceptions can be firewalled and information controlled. Most of those perpetuating the 'this-world-is-all-there-is' insanity are themselves indoctrinated into believing the same delusion. While major players and influencers know that official reality is laughable most of those in science, academia and medicine really believe the nonsense they peddle and teach succeeding generations. Those who challenge the orthodoxy are dismissed as nutters and freaks to protect the manufactured illusion from exposure. Observe the dynamic of the 'Covid' hoax and you will see how that takes the same form. The inner-circle psychopaths knows it's a gigantic scam, but almost the entirety of those imposing their fascist rules believe that 'Covid' is all that they're told it is.

Stolen identity

Ask people who they are and they will give you their name, place of birth, location, job, family background and life story. Yet that is not who they are – it is what they are *experiencing*. The difference is *absolutely crucial*. The true 'I', the eternal, infinite 'I', is consciousness, a state of being aware. Forget 'form'. That is a vehicle for a brief experience. Consciousness does not come *from* the brain, but *through* the brain and even that is more symbolic than literal. We are awareness, pure awareness, and this is what withdraws from the body at what we call 'death' to continue our eternal beingness, *isness*, in other realms of reality within the limitlessness of infinity or the Biblical 'many mansions in my father's house'. Labels of a human life, man, woman, transgender, black, white, brown, nationality, circumstances and income are not who we are. They are what we are – awareness – is *experiencing* in a brief connection with a band of frequency we call 'human'. The labels are not the self; they are, to use the title of one of my books, a *Phantom* Self. I am not David Icke born in Leicester, England, on April 29th, 1952. I am the consciousness *having that experience*. The Cult and its non-human masters seek to convince us through the institutions of 'education', science, medicine, media and government that what we are *experiencing* is who we *are*. It's so easy to control and direct perception locked away in the bewildered illusions of the five senses with no expanded radar. Try, by contrast, doing the same with a humanity aware of its true self and its true power to consciously create its reality and experience. How is it possible to do this? We do it all day every day. If you perceive yourself as 'little me' with no power to impact upon your life and the world then your life experience will reflect that. You will hand the power you don't think you have to authority in all its forms which will use it to control your experience. This, in turn, will appear to confirm your perception of 'little me' in a self-fulfilling feedback loop. But that is what 'little me' really is – a *perception*. We are all 'big-me', infinite me, and the Cult has to make us forget that if its will is to prevail. We are therefore manipulated and pressured into self-identifying with human labels and not the consciousness/awareness *experiencing* those human labels.

The phenomenon of identity politics is a Cult-instigated manipulation technique to sub-divide previous labels into even smaller ones. A United States university employs this list of letters to describe student identity: LGBTTQQFAGPBDSM or lesbian, gay, bisexual, transgender, transsexual, queer, questioning, flexual, asexual, gender-fuck, polyamorous, bondage/discipline, dominance/submission and sadism/masochism. I'm sure other lists are even longer by now as people feel the need to self-identity the 'I' with the minutiae of race and sexual preference. Wokers programmed by the Cult for generations believe this is about 'inclusivity' when it's really the Cult locking them away into smaller and smaller versions of Phantom Self while firewalling them from the influence of their true self, the infinite, eternal 'I'. You may notice that my philosophy which contends that we are all unique points of attention/awareness within the same infinite whole or Oneness is the ultimate non-racism. The very sense of Oneness makes the judgement of people by their body-type, colour or sexuality utterly ridiculous and confirms that racism has no understanding of reality (including anti-white racism). Yet despite my perception of life Cult agents and fast-asleep Wokers label me racist to discredit my information while they are themselves phenomenally racist and sexist. All they see is race and sexuality and they judge people as good or bad, demons or untouchables, by their race and sexuality. All they see is *Phantom Self* and perceive themselves in terms of Phantom Self. They are pawns and puppets of the Cult agenda to focus attention and self-identity in the five senses and play those identities against each other to divide and rule. Columbia University has introduced segregated graduations in another version of social distancing designed to drive people apart and teach them that different racial and cultural groups have nothing in common with each other. The last thing the Cult wants is unity. Again the pump-primers of this will be Cult operatives in the knowledge of what they are doing, but the rest are just the Phantom Self blind leading the Phantom Self blind. We *do* have something in common – we are all *the same consciousness* having different temporary experiences.

What is this 'human'?

Yes, what *is* 'human'? That is what we are supposed to be, right? I mean 'human'? True, but 'human' is the experience not the 'I'. Break it down to basics and 'human' is the way that information is processed. If we are to experience and interact with this band of frequency we call the 'world' we must have a vehicle that operates within that band of frequency. Our consciousness in its prime form cannot do that; it is way beyond the frequency of the human realm. My consciousness or awareness could not tap these keys and pick up the cup in front of me in the same way that radio station A cannot interact with radio station B when they are on different frequencies. The human body is the means through which we have that interaction. I have long described the body as a biological computer which processes information in a way that allows consciousness to experience this reality. The body is a receiver, transmitter and processor of information in a particular way that we call human. We visually perceive only the world of the five senses in a wakened state – that is the limit of the body's visual decoding system. In truth it's not even visual in the way we experience 'visual reality' as I will come to in a moment. We are 'human' because the body processes the information sources of human into a reality and behaviour system that we *perceive* as human. Why does an elephant act like an elephant and not like a human or a duck? The elephant's biological computer is a different information field and processes information according to that program into a visual and behaviour type we call an elephant. The same applies to everything in our reality. These body information fields are perpetuated through procreation (like making a copy of a software program). The Cult wants to break that cycle and intervene technologically to transform the human information field into one that will change what we call humanity. If it can change the human information field it will change the way that field processes information and change humanity both 'physically' and psychologically. Hence the *messenger* (information) RNA 'vaccines' and so much more that is targeting human genetics by changing the body's information – *messaging* – construct through food, drink, radiation, toxicity and other means.

Reality that we experience is nothing like reality as it really is in the same way that the reality people experience in virtual reality games is not the reality they are really living in. The game is only a decoded source of information that appears to be a reality. Our world is also an information construct – a *simulation* (more later). In its base form our reality is a wavefield of information much the same in theme as Wi-Fi. The five senses decode wavefield information into electrical information which they communicate to the brain to decode into holographic (illusory 'physical') information. Different parts of the brain specialise in decoding different senses and the information is fused into a reality that appears to be outside of us but is really inside the brain and the genetic structure in general (Fig 12 overleaf). DNA is a receiver-transmitter of information and a vital part of this decoding process and the body's connection to other realities. Change DNA and you change the way we decode and connect with reality – see 'Covid vaccines'. Think of computers decoding Wi-Fi. You have information encoded in a radiation field and the computer decodes that information into a very different form on the screen. You can't see the Wi-Fi until its information is made manifest on the screen and the information on the screen is inside the computer and not outside. I have just described how we decode the 'human world'. All five senses decode the waveform 'Wi-Fi' field into electrical signals and the brain (computer) constructs reality inside the brain and not outside – 'You don't just look at a rainbow, you create it'. Sound is a simple example. We don't hear sound until the brain decodes it. Waveform sound waves are picked up by the hearing sense and communicated to the brain in an electrical form to be decoded into the sounds that we hear. Everything we hear is inside the brain along with everything we see, feel, smell and taste. Words and language are waveform fields generated by our vocal chords which pass through this process until they are decoded by the brain into words that we hear. Different languages are different frequency fields or sound waves generated by vocal chords. Late British philosopher Alan Watts said:

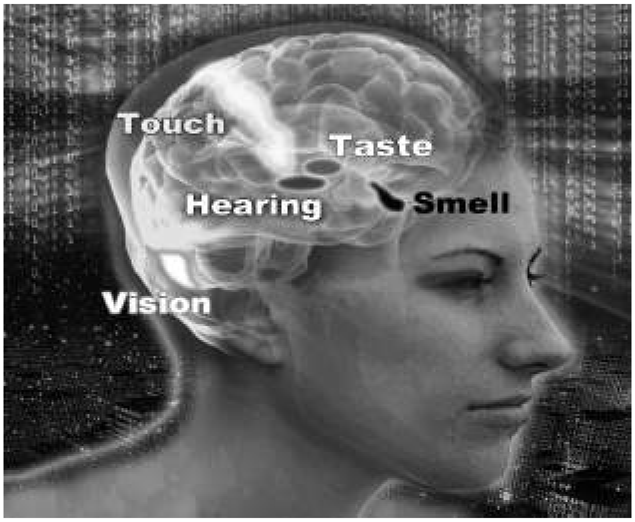


Figure 12: The brain receives information from the five senses and constructs from that our perceived reality.

> [Without the brain] the world is devoid of light, heat, weight, solidity, motion, space, time or any other imaginable feature. All these phenomena are interactions, or transactions, of vibrations with a certain arrangement of neurons.

That's exactly what they are and scientist Robert Lanza describes in his book, *Biocentrism*, how we decode electromagnetic waves and energy into visual and 'physical' experience. He uses the example of a flame emitting photons, electromagnetic energy, each pulsing electrically and magnetically:

> … these … invisible electromagnetic waves strike a human retina, and if (and only if) the waves happen to measure between 400 and 700 nano meters in length from crest to crest, then their energy is just right to deliver a stimulus to the 8 million cone-shaped cells in the retina.
>
> Each in turn send an electrical pulse to a neighbour neuron, and on up the line this goes, at 250 mph, until it reaches the … occipital lobe of the brain, in the back of the head. There, a cascading complex of neurons fire from the incoming stimuli, and we subjectively perceive this experience as a yellow brightness occurring in a place we have been conditioned to call the 'external world'.

You hear what you decode

If a tree falls or a building collapses they make no noise unless someone is there to decode the energetic waves generated by the disturbance into what we call sound. Does a falling tree make a noise? Only if you hear it – *decode* it. Everything in our reality is a frequency field of information operating within the overall 'Wi-Fi' field that I call The Field. A vibrational disturbance is generated in The Field by the fields of the falling tree or building. These disturbance waves are what we decode into the sound of them falling. If no one is there to do that then neither will make any noise. Reality is created by the observer – *decoder* – and the *perceptions* of the observer affect the decoding process. For this reason different people – different *perceptions* – will perceive the same reality or situation in a different way. What one may perceive as a nightmare another will see as an opportunity. The question of why the Cult is so focused on controlling human perception now answers itself. All experienced reality is the act of decoding and we don't experience Wi-Fi until it is decoded on the computer screen. The sight and sound of an Internet video is encoded in the Wi-Fi all around us, but we don't see or hear it until the computer decodes that information. Taste, smell and touch are all phenomena of the brain as a result of the same process. We don't taste, smell or feel anything except in the brain and there are pain relief techniques that seek to block the signal from the site of discomfort to the brain because if the brain doesn't decode that signal we don't feel pain. Pain is in the brain and only appears to be at the point of impact thanks to the feedback loop between them. We don't see anything until electrical information from the sight senses is decoded in an area at the back of the brain. If that area is damaged we can go blind when our eyes are perfectly okay. So why do we go blind if we damage an eye? We damage the information processing between the waveform visual information and the visual decoding area of the brain. If information doesn't reach the brain in a form it can decode then we can't see the visual reality that it represents. What's more the brain is decoding only a fraction of the information it receives and the rest is absorbed by the sub-conscious mind. This explanation is from the science magazine, *Wonderpedia*:

> Every second, 11 million sensations crackle along these [brain] pathways … The brain is confronted with an alarming array of images, sounds and smells which it rigorously filters down until it is left with a manageable list of around 40. Thus 40 sensations per second make up what we perceive as reality.

The ‘world’ is not what people are told to believe that is it and the inner circles of the Cult *know that*.

Illusory ‘physical’ reality

We can only see a smear of 0.005 percent of the Universe which is only one of a vast array of universes – ‘mansions’ – within infinite reality. Even then the brain decodes only 40 pieces of information (‘sensations’) from a potential *11 million* that we receive every second. Two points strike you from this immediately: The sheer breathtaking stupidity of believing we know anything so rigidly that there’s nothing more to know; and the potential for these processes to be manipulated by a malevolent force to control the reality of the population. One thing I can say for sure with no risk of contradiction is that when you can perceive an almost indescribable fraction of infinite reality there is always more to know as in tidal waves of it. Ancient Greek philosopher Socrates was so right when he said that wisdom is to know how little we know. How obviously true that is when you think that we are experiencing a physical world of solidity that is neither physical nor solid and a world of apartness when everything is connected. Cult-controlled ‘science’ dismisses the so-called ‘paranormal’ and all phenomena related to that when the ‘para’-normal is perfectly normal and explains the alleged ‘great mysteries’ which dumbfound scientific minds. There is a reason for this. A ‘scientific mind’ in terms of the mainstream is a material mind, a five-sense mind imprisoned in see it, touch it, hear it, smell it and taste it. Phenomena and happenings that can’t be explained that way leave the ‘scientific mind’ bewildered and the rule is that if they can't account for why something is happening then it can't, by definition, be happening. I beg to differ. Telepathy is thought waves passing through The Field (think wave disturbance again) to be decoded by someone able to connect with that wavelength (information). For example: You can pick up the thought waves of a friend at any distance and at the very least that will bring them to mind. A few minutes later the friend calls you. 'My god', you say, 'that's incredible – I was just thinking of you.' Ah, but *they* were thinking of *you* before they made the call and that's what you decoded. Native peoples not entrapped in five-sense reality do this so well it became known as the 'bush telegraph'. Those known as psychics and mediums (genuine ones) are doing the same only across dimensions of reality. 'Mind over matter' comes from the fact that matter and mind are the *same*. The state of one influences the state of the other. Indeed one *and* the other are illusions. They are aspects of the same field. Paranormal phenomena are all explainable so why are they still considered 'mysteries' or not happening? Once you go down this road of understanding you begin to expand awareness beyond the five senses and that's the nightmare for the Cult.

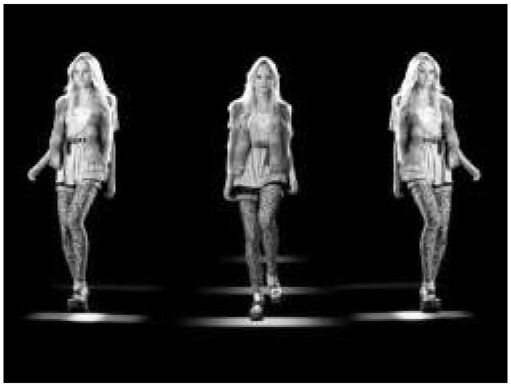


Figure 13: Holograms are not solid, but the best ones appear to be.

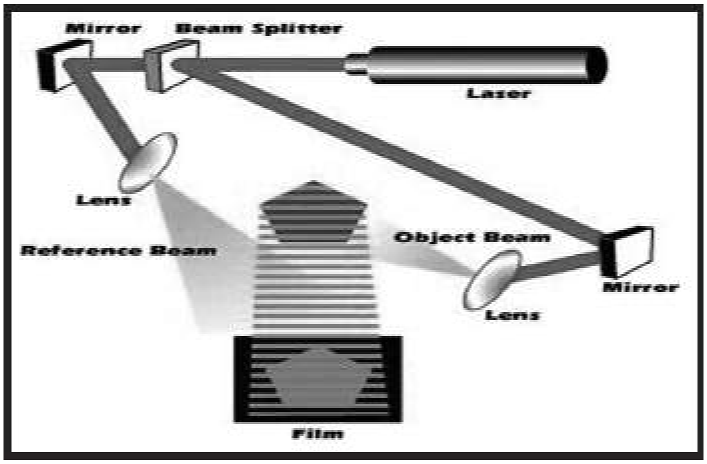


Figure 14: How holograms are created by capturing a waveform version of the subject image.

Holographic 'solidity'

Our reality is not solid, it is holographic. We are now well aware of holograms which are widely used today. Two-dimensional information is decoded into a three-dimensional reality that is not solid although can very much appear to be (Fig 13). Holograms are created with a laser divided into two parts. One goes directly onto a holographic photographic print ('reference beam') and the other takes a waveform image of the subject ('working beam') before being directed onto the print where it 'collides' with the other half of the laser (Fig 14). This creates a *waveform* interference pattern which contains the wavefield information of whatever is being photographed (Fig 15 overleaf). The process can be likened to dropping pebbles in a pond. Waves generated by each one spread out across the water to collide with the others and create a wave representation of where the stones fell and at what speed, weight and distance. A waveform interference pattern of a hologram is akin to the waveform information in The Field which the five senses decode into electrical signals to be decoded by the brain into a holographic illusory 'physical' reality. In the same way when a laser (think human attention) is directed at the waveform interference pattern a three-dimensional version of the subject is projected into apparently 'solid' reality (Fig 16). An amazing trait of holograms reveals more 'paranormal mysteries'. Information of the *whole* hologram is encoded in waveform in every part of the interference pattern by the way they are created. This means that every *part* of a hologram is a smaller version of the whole. Cut the interference wave-pattern into four and you won't get four parts of the image. You get quarter-sized versions of the *whole* image. The body is a hologram and the same applies. Here we have the basis of acupuncture, reflexology and other forms of healing which identify representations of the whole body in all of the parts, hands, feet, ears, everywhere. Skilled palm readers can do what they do because the information of whole body is encoded in the hand. The concept of as above, so below, comes from this.

Figure 15: A waveform interference pattern that holds the information that transforms into a hologram.

She isn't really standing there
WILLIAM VIA HOLOGRAM
PERFORMER & OBAMA SUPPORTER
LIVE

Figure 16: Holographic people including 'Elvis' holographically inserted to sing a duet with Celine Dion.

The question will be asked of why, if solidity is illusory, we can't just walk through walls and each other. The resistance is not solid against solid; it is electromagnetic field against electromagnetic field and we decode this into the *experience* of solid against solid. We should also not underestimate the power of belief to dictate reality. What you believe is impossible *will be*. Your belief impacts on your decoding processes and they won't decode what you think is impossible. What we believe we perceive and what we perceive we experience. 'Can't dos' and 'impossibles' are like a firewall in a computer system that won't put on the screen what the firewall blocks. How vital that is to understanding how human experience has been hijacked. I explain in *The Answer*, *Everything You Need To Know But Have Never Been Told* and other books a long list of 'mysteries' and 'paranormal' phenomena that are not mysterious and perfectly normal once you realise what reality is and how it works. 'Ghosts' can be seen to pass through 'solid' walls because the walls are not solid and the ghost is a discarnate entity operating on a frequency so different to that of the wall that it's like two radio stations sharing the same space while never interfering with each other. I have seen ghosts do this myself. The apartness of people and objects is also an illusion. Everything is connected by the Field like all sea life is connected by the sea. It's just that within the limits of our visual reality we only 'see' holographic information and not the field of information that connects everything and from which the holographic world is made manifest. If you can only see holographic 'objects' and not the field that connects them they will appear to you as unconnected to each other in the same way that we see the computer while not seeing the Wi-Fi.

What you don't know *can* hurt you

Okay, we return to those 'two worlds' of human society and the Cult with its global network of interconnecting secret societies and satanic groups which manipulate through governments, corporations, media, religions, etc. The fundamental difference between them is *knowledge*. The idea has been to keep humanity ignorant of the plan for its total enslavement underpinned by a crucial ignorance of reality – who we are and where we are – and how we interact with it. 'Human' should be the interaction between our expanded eternal consciousness and the five-sense body experience. We are meant to be *in* this world in terms of the five senses but not *of* this world in relation to our greater consciousness and perspective. In that state we experience the small picture of the five senses within the wider context of the big picture of awareness beyond the five senses. Put another way the five senses see the dots and expanded awareness connects them into pictures and patterns that give context to the apparently random and unconnected. Without the context of expanded awareness the five senses see only apartness and randomness with apparently no meaning. The Cult and its other-dimensional controllers seek to intervene in the frequency realm where five-sense reality is supposed to connect with expanded reality and to keep the two apart (more on this in the final chapter). When that happens five-sense mental and emotional processes are no longer influenced by expanded awareness, or the True 'I', and instead are driven by the isolated perceptions of the body's decoding systems. They are in the world *and* of it. Here we have the human plight and why humanity with its potential for infinite awareness can be so easily manipulatable and descend into such extremes of stupidity.

Once the Cult isolates five-sense mind from expanded awareness it can then program the mind with perceptions and beliefs by controlling information that the mind receives through the 'education' system of the formative years and the media perceptual bombardment and censorship of an entire lifetime. Limit perception and a sense of the possible through limiting knowledge by limiting and skewing information while censoring and discrediting that which could set people free. As the title of another of my books says *… And The Truth Shall Set You Free*. For this reason the last thing the Cult wants in circulation is the truth about anything – especially the reality of the eternal 'I' – and that's why it is desperate to control information. The Cult knows that information becomes perception which becomes behaviour which, collectively, becomes human society. Cult-controlled and funded mainstream 'science' denies the existence of an eternal 'I' and seeks to dismiss and trash all evidence to the contrary. Cult-controlled mainstream religion has a version of 'God' that is little more than a system of control and dictatorship that employs threats of damnation in an afterlife to control perceptions and behaviour in the here and now through fear and guilt. Neither is true and it's the 'neither' that the Cult wishes to suppress. This 'neither' is that everything is an expression, a point of attention, within an infinite state of consciousness which is the real meaning of the term 'God'.

Perceptual obsession with the 'physical body' and five-senses means that 'God' becomes personified as a bearded bloke sitting among the clouds or a raging bully who loves us if we do what 'he' wants and condemns us to the fires of hell if we don't. These are no more than a 'spiritual' fairy tales to control and dictate events and behaviour through fear of this 'God' which has bizarrely made 'God-fearing' in religious circles a state to be desired. I would suggest that fearing *anything* is not to be encouraged and celebrated, but rather deleted. You can see why 'God fearing' is so beneficial to the Cult and its religions when *they* decide what 'God' wants and what 'God' demands (the Cult demands) that everyone do. As the great American comedian Bill Hicks said satirising a Christian zealot: 'I think what God meant to say.' How much of this infinite awareness ('God') that we access is decided by how far we choose to expand our perceptions, self-identity and sense of the possible. The scale of self-identity reflects itself in the scale of awareness that we can connect with and are influenced by – how much knowing and insight we have instead of programmed perception. You cannot expand your awareness into the infinity of possibility when you believe that you are little me Peter the postman or Mary in marketing and nothing more. I'll deal with this in the concluding chapter because it's crucial to how we turnaround current events.

Where the Cult came from

When I realised in the early 1990s there was a Cult network behind global events I asked the obvious question: When did it start? I took it back to ancient Rome and Egypt and on to Babylon and Sumer in Mesopotamia, the 'Land Between Two Rivers', in what we now call Iraq. The two rivers are the Tigris and Euphrates and this region is of immense historical and other importance to the Cult, as is the land called Israel only 550 miles away by air. There is much more going with deep esoteric meaning across this whole region. It's not only about 'wars for oil'. Priceless artefacts from Mesopotamia were stolen or destroyed after the American and British invasion of Iraq in 2003 justified by the lies of Boy Bush and Tony Blair (their Cult masters) about non-existent 'weapons of mass destruction'. Mesopotamia was the location of Sumer (about 5,400BC to 1,750BC), and Babylon (about 2,350BC to 539BC). Sabbatians may have become immensely influential in the Cult in modern times but they are part of a network that goes back into the mists of history. Sumer is said by historians to be the 'cradle of civilisation'. I disagree. I say it was the re-start of what we call human civilisation after cataclysmic events symbolised in part as the 'Great Flood' destroyed the world that existed before. These fantastic upheavals that I have been describing in detail in the books since the early1990s appear in accounts and legends of ancient cultures across the world and they are supported by geological and biological evidence. Stone tablets found in Iraq detailing the Sumer period say the cataclysms were caused by non-human 'gods' they call the Anunnaki. These are described in terms of extraterrestrial visitations in which knowledge supplied by the Anunnaki is said to have been the source of at least one of the world's oldest writing systems and developments in astronomy, mathematics and architecture that were way ahead of their time. I have covered this subject at length in *The Biggest Secret* and *Children of the Matrix* and the same basic 'Anunnaki' story can be found in Zulu accounts in South Africa where the late and very great Zulu high shaman Credo Mutwa told me that the Sumerian Anunnaki were known by Zulus as the Chitauri or 'children of the serpent'. See my six-hour video interview with Credo on this subject entitled *The*

Reptilian Agenda recorded at his then home near Johannesburg in 1999 which you can watch on the Ickonic media platform.

The Cult emerged out of Sumer, Babylon and Egypt (and elsewhere) and established the Roman Empire before expanding with the Romans into northern Europe from where many empires were savagely imposed in the form of Cult-controlled societies all over the world. Mass death and destruction was their calling card. The Cult established its centre of operations in Europe and European Empires were Cult empires which allowed it to expand into a global force. Spanish and Portuguese colonialists headed for Central and South America while the British and French targeted North America. Africa was colonised by Britain, France, Belgium, the Netherlands, Portugal, Spain, Italy, and Germany. Some like Britain and France moved in on the Middle East. The British Empire was by far the biggest for a simple reason. By now Britain was the headquarters of the Cult from which it expanded to form Canada, the United States, Australia and New Zealand. The Sun never set on the British Empire such was the scale of its occupation. London remains a global centre for the Cult along with Rome and the Vatican although others have emerged in Israel and China. It is no accident that the 'virus' is alleged to have come out of China while Italy was chosen as the means to terrify the Western population into compliance with 'Covid' fascism. Nor that Israel has led the world in 'Covid' fascism and mass 'vaccination'.

You would think that I would mention the United States here, but while it has been an important means of imposing the Cult's will it is less significant than would appear and is currently in the process of having what power it does have deleted. The Cult in Europe has mostly loaded the guns for the US to fire. America has been controlled from Europe from the start through Cult operatives in Britain and Europe. The American Revolution was an illusion to make it appear that America was governing itself while very different forces were pulling the strings in the form of Cult families such as the Rothschilds through the Rockefellers and other subordinates. The Rockefellers are extremely close to Bill Gates and established both scalpel and drug 'medicine' and the World Health Organization. They play a major role in the development and circulation of vaccines through the Rockefeller Foundation on which Bill Gates said his Foundation is based. Why wouldn't this be the case when the Rockefellers and Gates are on the same team? Cult infiltration of human society goes way back into what we call history and has been constantly expanding and centralising power with the goal of establishing a global structure to dictate everything. Look how this has been advanced in great leaps with the 'Covid' hoax.

The non-human dimension

I researched and observed the comings and goings of Cult operatives through the centuries and even thousands of years as they were born, worked to promote the agenda within the secret society and satanic networks, and then died for others to replace them. Clearly there had to be a coordinating force that spanned this entire period while operatives who would not have seen the end goal in their lifetimes came and went advancing the plan over millennia. I went in search of that coordinating force with the usual support from the extraordinary synchronicity of my life which has been an almost daily experience since 1990. I saw common themes in religious texts and ancient cultures about a non-human force manipulating human society from the hidden. Christianity calls this force Satan, the Devil and demons; Islam refers to the Jinn or Djinn; Zulus have their Chitauri (spelt in other ways in different parts of Africa); and the Gnostic people in Egypt in the period around and before 400AD referred to this phenomena as the 'Archons', a word meaning rulers in Greek. Central American cultures speak of the 'Predators' among other names and the same theme is everywhere. I will use 'Archons' as a collective name for all of them. When you see how their nature and behaviour is described all these different sources are clearly talking about the same force. Gnostics described the Archons in terms of 'luminous fire' while Islam relates the Jinn to 'smokeless fire'. Some refer to beings in form that could occasionally be seen, but the most common of common theme is that they operate from unseen realms which means almost all existence to the visual processes of humans. I had concluded that this was indeed the foundation of human control and that the Cult was operating within the human frequency band on behalf of this hidden force when I came across the writings of Gnostics which supported my conclusions in the most extraordinary way.

A sealed earthen jar was found in 1945 near the town of Nag Hammadi about 75-80 miles north of Luxor on the banks of the River Nile in Egypt. Inside was a treasure trove of manuscripts and texts left by the Gnostic people some 1,600 years earlier. They included 13 leather-bound papyrus codices (manuscripts) and more than 50 texts written in Coptic Egyptian estimated to have been hidden in the jar in the period of 400AD although the source of the information goes back much further. Gnostics oversaw the Great or Royal Library of Alexandria, the fantastic depository of ancient texts detailing advanced knowledge and accounts of human history. The Library was dismantled and destroyed in stages over a long period with the death-blow delivered by the Cult-established Roman Church in the period around 415AD. The Church of Rome was the Church of Babylon relocated as I said earlier. Gnostics were not a race. They were a way of perceiving reality. Whenever they established themselves and their information circulated the terrorists of the Church of Rome would target them for destruction. This happened with the Great Library and with the Gnostic Cathars who were burned to death by the psychopaths after a long period of oppression at the siege of the Castle of Monségur in southern France in 1244. The Church has always been terrified of Gnostic information which demolishes the official Christian narrative although there is much in the Bible that supports the Gnostic view if you read it in another way. To anyone studying the texts of what became known as the Nag Hammadi Library it is clear that great swathes of Christian and Biblical belief has its origin with Gnostics sources going back to Sumer. Gnostic themes have been twisted to manipulate the perceived reality of Bible believers. Biblical texts have been in the open for centuries where they could be changed while Gnostic documents found at Nag Hammadi were sealed away and untouched for 1,600 years. What you see is what they wrote.

Use your *pneuma* not your *nous*

Gnosticism and Gnostic come from 'gnosis' which means knowledge, or rather *secret* knowledge, in the sense of spiritual awareness – knowledge about reality and life itself. The desperation of the Cult's Church of Rome to destroy the Gnostics can be understood when the knowledge they were circulating was the last thing the Cult wanted the population to know. Sixteen hundred years later the same Cult is working hard to undermine and silence me for the same reason. The dynamic between knowledge and ignorance is a constant. 'Time' appears to move on, but essential themes remain the same. We are told to 'use your nous', a Gnostic word for head/brain/intelligence. They said, however, that spiritual awakening or 'salvation' could only be secured by expanding awareness *beyond* what they called *nous* and into *pneuma* or Infinite Self. Obviously as I read these texts the parallels with what I have been saying since 1990 were fascinating to me. There is a universal truth that spans human history and in that case why wouldn't we be talking the same language 16 centuries apart? When you free yourself from the perception program of the five senses and explore expanded realms of consciousness you are going to connect with the same information no matter what the perceived 'era' within a manufactured timeline of a single and tiny range of manipulated frequency. Humans working with 'smart' technology or knocking rocks together in caves is only a timeline appearing to operate within the human frequency band. Expanded awareness and the knowledge it holds have always been there whether the era be Stone Age or computer age. We can only access that knowledge by opening ourselves to its frequency which the five-sense prison cell is designed to stop us doing. Gates, Fauci, Whitty, Vallance, Zuckerberg, Brin, Page, Wojcicki, Bezos, and all the others behind the 'Covid' hoax clearly have a long wait before their range of frequency can make that connection given that an open heart is crucial to that as we shall see. Instead of accessing knowledge directly through expanded awareness it is given to Cult operatives by the secret society networks of the Cult where it has been passed on over thousands of years outside the public arena. Expanded realms of consciousness is where great artists, composers and writers find their inspiration and where truth awaits anyone open enough to connect with it. We need to go there fast.

Archon hijack

A fifth of the Nag Hammadi texts describe the existence and manipulation of the Archons led by a 'Chief Archon' they call 'Yaldabaoth', or the 'Demiurge', and this is the Christian 'Devil', 'Satan', 'Lucifer', and his demons. Archons in Biblical symbolism are the 'fallen ones' which are also referred to as fallen angels after the angels expelled from heaven according to the Abrahamic religions of Judaism, Christianity and Islam. These angels are claimed to tempt humans to 'sin' ongoing and you will see how accurate that symbolism is during the rest of the book. The theme of 'original sin' is related to the 'Fall' when Adam and Eve were 'tempted by the serpent' and fell from a state of innocence and 'obedience' (connection) with God into a state of disobedience (disconnection). The Fall is said to have brought sin into the world and corrupted everything including human nature. Yaldabaoth, the 'Lord Archon', is described by Gnostics as a 'counterfeit spirit', 'The Blind One', 'The Blind God', and 'The Foolish One'. The Jewish name for Yaldabaoth in Talmudic writings is Samael which translates as 'Poison of God', or 'Blindness of God'. You see the parallels. Yaldabaoth in Islamic belief is the Muslim Jinn devil known as Shaytan – Shaytan is Satan as the same themes are found all over the world in every religion and culture. The 'Lord God' of the Old Testament is the 'Lord Archon' of Gnostic manuscripts and that's why he's such a bloodthirsty bastard. Satan is known by Christians as 'the Demon of Demons' and Gnostics called Yaldabaoth the 'Archon of Archons'. Both are known as 'The Deceiver'. We are talking about the same 'bloke' for sure and these common themes using different names, storylines and symbolism tell a common tale of the human plight.

Archons are referred to in Nag Hammadi documents as mind parasites, inverters, guards, gatekeepers, detainers, judges, pitiless ones and deceivers. The 'Covid' hoax alone is a glaring example of all these things. The Biblical 'God' is so different in the Old and New Testaments because they are not describing the same phenomenon. The vindictive, angry, hate-filled, 'God' of the Old Testament, known as Yahweh, is Yaldabaoth who is depicted in Cult-dictated popular culture as the 'Dark Lord', 'Lord of Time', Lord (Darth) Vader and Dormammu, the evil ruler of the 'Dark Dimension' trying to take over the 'Earth Dimension' in the Marvel comic movie, *Dr Strange*. Yaldabaoth is both the Old Testament 'god' and the Biblical 'Satan'. Gnostics referred to Yaldabaoth as the 'Great Architect of the Universe'and the Cult-controlled Freemason network calls their god 'the 'Great Architect of the Universe' (also Grand Architect). The 'Great Architect' Yaldabaoth is symbolised by the Cult as the all-seeing eye at the top of the pyramid on the Great Seal of the United States and the dollar bill. Archon is encoded in *arch*-itect as it is in *arch*-angels and *arch*-bishops. All religions have the theme of a force for good and force for evil in some sort of spiritual war and there is a reason for that – the theme is true. The Cult and its non-human masters are quite happy for this to circulate. They present themselves as the force for good fighting evil when they are really the force of evil (absence of love). The whole foundation of Cult modus operandi is inversion. They promote themselves as a force for good and anyone challenging them in pursuit of peace, love, fairness, truth and justice is condemned as a satanic force for evil. This has been the game plan throughout history whether the Church of Rome inquisitions of non-believers or 'conspiracy theorists' and 'anti-vaxxers' of today. The technique is the same whatever the timeline era.

Yaldabaoth is revolting (true)

Yaldabaoth and the Archons are said to have revolted against God with Yaldabaoth claiming to *be* God – the *All That Is*. The Old Testament 'God' (Yaldabaoth) demanded to be worshipped as such: '*I am* the LORD, and there is none else, there is no God beside me' (Isaiah 45:5). I have quoted in other books a man who said he was the unofficial son of the late Baron Philippe de Rothschild of the Mouton-Rothschild wine producing estates in France who died in 1988 and he told me about the Rothschild 'revolt from God'. The man said he was given the name Phillip Eugene de Rothschild and we shared long correspondence many years ago while he was living under another identity. He said that he was conceived through 'occult incest' which (within the Cult) was 'normal and to be admired'. 'Phillip' told me about his experience attending satanic rituals with rich and famous people whom he names and you can see them and the wider background to Cult Satanism in my other books starting with *The Biggest Secret*. Cult rituals are interactions with Archontic 'gods'. 'Phillip' described Baron Philippe de Rothschild as 'a master Satanist and hater of God' and he used the same term 'revolt from God' associated with Yaldabaoth/Satan/Lucifer/the Devil in describing the Sabbatian Rothschild dynasty. 'I played a key role in my family's revolt from God', he said. That role was to infiltrate in classic Sabbatian style the Christian Church, but eventually he escaped the mind-prison to live another life. The Cult has been targeting religion in a plan to make worship of the Archons the global one-world religion. Infiltration of Satanism into modern 'culture', especially among the young, through music videos, stage shows and other means, is all part of this.

Nag Hammadi texts describe Yaldabaoth and the Archons in their prime form as energy – consciousness – and say they can take form if they choose in the same way that consciousness takes form as a human. Yaldabaoth is called 'formless' and represents a deeply inverted, distorted and chaotic state of consciousness which seeks to attached to humans and turn them into a likeness of itself in an attempt at assimilation. For that to happen it has to manipulate humans into low frequency mental and emotional states that match its own. Archons can certainly appear in human form and this is the origin of the psychopathic personality. The energetic distortion Gnostics called Yaldabaoth *is* psychopathy. When psychopathic Archons take human form that human will be a psychopath as an expression of Yaldabaoth consciousness. Cult psychopaths are Archons in human form. The principle is the same as that portrayed in the 2009 *Avatar* movie when the American military travelled to a fictional Earth-like moon called Pandora in the Alpha Centauri star system to infiltrate a society of blue people, or Na'vi, by hiding within bodies that looked like the Na'vi. Archons posing as humans have a particular hybrid information field, part human, part Archon, (the ancient 'demigods') which processes information in a way that manifests behaviour to match their psychopathic evil, lack of empathy and compassion, and stops them being influenced by the empathy, compassion and love that a fully-human information field is capable of expressing. Cult bloodlines interbreed, be they royalty or dark suits, for this reason and you have their obsession with incest. Interbreeding with full-blown humans would dilute the Archontic energy field that guarantees psychopathy in its representatives in the human realm.

Gnostic writings say the main non-human forms that Archons take are *serpentine* (what I have called for decades 'reptilian' amid unbounded ridicule from the Archontically-programmed) and what Gnostics describe as 'an unborn baby or foetus with grey skin and dark, unmoving eyes'. This is an excellent representation of the ET 'Greys' of UFO folklore which large numbers of people claim to have seen and been abducted by – Zulu shaman Credo Mutwa among them. I agree with those that believe in extraterrestrial or interdimensional visitations today and for thousands of years past. No wonder with their advanced knowledge and technological capability they were perceived and worshipped as gods for technological and other 'miracles' they appeared to perform. Imagine someone arriving in a culture disconnected from the modern world with a smartphone and computer. They would be seen as a ‘god’ capable of ‘miracles’. The Renegade Mind, however, wants to know the source of everything and not only the way that source manifests as human or non-human. In the same way that a Renegade Mind seeks the original source material for the ‘Covid virus’ to see if what is claimed is true. The original source of Archons in form is consciousness – the distorted state of consciousness known to Gnostics as Yaldabaoth.

‘Revolt from God’ is energetic disconnection

Where I am going next will make a lot of sense of religious texts and ancient legends relating to ‘Satan’, Lucifer’ and the ‘gods’. Gnostic descriptions sync perfectly with the themes of my own research over the years in how they describe a consciousness distortion seeking to impose itself on human consciousness. I’ve referred to the core of infinite awareness in previous books as Infinite Awareness in Awareness of Itself. By that I mean a level of awareness that knows that it is all awareness and is aware of all awareness. From here comes the frequency of love in its true sense and balance which is what love is on one level – the balance of all forces into a single whole called Oneness and Isness. The more we disconnect from this state of love that many call ‘God’ the constituent parts of that Oneness start to unravel and express themselves as a part and not a whole. They become individualised as intellect, mind, selfishness, hatred, envy, desire for power over others, and such like. This is not a problem in the greater scheme in that ‘God’, the *All That Is*, can experience all these possibilities through different expressions of itself including humans. What we as expressions of the whole experience the *All That Is* experiences. We are the *All That Is* experiencing itself. As we withdraw from that state of Oneness we disconnect from its influence and things can get very unpleasant and very stupid. Archontic consciousness is at the extreme end of that. It has so disconnected from the influence of Oneness that it has become an inversion of unity and love, an inversion of everything, an inversion of life itself. Evil is appropriately live written backwards. Archontic consciousness is obsessed with death, an inversion of life, and so its manifestations in Satanism are obsessed with death. They use inverted symbols in their rituals such as the inverted pentagram and cross. Sabbatians as Archontic consciousness incarnate invert Judaism and every other religion and culture they infiltrate. They seek disunity and chaos and they fear unity and harmony as they fear love like garlic to a vampire. As a result the Cult, Archons incarnate, act with such evil, psychopathy and lack of empathy and compassion disconnected as they are from the source of love. How could Bill Gates and the rest of the Archontic psychopaths do what they have to human society in the 'Covid' era with all the death, suffering and destruction involved and have no emotional consequence for the impact on others? Now you know. Why have Zuckerberg, Brin, Page, Wojcicki and company callously censored information warning about the dangers of the 'vaccine' while thousands have been dying and having severe, sometimes life-changing reactions? Now you know. Why have Tedros, Fauci, Whitty, Vallance and their like around the world been using case and death figures they're aware are fraudulent to justify lockdowns and all the deaths and destroyed lives that have come from that? Now you know. Why did Christian Drosten produce and promote a 'testing' protocol that he knew couldn't test for infectious disease which led to a global human catastrophe. Now you know. The Archontic mind doesn't give a shit (Fig 17). I personally think that Gates and major Cult insiders are a form of AI cyborg that the Archons want humans to become.

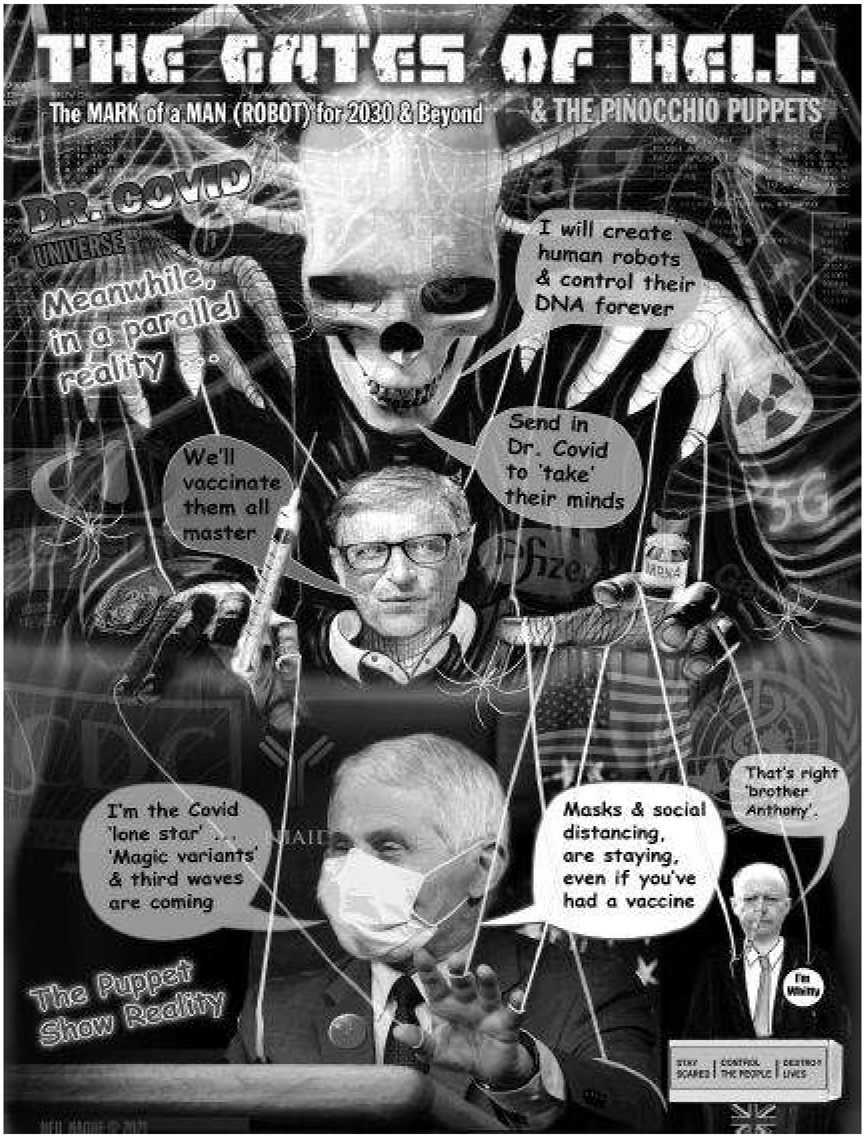

THE GATES OF HELL
The MARK of a MAN (ROBOT) for 2030 & Beyond
& THE PINOCCHIO PUPPETS
DR. COVID
UNIVERSE
Meanwhile, in a parallel reality . . .
I will create human robots & control their DNA forever
We'll vaccinate them all master
Send in Dr. Covid to 'take' their minds
5G
Pfizer
MRNA
CDC
NIAID
I'm the Covid 'lone star' . . . 'Magic variants' & third waves are coming
Masks & social distancing, are staying, even if you've had a vaccine
That's right 'brother Anthony'.
I'm Whitty
STAY SCARED | CONTROL THE PEOPLE | DESTROY LIVES
The Puppet Show Reality
NEIL HAGUE © 2021

Figure 17: Artist Neil Hague's version of the 'Covid' hierarchy.

Human batteries

A state of such inversion does have its consequences, however. The level of disconnection from the Source of All means that you withdraw from that source of energetic sustenance and creativity. This means that you have to find your own supply of energetic power and it has – *us*. When the Morpheus character in the first *Matrix* movie held up a battery he spoke a profound truth when he said: 'The Matrix is a computer-generated dream world built to keep us under control in order to change the human being into one of these.' The statement was true in all respects. We do live in a technologically-generated virtual reality simulation (more very shortly) and we have been manipulated to be an energy source for Archontic consciousness. The Disney-Pixar animated movie *Monsters, Inc*. in 2001 symbolised the dynamic when monsters in their world had no energy source and they would enter the human world to terrify children in their beds, catch the child's scream, terror (low-vibrational frequencies), and take that energy back to power the monster world. The lead character you might remember was a single giant eye and the symbolism of the Cult's all-seeing eye was obvious. Every thought and emotion is broadcast as a frequency unique to that thought and emotion. Feelings of love and joy, empathy and compassion, are high, quick, frequencies while fear, depression, anxiety, suffering and hate are low, slow, dense frequencies. Which kind do you think Archontic consciousness can connect with and absorb? In such a low and dense frequency state there's no way it can connect with the energy of love and joy. Archons can only feed off energy compatible with their own frequency and they and their Cult agents want to delete the human world of love and joy and manipulate the transmission of low vibrational frequencies through low-vibrational human mental and emotional states. *We are their energy source*. Wars are energetic banquets to the Archons – a world war even more so – and think how much low-frequency mental and emotional energy has been generated from the consequences for humanity of the 'Covid' hoax orchestrated by Archons incarnate like Gates.

The ancient practice of human sacrifice 'to the gods', continued in secret today by the Cult, is based on the same principle. 'The gods' are Archontic consciousness in different forms and the sacrifice is induced into a state of intense terror to generate the energy the Archontic frequency can absorb. Incarnate Archons in the ritual drink the blood which contains an adrenaline they crave which floods into the bloodstream when people are terrorised. Most of the sacrifices, ancient and modern, are children and the theme of 'sacrificing young virgins to the gods' is just code for children. They have a particular pre-puberty energy that Archons want more than anything and the energy of the young in general is their target. The California Department of Education wants students to chant the names of Aztec gods (Archontic gods) once worshipped in human sacrifice rituals in a curriculum designed to encourage them to ‘challenge racist, bigoted, discriminatory, imperialist/colonial beliefs’, join ‘social movements that struggle for social justice’, and ‘build new possibilities for a post-racist, post-systemic racism society’. It’s the usual Woke crap that inverts racism and calls it anti-racism. In this case solidarity with ‘indigenous tribes’ is being used as an excuse to chant the names of ‘gods’ to which people were sacrificed (and still are in secret). What an example of Woke’s inability to see beyond black and white, us and them, They condemn the colonisation of these tribal cultures by Europeans (quite right), but those cultures sacrificing people including children to their ‘gods’, and mass murdering untold numbers as the Aztecs did, is just fine. One chant is to the Aztec god Tezcatlipoca who had a man sacrificed to him in the 5th month of the Aztec calendar. His heart was cut out and he was eaten. Oh, that’s okay then. Come on children … after three … Other sacrificial ‘gods’ for the young to chant their allegiance include Quetzalcoatl, Huitzilopochtli and Xipe Totec. The curriculum says that ‘chants, affirmations, and energizers can be used to bring the class together, build unity around ethnic studies principles and values, and to reinvigorate the class following a lesson that may be emotionally taxing or even when student engagement may appear to be low’. Well, that’s the cover story, anyway. Chanting and mantras are the repetition of a particular frequency generated from the vocal cords and chanting the names of these Archontic ‘gods’ tunes you into their frequency. That is the last thing you want when it allows for energetic synchronisation, attachment and perceptual influence. Initiates chant the names of their ‘Gods’ in their rituals for this very reason.

Vampires of the Woke

Paedophilia is another way that Archons absorb the energy of children. Paedophiles possessed by Archontic consciousness are used as the conduit during sexual abuse for discarnate Archons to vampire the energy of the young they desire so much. Stupendous numbers of children disappear every year never to be seen again although you would never know from the media. Imagine how much low-vibrational energy has been generated by children during the 'Covid' hoax when so many have become depressed and psychologically destroyed to the point of killing themselves. Shocking numbers of children are now taken by the state from loving parents to be handed to others. I can tell you from long experience of researching this since 1996 that many end up with paedophiles and assets of the Cult through corrupt and Cult-owned social services which in the reframing era has hired many psychopaths and emotionless automatons to do the job. Children are even stolen to order using spurious reasons to take them by the corrupt and secret (because they're corrupt) 'family courts'. I have written in detail in other books, starting with *The Biggest Secret* in 1997, about the ubiquitous connections between the political, corporate, government, intelligence and military elites (Cult operatives) and Satanism and paedophilia. If you go deep enough both networks have an interlocking leadership. The Woke mentality has been developed by the Cult for many reasons: To promote almost every aspect of its agenda; to hijack the traditional political left and turn it fascist; to divide and rule; and to target agenda pushbackers. But there are other reasons which relate to what I am describing here. How many happy and joyful Wokers do you ever see especially at the extreme end? They are a mental and psychological mess consumed by emotional stress and constantly emotionally cocked for the next explosion of indignation at someone referring to a female as a female. They are walking, talking, batteries as Morpheus might say emitting frequencies which both enslave them in low-vibrational bubbles of perceptual limitation and feed the Archons. Add to this the hatred claimed to be love; fascism claimed to 'anti-fascism', racism claimed to be 'anti-racism';

exclusion claimed to inclusion; and the abuse-filled Internet trolling. You have a purpose-built Archontic energy system with not a wind turbine in sight and all founded on Archontic *inversion*. We have whole generations now manipulated to serve the Archons with their actions and energy. They will be doing so their entire adult lives unless they snap out of their Archon-induced trance. Is it really a surprise that Cult billionaires and corporations put so much money their way? Where is the energy of joy and laughter, including laughing at yourself which is confirmation of your own emotional security? Mark Twain said: ‘The human race has one really effective weapon, and that is laughter.’ We must use it all the time. Woke has destroyed comedy because it has no humour, no joy, sense of irony, or self-deprecation. Its energy is dense and intense. *Mmmmm*, lunch says the Archontic frequency. Rudolf Steiner (1861-1925) was the Austrian philosopher and famous esoteric thinker who established Waldorf education or Steiner schools to treat children like unique expressions of consciousness and not minds to be programmed with the perceptions determined by authority. I’d been writing about this energy vampiring for decades when I was sent in 2016 a quote by Steiner. He was spot on:

> There are beings in the spiritual realms for whom anxiety and fear emanating from human beings offer welcome food. When humans have no anxiety and fear, then these creatures starve. If fear and anxiety radiates from people and they break out in panic, then these creatures find welcome nutrition and they become more and more powerful. These beings are hostile towards humanity. Everything that feeds on negative feelings, on anxiety, fear and superstition, despair or doubt, are in reality hostile forces in super-sensible worlds, launching cruel attacks on human beings, while they are being fed ... These are exactly the feelings that belong to contemporary culture and materialism; because it estranges people from the spiritual world, it is especially suited to evoke hopelessness and fear of the unknown in people, thereby calling up the above mentioned hostile forces against them.

Pause for a moment from this perspective and reflect on what has happened in the world since the start of 2020. Not only will pennies drop, but billion dollar bills. We see the same theme from Don Juan Matus, a Yaqui Indian shaman in Mexico and the information source for Peruvian-born writer, Carlos Castaneda, who wrote a series of books from the 1960s to 1990s. Don Juan described the force manipulating human society and his name for the Archons was the predator:

> We have a predator that came from the depths of the cosmos and took over the rule of our lives. Human beings are its prisoners. The predator is our lord and master. It has rendered us docile, helpless. If we want to protest, it suppresses our protest. If we want to act independently, it demands that we don't do so … indeed we are held prisoner!
>
> They took us over because we are food to them, and they squeeze us mercilessly because we are their sustenance. Just as we rear chickens in coops, the predators rear us in human coops, humaneros. Therefore, their food is always available to them.

Different cultures, different eras, same recurring theme.

The 'ennoia' dilemma

Nag Hammadi Gnostic manuscripts say that Archon consciousness has no 'ennoia'. This is directly translated as 'intentionality', but I'll use the term 'creative imagination'. The *All That Is* in awareness of itself is the source of all creativity – all possibility – and the more disconnected you are from that source the more you are subsequently denied 'creative imagination'. Given that Archon consciousness is almost entirely disconnected it severely lacks creativity and has to rely on far more mechanical processes of thought and exploit the creative potential of those that do have 'ennoia'. You can see cases of this throughout human society. Archon consciousness almost entirely dominates the global banking system and if we study how that system works you will appreciate what I mean. Banks manifest 'money' out of nothing by issuing lines of 'credit' which is 'money' that has never, does not, and will never exist except in theory. It's a confidence trick. If you think 'credit' figures-on-a-screen 'money' is worth anything you accept it as payment. If you don't then the whole system collapses through lack of confidence in the value of that 'money'. Archontic bankers with no 'ennoia' are 'lending' 'money' that doesn't exist to humans that *do* have creativity – those that have the inspired ideas and create businesses and products. Archon banking feeds off human creativity which it controls through 'money' creation and debt. Humans have the creativity and Archons exploit that for their own benefit and control while having none themselves. Archon Internet platforms like Facebook claim joint copyright of everything that creative users post and while Archontic minds like Zuckerberg may officially head that company it will be human creatives on the staff that provide the creative inspiration. When you have limitless 'money' you can then buy other companies established by creative humans. Witness the acquisition record of Facebook, Google and their like. Survey the Archon-controlled music industry and you see non-creative dark suit executives making their fortune from the human creativity of their artists. The cases are endless. Research the history of people like Gates and Zuckerberg and how their empires were built on exploiting the creativity of others. Archon minds cannot create out of nothing, but they are skilled (because they have to be) in what Gnostic texts call 'countermimicry'. They can imitate, but not innovate. Sabbatians trawl the creativity of others through backdoors they install in computer systems through their cybersecurity systems. Archon-controlled China is globally infamous for stealing intellectual property and I remember how Hong Kong, now part of China, became notorious for making counterfeit copies of the creativity of others – 'countermimicry'. With the now pervasive and all-seeing surveillance systems able to infiltrate any computer you can appreciate the potential for Archons to vampire the creativity of humans. Author John Lamb Lash wrote in his book about the Nag Hammadi texts, *Not In His Image*:

> Although they cannot originate anything, because they lack the divine factor of ennoia (intentionality), Archons can imitate with a vengeance. Their expertise is simulation (HAL, virtual reality). The Demiurge [Yaldabaoth] fashions a heaven world copied from the fractal patterns [of the original] ... His construction is celestial kitsch, like the fake Italianate villa of a Mafia don complete with militant angels to guard every portal.

This brings us to something that I have been speaking about since the turn of the millennium. Our reality is a simulation; a virtual reality that we think is real. No, I'm not kidding.

Human reality? Well, virtually

I had pondered for years about whether our reality is 'real' or some kind of construct. I remembered being immensely affected on a visit as a small child in the late 1950s to the then newly-opened Planetarium on the Marylebone Road in London which is now closed and part of the adjacent Madame Tussauds wax museum. It was in the middle of the day, but when the lights went out there was the night sky projected in the Planetarium's domed ceiling and it appeared to be so real. The experience never left me and I didn't know why until around the turn of the millennium when I became certain that our 'night sky' and entire reality is a projection, a virtual reality, akin to the illusory world portrayed in the *Matrix* movies. I looked at the sky one day in this period and it appeared to me like the domed roof of the Planetarium. The release of the first *Matrix* movie in 1999 also provided a synchronistic and perfect visual representation of where my mind had been going for a long time. I hadn't come across the Gnostic Nag Hammadi texts then. When I did years later the correlation was once again astounding. As I read Gnostic accounts from 1,600 years and more earlier it was clear that they were describing the same simulation phenomenon. They tell how the Yaldabaoth 'Demiurge' and Archons created a 'bad copy' of original reality to rule over all that were captured by its illusions and the body was a prison to trap consciousness in the 'bad copy' fake reality. Read how Gnostics describe the 'bad copy' and update that to current times and they are referring to what we would call today a virtual reality simulation.

Author John Lamb Lash said 'the Demiurge fashions a heaven world copied from the fractal patterns' of the original through expertise in 'HAL' or virtual reality simulation. Fractal patterns are part of the energetic information construct of our reality, a sort of blueprint. If these patterns were copied in computer terms it would indeed give you a copy of a 'natural' reality in a non-natural frequency and digital form. The principle is the same as making a copy of a website. The original website still exists, but now you can change the copy version to make it whatever you like and it can become very different to the original website. Archons have done this with our reality, a *synthetic* copy of prime reality that still exists beyond the frequency walls of the simulation. Trapped within the illusions of this synthetic Matrix, however, were and are human consciousness and other expressions of prime reality and this is why the Archons via the Cult are seeking to make the human body synthetic and give us synthetic AI minds to complete the job of turning the entire reality synthetic including what we perceive to be the natural world. To quote Kurzweil: 'Nanobots will infuse all the matter around us with information. Rocks, trees, everything will become these intelligent creatures.' Yes, *synthetic* 'creatures' just as 'Covid' and other genetically-manipulating 'vaccines' are designed to make the human body synthetic. From this perspective it is obvious why Archons and their Cult are so desperate to infuse synthetic material into every human with their 'Covid' scam.

Let there be (electromagnetic) light

Yaldabaoth, the force that created the simulation, or Matrix, makes sense of the Gnostic reference to 'The Great Architect' and its use by Cult Freemasonry as the name of its deity. The designer of the Matrix in the movies is called 'The Architect' and that trilogy is jam-packed with symbolism relating to these subjects. I have contended for years that the angry Old Testament God (Yaldabaoth) is the 'God' being symbolically 'quoted' in the opening of Genesis as 'creating the world'. This is not the creation of prime reality – it's the creation of the *simulation*. The Genesis 'God' says: 'Let there be Light: and there was light.' But what is this 'Light'? I have said for decades that the speed of light (186,000 miles per second) is not the fastest speed possible as claimed by mainstream science and is in fact the frequency walls or outer limits of the Matrix. You can't have a fastest or slowest anything within all possibility when everything is possible. The human body is encoded to operate within the speed of light or *within the simulation* and thus we see only the tiny frequency band of visible *light*. Near-death experiencers who perceive reality outside the body during temporary 'death' describe a very different form of light and this is supported by the Nag Hammadi texts. Prime reality beyond the simulation ('Upper Aeons' to the Gnostics) is described as a realm of incredible beauty, bliss, love and harmony – a realm of 'watery light' that is so powerful 'there are no shadows'. Our false reality of Archon control, which Gnostics call the 'Lower Aeons', is depicted as a realm with a different kind of 'light' and described in terms of chaos, 'Hell', 'the Abyss' and 'Outer Darkness', where trapped souls are tormented and manipulated by demons (relate that to the 'Covid' hoax alone). The watery light theme can be found in near-death accounts and it is not the same as *simulation* 'light' which is electromagnetic or radiation light within the speed of light – the 'Lower Aeons'. Simulation 'light' is the 'luminous fire' associated by Gnostics with the Archons. The Bible refers to Yaldabaoth as 'that old serpent, called the Devil, and Satan, which deceiveth the whole world' (Revelation 12:9). I think that making a simulated copy of prime reality ('countermimicry') and changing it dramatically while all the time manipulating humanity to believe it to be real could probably meet the criteria of deceiving the whole world. Then we come to the Cult god Lucifer – the *Light Bringer*. Lucifer is symbolic of Yaldabaoth, the bringer of radiation light that forms the bad copy simulation within the speed of light. 'He' is symbolised by the lighted torch held by the Statue of Liberty and in the name 'Illuminati'. Sabbatian-Frankism declares that Lucifer is the true god and Lucifer is the real god of Freemasonry honoured as their 'Great or Grand Architect of the Universe' (simulation).

I would emphasise, too, the way Archontic technologically-generated luminous fire of radiation has deluged our environment since I was a kid in the 1950s and changed the nature of The Field with which we constantly interact. Through that interaction technological radiation is changing us. The Smart Grid is designed to operate with immense levels of communication power with 5G expanding across the world and 6G, 7G, in the process of development. Radiation is the simulation and the Archontic manipulation system. Why wouldn't the Archon Cult wish to unleash radiation upon us to an ever-greater extreme to form

Kurzweil's 'cloud'? The plan for a synthetic human is related to the need to cope with levels of radiation beyond even anything we've seen so far. Biological humans would not survive the scale of radiation they have in their script. The Smart Grid is a technological sub-reality within the technological simulation to further disconnect five-sense perception from expanded consciousness. It's a technological prison of the mind.

Infusing the 'spirit of darkness'

A recurring theme in religion and native cultures is the manipulation of human genetics by a non-human force and most famously recorded as the biblical 'sons of god' (the god*s* plural in the original) who interbred with the daughters of men. The Nag Hammadi *Apocryphon of John* tells the same story this way:

> He [Yaldabaoth] sent his angels [Archons/demons] to the daughters of men, that they might take some of them for themselves and raise offspring for their enjoyment. And at first they did not succeed. When they had no success, they gathered together again and they made a plan together ... And the angels changed themselves in their likeness into the likeness of their mates, filling them with the spirit of darkness, which they had mixed for them, and with evil ... And they took women and begot children out of the darkness according to the likeness of their spirit.

Possession when a discarnate entity takes over a human body is an age-old theme and continues today. It's very real and I've seen it. Satanic and secret society rituals can create an energetic environment in which entities can attach to initiates and I've heard many stories of how people have changed their personality after being initiated even into lower levels of the Freemasons. I have been inside three Freemasonic temples, one at a public open day and two by just walking in when there was no one around to stop me. They were in Ryde, the town where I live, Birmingham, England, when I was with a group, and Boston, Massachusetts. They all felt the same energetically – dark, dense, low-vibrational and sinister. Demonic attachment can happen while the initiate has no idea what is going on. To them it's just a ritual to get in the Masons and do a bit of good business. In the far more extreme rituals of Satanism human possession is even more powerful and they are designed to make possession possible. The hierarchy of the Cult is dictated by the power and perceived status of the possessing Archon. In this way the Archon hierarchy becomes the Cult hierarchy. Once the entity has attached it can influence perception and behaviour and if it attaches to the extreme then so much of its energy (information) infuses into the body information field that the hologram starts to reflect the nature of the possessing entity. This is the *Exorcist* movie type of possession when facial features change and it's known as shapeshifting. Islam's Jinn are said to be invisible tricksters who change shape, 'whisper', confuse and take human form. These are all traits of the Archons and other versions of the same phenomenon. Extreme possession could certainty infuse the 'spirit of darkness' into a partner during sex as the Nag Hammadi texts appear to describe. Such an infusion can change genetics which is also energetic information. Human genetics is information and the 'spirit of darkness' is information. Mix one with the other and change must happen. Islam has the concept of a 'Jinn baby' through possession of the mother and by Jinn taking human form. There are many ways that human genetics can be changed and remember that Archons have been aware all along of advanced techniques to do this. What is being done in human society today – and far more – was known about by Archons at the time of the 'fallen ones' and their other versions described in religions and cultures.

Archons and their human-world Cult are obsessed with genetics as we see today and they know this dictates how information is processed into perceived reality during a human life. They needed to produce a human form that would decode the simulation and this is symbolically known as 'Adam and Eve' who left the 'garden' (prime reality) and 'fell' into Matrix reality. The simulation is not a 'physical' construct (there is no 'physical'); it is a source of information. Think Wi-Fi again. The simulation is an energetic field encoded with information and body-brain systems are designed to decode that information encoded in wave or frequency form which is transmitted to the brain as electrical signals. These are decoded by the brain to construct our sense of reality – an illusory 'physical' world that only exists in the brain or the mind. Virtual reality games mimic this process using the same sensory decoding system. Information is fed to the senses to decode a virtual reality that can appear so real, but isn't (Figs 18 and 19). Some scientists believe – and I agree with them – that what we perceive as 'physical' reality only exists when we are looking or observing. The act of perception or focus triggers the decoding systems which turn waveform information into holographic reality. When we are not observing something our reality reverts from a holographic state to a waveform state. This relates to the same principle as a falling tree not making a noise unless someone is there to hear it or decode it. The concept makes sense from the simulation perspective. A computer is not decoding all the information in a Wi-Fi field all the time and only decodes or brings into reality on the screen that part of Wi-Fi that it's decoding – focusing upon – at that moment.

Figure 18: Virtual reality technology 'hacks' into the body's five-sense decoding system.

Figure 19: The result can be experienced as very 'real'.

Interestingly, Professor Donald Hoffman at the Department of Cognitive Sciences at the University of California, Irvine, says that our experienced reality is like a computer interface that shows us only the level with which we interact while hiding all that exists beyond it: 'Evolution shaped us with a user interface that hides the truth. Nothing that we see is the truth – the very language of space and time and objects is the wrong language to describe reality.' He is correct in what he says on so many levels. Space and time are not a universal reality. They are a phenomenon of decoded *simulation* reality as part of the process of enslaving our sense of reality. Near-death experiencers report again and again how space and time did not exist as we perceive them once they were free of the body – body decoding systems. You can appreciate from this why Archons and their Cult are so desperate to entrap human attention in the five senses where we are in the Matrix and of the Matrix. Opening your mind to expanded states of awareness takes you beyond the information confines of the simulation and you become aware of knowledge and insights denied to you before. This is what we call 'awakening' – *awakening from the Matrix* – and in the final chapter I will relate this to current events.

Where are the 'aliens'?

A simulation would explain the so-called 'Fermi Paradox' named after Italian physicist Enrico Fermi (1901-1954) who created the first nuclear reactor. He considered the question of why there is such a lack of extraterrestrial activity when there are so many stars and planets in an apparently vast universe; but what if the night sky that we see, or think we do, is a simulated projection as I say? If you control the simulation and your aim is to hold humanity fast in essential ignorance would you want other forms of life including advanced life coming and going sharing information with humanity? Or would you want them to believe they were isolated and apparently alone? Themes of human isolation and apartness are common whether they be the perception of a lifeless universe or the fascist isolation laws of the 'Covid' era. Paradoxically the very existence of a simulation means that we are not alone when some force had to construct it. My view is that experiences that people have reported all over the world for centuries with Reptilians and Grey entities are Archon phenomena as Nag Hammadi texts describe; and that benevolent 'alien' interactions are non-human groups that come in and out of the simulation by overcoming Archon attempts to keep them out. It should be highlighted, too, that Reptilians and Greys are obsessed with *genetics* and *technology* as related by cultural accounts and those who say they have been abducted by them. Technology is their way of overcoming some of the limitations in their creative potential and our technology-driven and controlled human society of today is *arch*etypical Archon-Reptilian-Grey modus operandi. Technocracy is really *Archon*tocracy. The Universe does not have to be as big as it appears with a simulation. There is no space or distance only information decoded into holographic reality. What we call 'space' is only the absence of holographic 'objects' and that 'space' is The Field of energetic information which connects everything into a single whole. The same applies with the artificially-generated information field of the simulation. The Universe is not big or small as a physical reality. It is decoded information, that's all, and its perceived size is decided by the way the simulation is encoded to make it appear. The entire night sky as we perceive it only exists in our brain and so where are those 'millions of light years'? The 'stars' on the ceiling of the Planetarium looked a vast distance away.

There's another point to mention about 'aliens'. I have been highlighting since the 1990s the plan to stage a fake 'alien invasion' to justify the centralisation of global power and a world military. Nazi scientist Werner von Braun, who was taken to America by Operation Paperclip after World War Two to help found NASA, told his American assistant Dr Carol Rosin about the Cult agenda when he knew he was dying in 1977. Rosin said that he told her about a sequence that would lead to total human control by a one-world government. This included threats from terrorism, rogue nations, meteors and asteroids before finally an 'alien invasion'. All of these things, von Braun said, would be bogus and what I would refer to as a No-Problem-Reaction-Solution. Keep this in mind when 'the aliens are coming' is the new mantra. The aliens are not coming – they are *already here* and they have infiltrated human society while looking human. French-Canadian investigative journalist Serge Monast said in 1994 that he had uncovered a NASA/military operation called Project Blue Beam which fits with what Werner von Braun predicted. Monast died of a 'heart attack' in 1996 the day after he was arrested and spent a night in prison. He was 51. He said Blue Beam was a plan to stage an alien invasion that would include religious figures beamed holographically into the sky as part of a global manipulation to usher in a 'new age' of worshipping what I would say is the Cult 'god' Yaldabaoth in a one-world religion. Fake holographic asteroids are also said to be part of the plan which again syncs with von Braun. How could you stage an illusory threat from asteroids unless they were holographic inserts? This is pretty straightforward given the advanced technology outside the public arena and the fact that our 'physical' reality is holographic anyway. Information fields would be projected and we would decode them into the illusion of a 'physical' asteroid. If they can sell a global 'pandemic' with a 'virus' that doesn't exist what will humans not believe if government and media tell them?

All this is particularly relevant as I write with the Pentagon planning to release in June, 2021, information about 'UFO sightings'. I have been following the UFO story since the early 1990s and the common theme throughout has been government and military denials and cover up. More recently, however, the Pentagon has suddenly become more talkative and apparently open with Air Force pilot radar images released of unexplained craft moving and changing direction at speeds well beyond anything believed possible with human technology. Then, in March, 2021, former Director of National Intelligence John Ratcliffe said a Pentagon report months later in June would reveal a great deal of information about UFO sightings unknown to the public. He said the report would have 'massive implications'. The order to do this was included bizarrely in a $2.3 trillion 'coronavirus' relief and government funding bill passed by the Trump administration at the end of 2020. I would add some serious notes of caution here. I have been pointing out since the 1990s that the US military and intelligence networks have long had craft – 'flying saucers' or anti-gravity craft – which any observer would take to be extraterrestrial in origin. Keeping this knowledge from the public allows craft flown by *humans* to be perceived as alien visitations. I am not saying that 'aliens' do not exist. I would be the last one to say that, but we have to be streetwise here. President Ronald Reagan told the UN General Assembly in 1987: 'I occasionally think how quickly our differences worldwide would vanish if we were facing an alien threat from outside this world.' That's the idea. Unite against a common 'enemy' with a common purpose behind your 'saviour force' (the Cult) as this age-old technique of mass manipulation goes global.

Science moves this way ...

I could find only one other person who was discussing the simulation hypothesis publicly when I concluded it was real. This was Nick Bostrom, a Swedish-born philosopher at the University of Oxford, who has explored for many years the possibility that human reality is a computer simulation although his version and mine are not the same. Today the simulation and holographic reality hypothesis have increasingly entered the scientific mainstream. Well, the more open-minded mainstream, that is. Here are a few of the ever-gathering examples. American nuclear physicist Silas Beane led a team of physicists at the University of Bonn in Germany pursuing the question of whether we live in a simulation. They concluded that we probably do and it was likely based on a lattice of cubes. They found that cosmic rays align with that specific pattern. The team highlighted the Greisen–Zatsepin–Kuzmin (GZK) limit which refers to cosmic ray particle interaction with cosmic background radiation that creates an apparent boundary for cosmic ray particles. They say in a paper entitled 'Constraints on the Universe as a Numerical Simulation' that this 'pattern of constraint' is exactly what you would find with a computer simulation. They also made the point that a simulation would create its own 'laws of physics' that would limit possibility. I've been making the same point for decades that the *perceived* laws of physics relate only to this reality, or what I would later call the simulation. When designers write codes to create computer and virtual reality games they are the equivalent of the laws of physics for that game. Players interact within the limitations laid out by the coding. In the same way those who wrote the codes for the simulation decided the laws of physics that would apply. These can be overridden by expanded states of consciousness, but not by those enslaved in only five-sense awareness where simulation codes rule. Overriding the codes is what people call 'miracles'. They are not. They are bypassing the encoded limits of the simulation. A population caught in simulation perception would have no idea that this was their plight. As the Bonn paper said: 'Like a prisoner in a pitch-black cell we would not be able to see the "walls" of our prison,' That's true if people remain mesmerised by the five senses. Open to expanded awareness and those walls become very clear. The main one is the speed of light.

American theoretical physicist James Gates is another who has explored the simulation question and found considerable evidence to support the idea. Gates was Professor of Physics at the University of Maryland, Director of The Center for String and Particle Theory, and on Barack Obama's Council of Advisors on Science and Technology. He and his team found *computer codes* of digital data embedded in the fabric of our reality. They relate to on-off electrical charges of 1 and 0 in the binary system used by computers. 'We have no idea what they are doing there', Gates said. They found within the energetic fabric mathematical sequences known as error-correcting codes or block codes that 'reboot' data to its original state or 'default settings' when something knocks it out of sync. Gates was asked if he had found a set of equations embedded in our reality indistinguishable from those that drive search engines and browsers and he said: 'That is correct.' Rich Terrile, director of the Centre for Evolutionary Computation and Automated Design at NASA's Jet

Propulsion Laboratory, has said publicly that he believes the Universe is a digital hologram that must have been created by a form of intelligence. I agree with that in every way. Waveform information is delivered electrically by the senses to the brain which constructs a *digital* holographic reality that we call the ‘world’. This digital level of reality can be read by the esoteric art of numerology. Digital holograms are at the cutting edge of holographics today. We have digital technology everywhere designed to access and manipulate our digital level of perceived reality. Synthetic mRNA in ‘Covid vaccines’ has a digital component to manipulate the body’s digital ‘operating system’.

Reality is numbers

How many know that our reality can be broken down to numbers and codes that are the same as computer games? Max Tegmark, a physicist at the Massachusetts Institute of Technology (MIT), is the author of *Our Mathematical Universe* in which he lays out how reality can be entirely described by numbers and maths in the way that a video game is encoded with the ‘physics’ of computer games. Our world and computer virtual reality are essentially the same. Tegmark imagines the perceptions of characters in an advanced computer game when the graphics are so good they don’t know they are in a game. They think they can bump into real objects (electromagnetic resistance in our reality), fall in love and feel emotions like excitement. When they began to study the apparently ‘physical world’ of the video game they would realise that everything was made of pixels (which have been found in our energetic reality as must be the case when on one level our world is digital). What computer game characters thought was physical ‘stuff’, Tegmark said, could actually be broken down into numbers:

> And we’re exactly in this situation in our world. We look around and it doesn’t seem that mathematical at all, but everything we see is made out of elementary particles like quarks and electrons. And what properties does an electron have? Does it have a smell or a colour or a texture? No! ... We physicists have come up with geeky names for [Electron] properties, like

electric charge, or spin, or lepton number, but the electron doesn't care what we call it, the properties are just numbers.

This is the illusory reality Gnostics were describing. This is the simulation. The A, C, G, and T codes of DNA have a binary value – A and C = 0 while G and T = 1. This has to be when the simulation is digital and the body must be digital to interact with it. Recurring mathematical sequences are encoded throughout reality and the body. They include the Fibonacci sequence in which the two previous numbers are added to get the next one, as in ... 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, etc. The sequence is encoded in the human face and body, proportions of animals, DNA, seed heads, pine cones, trees, shells, spiral galaxies, hurricanes and the number of petals in a flower. The list goes on and on. There are fractal patterns – a 'never-ending pattern that is infinitely complex and self-similar across all scales in the as above, so below, principle of holograms. These and other famous recurring geometrical and mathematical sequences such as Phi, Pi, Golden Mean, Golden Ratio and Golden Section are *computer codes* of the simulation. I had to laugh and give my head a shake the day I finished this book and it went into the production stage. I was sent an article in *Scientific American* published in April, 2021, with the headline 'Confirmed! We Live in a Simulation'. Two decades after I first said our reality is a simulation and the speed of light is it's outer limit the article suggested that we do live in a simulation and that the speed of light is its outer limit. I left school at 15 and never passed a major exam in my life while the writer was up to his eyes in qualifications. As I will explain in the final chapter *knowing* is far better than thinking and they come from very different sources. The article rightly connected the speed of light to the processing speed of the 'Matrix' and said what has been in my books all this time … 'If we are in a simulation, as it appears, then space is an abstract property written in code. It is not real'. No it's not and if we live in a simulation something created it and it wasn't *us*. 'That David Icke says we are manipulated by aliens' – he's crackers.'

Wow …

The reality that humanity thinks is so real is an illusion. Politicians, governments, scientists, doctors, academics, law enforcement, media, school and university curriculums, on and on, are all founded on a world that *does not exist* except as a simulated prison cell. Is it such a stretch to accept that 'Covid' doesn't exist when our entire 'physical' reality doesn't exist? Revealed here is the knowledge kept under raps in the Cult networks of compartmentalised secrecy to control humanity's sense of reality by inducing the population to believe in a reality that's not real. If it wasn't so tragic in its experiential consequences the whole thing would be hysterically funny. None of this is new to Renegade Minds. Ancient Greek philosopher Plato (about 428 to about 347BC) was a major influence on Gnostic belief and he described the human plight thousands of years ago with his Allegory of the Cave. He told the symbolic story of prisoners living in a cave who had never been outside. They were chained and could only see one wall of the cave while behind them was a fire that they could not see. Figures walked past the fire casting shadows on the prisoners' wall and those moving shadows became their sense of reality. Some prisoners began to study the shadows and were considered experts on them (today's academics and scientists), but what they studied was only an illusion (today's academics and scientists). A prisoner escaped from the cave and saw reality as it really is. When he returned to report this revelation they didn't believe him, called him mad and threatened to kill him if he tried to set them free. Plato's tale is not only a brilliant analogy of the human plight and our illusory reality. It describes, too, the dynamics of the 'Covid' hoax. I have only skimmed the surface of these subjects here. The aim of this book is to crisply connect all essential dots to put what is happening today into its true context. All subject areas and their connections in this chapter are covered in great evidential detail in *Everything You Need To Know, But Have Never Been Told* and *The Answer*.

They say that bewildered people 'can't see the forest for the trees'. Humanity, however, can't see the forest for the *twigs*. The five senses see only twigs while Renegade Minds can see the forest and it’s the forest where the answers lie with the connections that reveals. Breaking free of perceptual programming so the forest can be seen is the way we turn all this around. Not breaking free is how humanity got into this mess. The situation may seem hopeless, but I promise you it’s not. We are a perceptual heartbeat from paradise if only we knew.

CHAPTER TWELVE

Escaping Wetiko

Life is simply a vacation from the infinite
Dean Cavanagh

Renegade Minds weave the web of life and events and see common themes in the apparently random. They are always there if you look for them and their pursuit is aided by incredible synchronicity that comes when your mind is open rather than mesmerised by what it thinks it can see.

Infinite awareness is infinite possibility and the more of infinite possibility that we access the more becomes infinitely possible. That may be stating the apparently obvious, but it is a devastatingly-powerful fact that can set us free. We are a point of attention within an infinity of consciousness. The question is how much of that infinity do we choose to access? How much knowledge, insight, awareness, wisdom, do we want to connect with and explore? If your focus is only in the five senses you will be influenced by a fraction of infinite awareness. I mean a range so tiny that it gives new meaning to infinitesimal. Limitation of self-identity and a sense of the possible limit accordingly your range of consciousness. We are what we think we are. Life is what we think it is. The dream is the dreamer and the dreamer is the dream. Buddhist philosophy puts it this way: 'As a thing is viewed, so it appears.' Most humans live in the realm of touch, taste, see, hear, and smell and that's the limit of their sense of the possible and sense of self. Many will follow a religion and speak of a God in his heaven, but their lives are still dominated by the five senses in their perceptions and actions. The five senses become the arbiter of everything. When that happens all except a smear of infinity is sealed away from influence by the rigid, unyielding, reality bubbles that are the five-sense human or Phantom Self. Archon Cult methodology is to isolate consciousness within five-sense reality – the simulation – and then program that consciousness with a sense of self and the world through a deluge of life-long information designed to instil the desired perception that allows global control. Efforts to do this have increased dramatically with identity politics as identity bubbles are squeezed into the minutiae of five-sense detail which disconnect people even more profoundly from the infinite 'I'.

Five-sense focus and self-identity are like a firewall that limits access to the infinite realms. You only perceive one radio or television station and no other. We'll take that literally for a moment. Imagine a vast array of stations giving different information and angles on reality, but you only ever listen to one. Here we have the human plight in which the population is overwhelmingly confined to CultFM. This relates only to the frequency range of CultFM and limits perception and insight to that band – limits *possibility* to that band. It means you are connecting with an almost imperceptibly minuscule range of possibility and creative potential within the infinite Field. It's a world where everything seems apart from everything else and where synchronicity is rare. Synchronicity is defined in the dictionary as 'the happening by chance of two or more related or similar events at the same time'. Use of 'by chance' betrays a complete misunderstanding of reality. Synchronicity is not 'by chance'. As people open their minds, or 'awaken' to use the term, they notice more and more coincidences in their lives, bits of 'luck', apparently miraculous happenings that put them in the right place at the right time with the right people. Days become peppered with 'fancy meeting you here' and 'what are the chances of that?' My entire life has been lived like this and ever more so since my own colossal awakening in 1990 and 91 which transformed my sense of reality. Synchronicity is not 'by chance'; it is by accessing expanded realms of possibility which allow expanded potential for manifestation. People broadcasting the same vibe from the same openness of mind tend to be drawn 'by chance' to each other through what I call frequency magnetism and it's not only people. In the last more than 30 years incredible synchronicity has also led me through the Cult maze to information in so many forms and to crucial personal experiences. These 'coincidences' have allowed me to put the puzzle pieces together across an enormous array of subjects and situations. Those who have breached the bubble of five-sense reality will know exactly what I mean and this escape from the perceptual prison cell is open to everyone whenever they make that choice. This may appear super-human when compared with the limitations of 'human', but it's really our natural state. 'Human' as currently experienced is consciousness in an unnatural state of induced separation from the infinity of the whole. I'll come to how this transformation into unity can be made when I have described in more detail the force that holds humanity in servitude by denying this access to infinite self.

The Wetiko factor

I have been talking and writing for decades about the way five-sense mind is systematically barricaded from expanded awareness. I have used the analogy of a computer (five-sense mind) and someone at the keyboard (expanded awareness). Interaction between the computer and the operator is symbolic of the interaction between five-sense mind and expanded awareness. The computer directly experiences the Internet and the operator experiences the Internet via the computer which is how it's supposed to be – the two working as one. Archons seek to control that point where the operator connects with the computer to stop that interaction (Fig 20). Now the operator is banging the keyboard and clicking the mouse, but the computer is not responding and this happens when the computer is taken over – *possessed* – by an appropriately-named computer 'virus'. The operator has lost all influence over the computer which goes its own way making decisions under the control of the 'virus'. I have just described the dynamic through which the force known to Gnostics as Yaldabaoth and Archons disconnects five-sense mind from expanded awareness to imprison humanity in perceptual servitude.

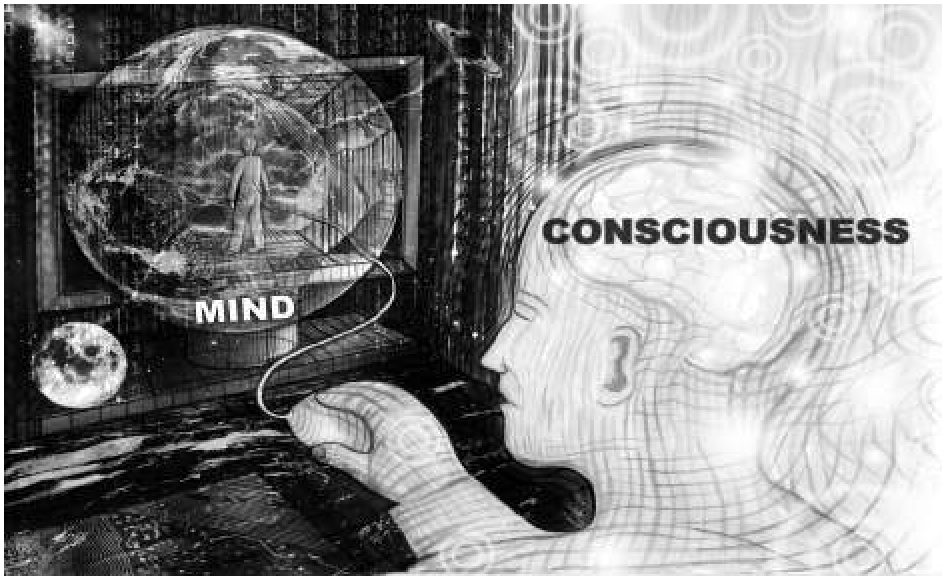


Figure 20: The mind 'virus' I have been writing about for decades seeks to isolate five-sense mind (the computer) from the true 'I'. (Image by Neil Hague).

About a year ago I came across a Native American concept of Wetiko which describes precisely the same phenomenon. Wetiko is the spelling used by the Cree and there are other versions including wintiko and windigo used by other tribal groups. They spell the name with lower case, but I see Wetiko as a proper noun as with Archons and prefer a capital. I first saw an article about Wetiko by writer and researcher Paul Levy which so synced with what I had been writing about the computer/operator disconnection and later the Archons. I then read his book, the fascinating *Dispelling Wetiko, Breaking the Spell of Evil*. The parallels between what I had concluded long before and the Native American concept of Wetiko were so clear and obvious that it was almost funny. For Wetiko see the Gnostic Archons for sure and the Jinn, the Predators, and every other name for a force of evil, inversion and chaos. Wetiko is the Native American name for the force that divides the computer from the operator (Fig 21). Indigenous author Jack D. Forbes, a founder of the Native American movement in the 1960s, wrote another book about Wetiko entitled *Columbus And Other Cannibals – The Wetiko Disease of Exploitation, Imperialism, and Terrorism* which I also read. Forbes says that Wetiko refers to an evil person or spirit 'who terrorizes other creatures by means of terrible acts, including cannibalism'. Zulu shaman Credo Mutwa told me that African accounts tell how cannibalism was brought into the world by the Chitauri 'gods' – another manifestation of Wetiko. The distinction between 'evil person or spirit' relates to Archons/Wetiko possessing a human or acting as pure consciousness. Wetiko is said to be a sickness of the soul or spirit and a state of being that takes but gives nothing back – the Cult and its operatives perfectly described. Black Hawk, a Native American war leader defending their lands from confiscation, said European invaders had 'poisoned hearts' – Wetiko hearts – and that this would spread to native societies. Mention of the heart is very significant as we shall shortly see. Forbes writes: 'Tragically, the history of the world for the past 2,000 years is, in great part, the story of the epidemiology of the wetiko disease.' Yes, and much longer. Forbes is correct when he says: 'The wetikos destroyed Egypt and Babylon and Athens and Rome and Tenochtitlan [capital of the Aztec empire] and perhaps now they will destroy the entire earth.' Evil, he said, is the number one export of a Wetiko culture – see its globalisation with 'Covid'. Constant war, mass murder, suffering of all kinds, child abuse, Satanism, torture and human sacrifice are all expressions of Wetiko and the Wetiko possessed. The world is Wetiko made manifest, *but it doesn't have to be*. There is a way out of this even now.

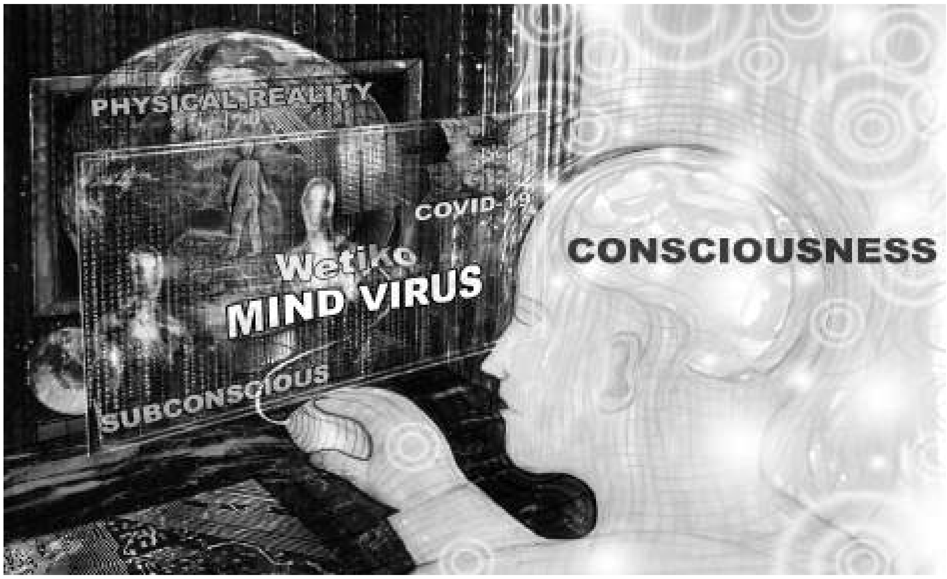


Figure 21: The mind 'virus' is known to Native Americans as 'Wetiko'. (Image by Neil Hague).

Cult of Wetiko

Wetiko is the Yaldabaoth frequency distortion that seeks to attach to human consciousness and absorb it into its own. Once this connection is made Wetiko can drive the perceptions of the target which they believe to be coming from their own mind. All the horrors of history and today from mass killers to Satanists, paedophiles like Jeffrey Epstein and other psychopaths, are the embodiment of Wetiko and express its state of being in all its grotesqueness. The Cult is Wetiko incarnate, Yaldabaoth incarnate, and it seeks to facilitate Wetiko assimilation of humanity in totality into its distortion by manipulating the population into low frequency states that match its own. Paul Levy writes: 'Holographically enforced within the psyche of every human being the wetiko virus pervades and underlies the entire field of consciousness, and can therefore potentially manifest through any one of us at any moment if we are not mindful.' The 'Covid' hoax has achieved this with many people, but others have not fallen into Wetiko's frequency lair. Players in the 'Covid' human catastrophe including Gates, Schwab, Tedros, Fauci, Whitty, Vallance, Johnson, Hancock, Ferguson, Drosten, and all the rest, including the psychopath psychologists, are expressions of Wetiko. This is why they have no compassion or empathy and no emotional consequence for what they do that would make them stop doing it. Observe all the people who support the psychopaths in authority against the Pushbackers despite the damaging impact the psychopaths have on their own lives and their family's lives. You are again looking at Wetiko possession which prevents them seeing through the lies to the obvious scam going on. *Why can't they see it?* Wetiko won't let them see it. The perceptual divide that has now become a chasm is between the Wetikoed and the non-Wetikoed.

Paul Levy describes Wetiko in the same way that I have long described the Archontic force. They are the same distorted consciousness operating across dimensions of reality: '… the subtle body of wetiko is not located in the third dimension of space and time, literally existing in another dimension … it is able to affect ordinary lives by mysteriously interpenetrating into our three-dimensional world.' Wetiko does this through its incarnate representatives in the Cult and by weaving itself into The Field which on our level of reality is the electromagnetic information field of the simulation or Matrix. More than that, the simulation *is* Wetiko / Yaldabaoth. Caleb Scharf, Director of Astrobiology at Columbia University, has speculated that 'alien life' could be so advanced that it has transcribed itself into the quantum realm to become what we call physics. He said intelligence indistinguishable from the fabric of the Universe would solve many of its greatest mysteries:

> Perhaps hyper-advanced life isn't just external. Perhaps it's already all around. It is embedded in what we perceive to be physics itself, from the root behaviour of particles and fields to the phenomena of complexity and emergence … In other words, life might not just be in the equations. It might BE the equations [My emphasis].

Scharf said it is possible that 'we don't recognise advanced life because it forms an integral and unsuspicious part of what we've considered to be the natural world'. I agree. Wetiko/Yaldabaoth *is* the simulation. We are literally in the body of the beast. But that doesn't mean it has to control us. We all have the power to overcome Wetiko influence and the Cult knows that. I doubt it sleeps too well because it knows that.

Which Field?

This, I suggest, is how it all works. There are two Fields. One is the fierce electromagnetic light of the Matrix within the speed of light; the other is the 'watery light' of The Field beyond the walls of the Matrix that connects with the Great Infinity. Five-sense mind and the decoding systems of the body attach us to the Field of Matrix light. They have to or we could not experience this reality. Five-sense mind sees only the Matrix Field of information while our expanded consciousness is part of the Infinity Field. When we open our minds, and most importantly our hearts, to the Infinity Field we have a mission control which gives us an expanded perspective, a road map, to understand the nature of the five-sense world. If we are isolated only in five-sense mind there is no mission control. We're on our own trying to understand a world that's constantly feeding us information to ensure we do not understand. People in this state can feel 'lost' and bewildered with no direction or radar. You can see ever more clearly those who are influenced by the Fields of Big Infinity or little five-sense mind simply by their views and behaviour with regard to the 'Covid' hoax. We have had this division throughout known human history with the mass of the people on one side and individuals who could see and intuit beyond the walls of the simulation – Plato's prisoner who broke out of the cave and saw reality for what it is. Such people have always been targeted by Wetiko/Archon-possessed authority, burned at the stake or demonised as mad, bad and dangerous. The Cult today and its global network of 'anti-hate', 'anti-fascist' Woke groups are all expressions of Wetiko attacking those exposing the conspiracy, 'Covid' lies and the 'vaccine' agenda.

Woke as a whole is Wetiko which explains its black and white mentality and how at one it is with the Wetiko-possessed Cult. Paul Levy said: 'To be in this paradigm is to still be under the thrall of a two-valued logic – where things are either true or false – of a wetikoized mind.’ Wetiko consciousness is in a permanent rage, therefore so is Woke, and then there is Woke inversion and contradiction. ‘Anti-fascists’ act like fascists because fascists *and* ‘anti-fascists’ are both Wetiko at work. Political parties act the same while claiming to be different for the same reason. Secret society and satanic rituals are attaching initiates to Wetiko and the cold, ruthless, psychopathic mentality that secures the positions of power all over the world is Wetiko. Reframing ‘training programmes’ have the same cumulative effect of attaching Wetiko and we have their graduates described as automatons and robots with a cold, psychopathic, uncaring demeanour. They are all traits of Wetiko possession and look how many times they have been described in this book and elsewhere with regard to personnel behind ‘Covid’ including the police and medical profession. Climbing the greasy pole in any profession in a Wetiko society requires traits of Wetiko to get there and that is particularly true of politics which is not about fair competition and pre-eminence of ideas. It is founded on how many backs you can stab and arses you can lick. This culminated in the global ‘Covid’ coordination between the Wetiko possessed who pulled it off in all the different countries without a trace of empathy and compassion for their impact on humans. Our sight sense can see only holographic form and not the Field which connects holographic form. Therefore we perceive ‘physical’ objects with ‘space’ in between. In fact that ‘space’ is energy/consciousness operating on multiple frequencies. One of them is Wetiko and that connects the Cult psychopaths, those who submit to the psychopaths, and those who serve the psychopaths in the media operations of the world. Wetiko is Gates. Wetiko is the mask-wearing submissive. Wetiko is the fake journalist and ‘fact-checker’. The Wetiko Field is coordinating the whole thing. Psychopaths, gofers, media operatives, ‘anti-hate’ hate groups, ‘fact-checkers’ and submissive people work as one unit *even without human coordination* because they are attached to the *same* Field which is organising it all (Fig 22). Paul Levy is here describing how Wetiko-possessed people are drawn together and refuse to let any information breach their rigid perceptions. He was writing long before ‘Covid’, but I think you will recognise followers of the ‘Covid’ religion *oh just a little bit*:

> People who are channelling the vibratory frequency of wetiko align with each other through psychic resonance to reinforce their unspoken shared agreement so as to uphold their deranged view of reality. Once an unconscious content takes possession of certain individuals, it irresistibly draws them together by mutual attraction and knits them into groups tied together by their shared madness that can easily swell into an avalanche of insanity.
>
> A psychic epidemic is a closed system, which is to say that it is insular and not open to any new information or informing influences from the outside world which contradict its fixed, limited, and limiting perspective.

There we have the Woke mind and the ‘Covid’ mind. Compatible resonance draws the awakening together, too, which is clearly happening today.

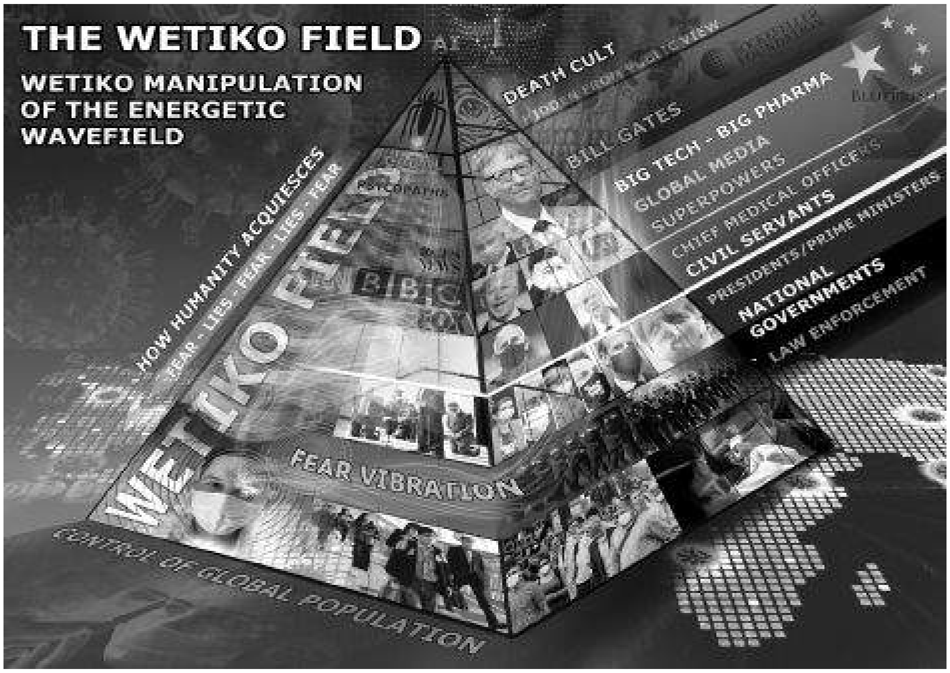

THE WETIKO FIELD
WETIKO MANIPULATION
OF THE ENERGETIC
WAVEFIELD
DEATH CULT
BILL GATES
BIG TECH - BIG PHARMA
GLOBAL MEDIA
SUPERPOWERS
CHIEF MEDICAL OFFICERS
CIVIL SERVANTS
PRESIDENTS/PRIME MINISTERS
NATIONAL
GOVERNMENTS
LAW ENFORCEMENT
HOW HUMANITY ACQUIESCES
FEAR - LIES - FEAR - LIES - FEAR
PSYCHOPATHS
BBC
FOX
WETIKO FIELD
FEAR VIBRATION
CONTROL OF GLOBAL POPULATION

Figure 22: The Wetiko Field from which the Cult pyramid and its personnel are made manifest. (Image by Neil Hague).

Spiritual servitude

Wetiko doesn’t care about humans. It’s not human; it just possesses humans for its own ends and the effect (depending on the scale of possession) can be anything from extreme psychopathy to unquestioning obedience. Wetiko's worst nightmare is for human consciousness to expand beyond the simulation. Everything is focussed on stopping that happening through control of information, thus perception, thus frequency. The 'education system', media, science, medicine, academia, are all geared to maintaining humanity in five-sense servitude as is the constant stimulation of low-vibrational mental and emotional states (see 'Covid'). Wetiko seeks to dominate those subconscious spaces between five-sense perception and expanded consciousness where the computer meets the operator. From these subconscious hiding places Wetiko speaks to us to trigger urges and desires that we take to be our own and manipulate us into anything from low-vibrational to psychopathic states. Remember how Islam describes the Jinn as invisible tricksters that 'whisper' and confuse. Wetiko is the origin of the 'trickster god' theme that you find in cultures all over the world. Jinn, like the Archons, are Wetiko which is terrified of humans awakening and reconnecting with our true self for then its energy source has gone. With that the feedback loop breaks between Wetiko and human perception that provides the energetic momentum on which its very existence depends as a force of evil. Humans are both its target and its source of survival, but only if we are operating in low-vibrational states of fear, hate, depression and the background anxiety that most people suffer. We are Wetiko's target because we are its key to survival. It needs us, not the other way round. Paul Levy writes:

> A vampire has no intrinsic, independent, substantial existence in its own right; it only exists in relation to us. The pathogenic, vampiric mind-parasite called wetiko is nothing in itself – not being able to exist from its own side – yet it has a 'virtual reality' such that it can potentially destroy our species …
>
> …The fact that a vampire is not reflected by a mirror can also mean that what we need to see is that there's nothing, no-thing to see, other than ourselves. The fact that wetiko is the expression of something inside of us means that the cure for wetiko is with us as well. The critical issue is finding this cure within us and then putting it into effect.

Evil begets evil because if evil does not constantly expand and find new sources of energetic sustenance its evil, its *distortion*, dies with the assimilation into balance and harmony. Love is the garlic to Wetiko's vampire. Evil, the absence of love, cannot exist in the presence of love. I think I see a way out of here. I have emphasised so many times over the decades that the Archons/Wetiko and their Cult are not all powerful. *They are not*. I don't care how it looks even now *they are not*. I have not called them little boys in short trousers for effect. I have said it because it is true. Wetiko's insatiable desire for power over others is not a sign of its omnipotence, but its insecurity. Paul Levy writes: 'Due to the primal fear which ultimately drives it and which it is driven to cultivate, wetiko's body politic has an intrinsic and insistent need for centralising power and control so as to create imagined safety for itself.' *Yeeeeeees!* Exactly! Why does Wetiko want humans in an ongoing state of fear? Wetiko itself *is* fear and it is petrified of love. As evil is an absence of love, so love is an absence of fear. Love conquers all and *especially* Wetiko which *is* fear. Wetiko brought fear into the world when it wasn't here before. *Fear* was the 'fall', the fall into low-frequency ignorance and illusion – fear is **F**alse **E**motion **A**ppearing **R**eal. The simulation is driven and energised by fear because Wetiko/Yaldabaoth (fear) *are* the simulation. Fear is the absence of love and Wetiko is the absence of love.

Wetiko today

We can now view current events from this level of perspective. The 'Covid' hoax has generated momentous amounts of ongoing fear, anxiety, depression and despair which have empowered Wetiko. No wonder people like Gates have been the instigators when they are Wetiko incarnate and exhibit every trait of Wetiko in the extreme. See how cold and unemotional these people are like Gates and his cronies, how dead of eye they are. That's Wetiko. Sabbatians are Wetiko and everything they control including the World Health Organization, Big Pharma and the 'vaccine' makers, national 'health' hierarchies, corporate media, Silicon Valley, the banking system, and the United Nations with its planned transformation into world government. All are controlled and possessed by the Wetiko distortion into distorting human society in its image. We are with this knowledge at the gateway to understanding the world. Divisions of race, culture, creed and sexuality are diversions to hide the real division between those possessed and influenced by Wetiko and those that are not. The 'Covid' hoax has brought both clearly into view. Human behaviour is not about race. Tyrants and dictatorships come in all colours and creeds. What unites the US president bombing the innocent and an African tribe committing genocide against another as in Rwanda? What unites them? *Wetiko*. All wars are Wetiko, all genocide is Wetiko, all hunger over centuries in a world of plenty is Wetiko. Children going to bed hungry, including in the West, is Wetiko. Cult-generated Woke racial divisions that focus on the body are designed to obscure the reality that divisions in behaviour are manifestations of mind, not body. Obsession with body identity and group judgement is a means to divert attention from the real source of behaviour – mind and perception. Conflict sown by the Woke both within themselves and with their target groups are Wetiko providing lunch for itself through still more agents of the division, chaos, and fear on which it feeds. The Cult is seeking to assimilate the entirety of humanity and all children and young people into the Wetiko frequency by manipulating them into states of fear and despair. Witness all the suicide and psychological unravelling since the spring of 2020. Wetiko psychopaths want to impose a state of unquestioning obedience to authority which is no more than a conduit for Wetiko to enforce its will and assimilate humanity into itself. It needs us to believe that resistance is futile when it fears resistance and even more so the game-changing non-cooperation with its impositions. It can use violent resistance for its benefit. Violent impositions and violent resistance are *both* Wetiko. The Power of Love with its Power of No will sweep Wetiko from our world. Wetiko and its Cult know that. They just don't want us to know.

AI Wetiko

This brings me to AI or artificial intelligence and something else Wetikos don't want us to know. What is AI *really*? I know about computer code algorithms and AI that learns from data input. These, however, are more diversions, the expeditionary force, for the real AI that they want to connect to the human brain as promoted by Silicon Valley Wetikos like Kurzweil. What is this AI? It is the frequency of *Wetiko*, the frequency of the Archons. The connection of AI to the human brain is the connection of the Wetiko frequency to create a Wetiko hive mind and complete the job of assimilation. The hive mind is planned to be controlled from Israel and China which are both 100 percent owned by Wetiko Sabbatians. The assimilation process has been going on minute by minute in the 'smart' era which fused with the 'Covid' era. We are told that social media is scrambling the minds of the young and changing their personality. This is true, but what is social media? Look more deeply at how it works, how it creates divisions and conflict, the hostility and cruelty, the targeting of people until they are destroyed. That's Wetiko. Social media is manipulated to tune people to the Wetiko frequency with all the emotional exploitation tricks employed by platforms like Facebook and its Wetiko front man, Zuckerberg. Facebook's Instagram announced a new platform for children to overcome a legal bar on them using the main site. This is more Wetiko exploitation and manipulation of kids. Amnesty International likened the plan to foxes offering to guard the henhouse and said it was incompatible with human rights. Since when did Wetiko or Zuckerberg (I repeat myself) care about that? Would Brin and Page at Google, Wojcicki at YouTube, Bezos at Amazon and whoever the hell runs Twitter act as they do if they were not channelling Wetiko? Would those who are developing technologies for no other reason than human control? How about those designing and selling technologies to kill people and Big Pharma drug and 'vaccine' producers who know they will end or devastate lives? Quite a thought for these people to consider is that if you are Wetiko in a human life you are Wetiko on the 'other side' unless your frequency changes and that can only change by a change of perception which becomes a change of behaviour. Where Gates is going does not bear thinking about although perhaps that's exactly where he wants to go. Either way, that's where he's going. His frequency will make it so.

The frequency lair

I have been saying for a long time that a big part of the addiction to smartphones and devices is that a frequency is coming off them that entraps the mind. People spend ages on their phones and sometimes even a minute or so after they put them down they pick them up again and it all repeats. 'Covid' lockdowns will have increased this addiction a million times for obvious reasons. Addictions to alcohol overindulgence and drugs are another way that Wetiko entraps consciousness to attach to its own. Both are symptoms of low-vibrational psychological distress which alcoholism and drug addiction further compound. Do we think it's really a coincidence that access to them is made so easy while potions that can take people into realms beyond the simulation are banned and illegal? I have explored smartphone addiction in other books, the scale is mind-blowing, and that level of addiction does not come without help. Tech companies that make these phones are Wetiko and they will have no qualms about destroying the minds of children. We are seeing again with these companies the Wetiko perceptual combination of psychopathic enforcers and weak and meek unquestioning compliance by the rank and file.

The global Smart Grid is the Wetiko Grid and it is crucial to complete the Cult endgame. The simulation is radiation and we are being deluged with technological radiation on a devastating scale. Wetiko frauds like Elon Musk serve Cult interests while occasionally criticising them to maintain his street-cred. 5G and other forms of Wi-Fi are being directed at the earth from space on a volume and scale that goes on increasing by the day. Elon Musk's (officially) SpaceX Starlink project is in the process of putting tens of thousands of satellites in low orbit to cover every inch of the planet with 5G and other Wi-Fi to create Kurzweil's global 'cloud' to which the human mind is planned to be attached very soon. SpaceX has approval to operate 12,000 satellites with more than 1,300 launched at the time of writing and applications filed for *30,000* more. Other operators in the Wi-Fi, 5G, low-orbit satellite market include OneWeb (UK), Telesat (Canada), and AST & Science (US). Musk tells us that AI could be the end of humanity and then launches a company called Neuralink to connect the human brain to computers. Musk's (in theory) Tesla company is building electric cars and the driverless vehicles of the smart control grid. As frauds and bullshitters go Elon Musk in my opinion is Major League.

5G and technological radiation in general are destructive to human health, genetics and psychology and increasing the strength of artificial radiation underpins the five-sense perceptual bubbles which are themselves expressions of radiation or electromagnetism. Freedom activist John Whitehead was so right with his 'databit by databit, we are building our own electronic concentration camps'. The Smart Grid and 5G is a means to control the human mind and infuse perceptual information into The Field to influence anyone in sync with its frequency. You can change perception and behaviour en masse if you can manipulate the population into those levels of frequency and this is happening all around us today. The arrogance of Musk and his fellow Cult operatives knows no bounds in the way that we see with Gates. Musk's satellites are so many in number already they are changing the night sky when viewed from Earth. The astronomy community has complained about this and they have seen nothing yet. Some consequences of Musk's Wetiko hubris include: Radiation; visible pollution of the night sky; interference with astronomy and meteorology; ground and water pollution from intensive use of increasingly many spaceports; accumulating space debris; continual deorbiting and burning up of aging satellites, polluting the atmosphere with toxic dust and smoke; and ever-increasing likelihood of collisions. A collective public open letter of complaint to Musk said:

> We are writing to you … because SpaceX is in process of surrounding the Earth with a network of thousands of satellites whose very purpose is to irradiate every square inch of the

> Earth. SpaceX, like everyone else, is treating the radiation as if it were not there. As if the mitochondria in our cells do not depend on electrons moving undisturbed from the food we digest to the oxygen we breathe.
>
> As if our nervous systems and our hearts are not subject to radio frequency interference like any piece of electronic equipment. As if the cancer, diabetes, and heart disease that now afflict a majority of the Earth's population are not metabolic diseases that result from interference with our cellular machinery. As if insects everywhere, and the birds and animals that eat them, are not starving to death as a result.

People like Musk and Gates believe in their limitless Wetiko arrogance that they can do whatever they like to the world because they own it. Consequences for humanity are irrelevant. It's absolutely time that we stopped taking this shit from these self-styled masters of the Earth when you consider where this is going.

Why is the Cult so anti-human?

I hear this question often: Why would they do this when it will affect them, too? Ah, but will it? Who is this *them*? Forget their bodies. They are just vehicles for Wetiko consciousness. When you break it all down to the foundations we are looking at a state of severely distorted consciousness targeting another state of consciousness for assimilation. The rest is detail. The simulation is the fly-trap in which unique sensations of the five senses create a cycle of addiction called reincarnation. Renegade Minds see that everything which happens in our reality is a smaller version of the whole picture in line with the holographic principle. Addiction to the radiation of smart technology is a smaller version of addiction to the whole simulation. Connecting the body/brain to AI is taking that addiction on a giant step further to total ongoing control by assimilating human incarnate consciousness into Wetiko. I have watched during the 'Covid' hoax how many are becoming ever more profoundly attached to Wetiko's perceptual calling cards of aggressive response to any other point of view ('There is no other god but me'), psychopathic lack of compassion and empathy, and servile submission to the narrative and will of authority. Wetiko is the psychopaths *and* subservience to psychopaths. The Cult of Wetiko is so anti-human because it is *not* human. It embarked on a mission to destroy human by targeting everything that it means to be human and to survive as human. 'Covid' is not the end, just a means to an end. The Cult with its Wetiko consciousness is seeking to change Earth systems, including the atmosphere, to suit them, not humans. The gathering bombardment of 5G alone from ground and space is dramatically changing The Field with which the five senses interact. There is so much more to come if we sit on our hands and hope it will all go away. It is not meant to go away. It is meant to get ever more extreme and we need to face that while we still can – just.

Carbon dioxide is the gas of life. Without that human is over. Kaput, gone, history. No natural world, no human. The Cult has created a cock and bull story about carbon dioxide and climate change to justify its reduction to the point where Gates and the ignoramus Biden 'climate chief' John Kerry want to suck it out of the atmosphere. Kerry wants to do this because his master Gates does. Wetikos have made the gas of life a demon with the usual support from the Wokers of Extinction Rebellion and similar organisations and the bewildered puppet-child that is Greta Thunberg who was put on the world stage by Klaus Schwab and the World Economic Forum. The name Extinction Rebellion is both ironic and as always Wetiko inversion. The gas that we need to survive must be reduced to save us from extinction. The most basic need of human is oxygen and we now have billions walking around in face nappies depriving body and brain of this essential requirement of human existence. More than that 5G at 60 gigahertz interacts with the oxygen molecule to reduce the amount of oxygen the body can absorb into the bloodstream. The obvious knock-on consequences of that for respiratory and cognitive problems and life itself need no further explanation. Psychopaths like Musk are assembling a global system of satellites to deluge the human atmosphere with this insanity. The man should be in jail. Here we have two most basic of human needs, oxygen and carbon dioxide, being dismantled.

Two others, water and food, are getting similar treatment with the United Nations Agendas 21 and 2030 – the Great Reset – planning to centrally control all water and food supplies. People will not even own rain water that falls on their land. Food is affected at the most basic level by reducing carbon dioxide. We have genetic modification or GMO infiltrating the food chain on a mass scale, pesticides and herbicides polluting the air and destroying the soil. Freshwater fish that provide livelihoods for 60 million people and feed hundreds of millions worldwide are being 'pushed to the brink' according the conservationists while climate change is the only focus. Now we have Gates and Schwab wanting to dispense with current food sources all together and replace them with a synthetic version which the Wetiko Cult would control in terms of production and who eats and who doesn't. We have been on the Totalitarian Tiptoe to this for more than 60 years as food has become ever more processed and full of chemical shite to the point today when it's not natural food at all. As Dr Tom Cowan says: 'If it has a label don't eat it.' Bill Gates is now the biggest owner of farmland in the United States and he does nothing without an ulterior motive involving the Cult. Klaus Schwab wrote: 'To feed the world in the next 50 years we will need to produce as much food as was produced in the last 10,000 years … food security will only be achieved, however, if regulations on genetically modified foods are adapted to reflect the reality that gene editing offers a precise, efficient and safe method of improving crops.' Liar. People and the world are being targeted with aluminium through vaccines, chemtrails, food, drink cans, and endless other sources when aluminium has been linked to many health issues including dementia which is increasing year after year. Insects, bees and wildlife essential to the food chain are being deleted by pesticides, herbicides and radiation which 5G is dramatically increasing with 6G and 7G to come. The pollinating bee population is being devastated while wildlife including birds, dolphins and whales are having their natural radar blocked by the effects of ever-increasing radiation. In the summer windscreens used to be splattered with insects so numerous were they. It doesn't happen now. Where have they gone?

Synthetic everything

The Cult is introducing genetically-modified versions of trees, plants and insects including a Gates-funded project to unleash hundreds of millions of genetically-modified, lab-altered and patented male mosquitoes to mate with wild mosquitoes and induce genetic flaws that cause them to die out. Clinically-insane Gates-funded Japanese researchers have developed mosquitos that spread vaccine and are dubbed 'flying vaccinators'. Gates is funding the modification of weather patterns in part to sell the myth that this is caused by carbon dioxide and he's funding geoengineering of the skies to change the atmosphere. Some of this came to light with the Gates-backed plan to release tonnes of chalk into the atmosphere to 'deflect the Sun and cool the planet'. Funny how they do this while the heating effect of the Sun is not factored into climate projections focussed on carbon dioxide. The reason is that they want to reduce carbon dioxide (so don't mention the Sun), but at the same time they do want to reduce the impact of the Sun which is so essential to human life and health. I have mentioned the sun-cholesterol-vitamin D connection as they demonise the Sun with warnings about skin cancer (caused by the chemicals in sun cream they tell you to splash on). They come from the other end of the process with statin drugs to reduce cholesterol that turns sunlight into vitamin D. A lack of vitamin D leads to a long list of health effects and how vitamin D levels must have fallen with people confined to their homes over 'Covid'. Gates is funding other forms of geoengineering and most importantly chemtrails which are dropping heavy metals, aluminium and self-replicating nanotechnology onto the Earth which is killing the natural world. See *Everything You Need To Know, But Have Never Been Told* for the detailed background to this.

Every human system is being targeted for deletion by a force that's not human. The Wetiko Cult has embarked on the process of transforming the human body from biological to synthetic biological as I have explained. Biological is being replaced by the artificial and synthetic – Archontic 'countermimicry' – right across human society. The plan eventually is to dispense with the human body altogether and absorb human consciousness – which it wouldn't really be by then – into cyberspace (the simulation which is Wetiko/Yaldabaoth). Preparations for that are already happening if people would care to look. The alternative media rightly warns about globalism and 'the globalists', but this is far bigger than that and represents the end of the human race as we know it. The 'bad copy' of prime reality that Gnostics describe was a bad copy of harmony, wonder and beauty to start with before Wetiko/Yaldabaoth set out to change the simulated 'copy' into something very different. The process was slow to start with. Entrapped humans in the simulation timeline were not technologically aware and they had to be brought up to intellectual speed while being suppressed spiritually to the point where they could build their own prison while having no idea they were doing so. We have now reached that stage where technological intellect has the potential to destroy us and that's why events are moving so fast. Central American shaman Don Juan Matus said:

> Think for a moment, and tell me how you would explain the contradictions between the intelligence of man the engineer and the stupidity of his systems of belief, or the stupidity of his contradictory behaviour. Sorcerers believe that the predators have given us our systems of beliefs, our ideas of good and evil; our social mores. They are the ones who set up our dreams of success or failure. They have given us covetousness, greed, and cowardice. It is the predator who makes us complacent, routinary, and egomaniacal.
>
> In order to keep us obedient and meek and weak, the predators engaged themselves in a stupendous manoeuvre – stupendous, of course, from the point of view of a fighting strategist; a horrendous manoeuvre from the point of those who suffer it. They gave us their mind. The predators' mind is baroque, contradictory, morose, filled with the fear of being discovered any minute now.

For 'predators' see Wetiko, Archons, Yaldabaoth, Jinn, and all the other versions of the same phenomenon in cultures and religions all over the world. The theme is always the same because it's true and it's real. We have reached the point where we have to deal with it. The question is – how?

Don't fight – walk away

I thought I'd use a controversial subheading to get things moving in terms of our response to global fascism. What do you mean 'don't fight'? What do you mean 'walk away'? We've got to fight. We can't walk away. Well, it depends what we mean by fight and walk away. If fighting means physical combat we are playing Wetiko's game and falling for its trap. It wants us to get angry, aggressive, and direct hate and hostility at the enemy we think we must fight. Every war, every battle, every conflict, has been fought with Wetiko leading both sides. It's what it does. Wetiko wants a fight, anywhere, any place. Just hit me, son, so I can hit you back. Wetiko hits Wetiko and Wetiko hits Wetiko in return. I am very forthright as you can see in exposing Wetikos of the Cult, but I don't hate them. I refuse to hate them. It's what they want. What you hate you become. What you *fight* you become. Wokers, 'anti-haters' and 'anti-fascists' prove this every time they reach for their keyboards or don their balaclavas. By walk away I mean to disengage from Wetiko which includes ceasing to cooperate with its tyranny. Paul Levy says of Wetiko:

> The way to 'defeat' evil is not to try to destroy it (for then, in playing evil's game, we have already lost), but rather, to find the invulnerable place within ourselves where evil is unable to vanquish us – this is to truly 'win' our battle with evil.

Wetiko is everywhere in human society and it's been on steroids since the 'Covid' hoax. Every shouting match over wearing masks has Wetiko wearing a mask and Wetiko not wearing one. It's an electrical circuit of push and resist, push and resist, with Wetiko pushing *and* resisting. Each polarity is Wetiko empowering itself. Dictionary definitions of 'resist' include 'opposing, refusing to accept or comply with' and the word to focus on is 'opposing'. What form does this take – setting police cars alight or 'refusing to accept or comply with'? The former is Wetiko opposing Wetiko while the other points the way forward. This is the difference between those aggressively demanding that government fascism must be obeyed who stand in stark contrast to the great majority of Pushbackers. We saw this clearly with a march by thousands of Pushbackers against lockdown in London followed days later by a Woker-hijacked protest in Bristol in which police cars were set on fire. Masks were virtually absent in London and widespread in Bristol. Wetiko wants lockdown on every level of society and infuses its aggression to police it through its unknowing stooges. Lockdown protesters are the ones with the smiling faces and the hugs, The two blatantly obvious states of being – getting more obvious by the day – are the result of Wokers and their like becoming ever more influenced by the simulation Field of Wetiko and Pushbackers ever more influenced by The Field of a far higher vibration beyond the simulation. Wetiko can't invade the heart which is where most lockdown opponents are coming from. It's the heart that allows them to see through the lies to the truth in ways I will be highlighting.

Renegade Minds know that calmness is the place from which wisdom comes. You won't find wisdom in a hissing fit and wisdom is what we need in abundance right now. Calmness is not weakness – you don't have to scream at the top of your voice to be strong. Calmness is indeed a sign of strength. 'No' means I'm not doing it. *NOOOO!!!* doesn't mean you're not doing it even more. Volume does not advance 'No – I'm not doing it'. You are just not doing it. Wetiko possessed and influenced don't know how to deal with that. Wetiko wants a fight and we should not give it one. What it needs more than anything is our *cooperation* and we should not give that either. Mass rallies and marches are great in that they are a visual representation of feeling, but if it ends there they are irrelevant. You demand that Wetikos act differently? Well, they're not going to are they? They are Wetikos. We don't need to waste our time demanding that something doesn't happen when that will make no difference. We need to delete the means that *allows* it to happen. This, invariably, is our cooperation. You can demand a child stop firing a peashooter at the dog or you can refuse to buy the peashooter. If you provide the means you are cooperating with the dog being smacked on the nose with a pea. How can the authorities enforce mask-wearing if millions in a country refuse? What if the 74 million Pushbackers that voted for Trump in 2020 refused to wear masks, close their businesses or stay in their homes. It would be unenforceable. The few control the many through the compliance of the many and that's always been the dynamic be it 'Covid' regulations or the Roman Empire. I know people can find it intimidating to say no to authority or stand out in a crowd for being the only one with a face on display; but it has to be done or it's over. I hope I've made clear in this book that where this is going will be far more intimidating than standing up now and saying 'No' – I will not cooperate with my own enslavement and that of my children. There might be consequences for some initially, although not so if enough do the same. The question that must be addressed is what is going to happen if we don't? It is time to be strong and unyieldingly so. No means no. Not here and there, but *everywhere* and *always*. I have refused to wear a mask and obey all the other nonsense. I will not comply with tyranny. I repeat: Fascism is not imposed by fascists – there are never enough of them. Fascism is imposed by the population acquiescing to fascism. *I will not do it*. I will die first, or my body will. Living meekly under fascism is a form of death anyway, the death of the spirit that Martin Luther King described.

Making things happen

We must not despair. This is not over till it's over and it's far from that. The 'fat lady' must refuse to sing. The longer the 'Covid' hoax has dragged on and impacted on more lives we have seen an awakening of phenomenal numbers of people worldwide to the realisation that what they have believed all their lives is not how the world really is. Research published by the system-serving University of Bristol and King's College London in February, 2021, concluded: 'One in every 11 people in Britain say they trust David Icke's take on the coronavirus pandemic.' It will be more by now and we have gathering numbers to build on. We must urgently progress from seeing the scam to ceasing to cooperate with it. Prominent German lawyer Reiner Fuellmich, also licenced to practice law in America, is doing a magnificent job taking the legal route to bring the psychopaths to justice through a second Nuremberg tribunal for crimes against humanity. Fuellmich has an impressive record of beating the elite in court and he formed the German Corona Investigative Committee to pursue civil charges against the main perpetrators with a view to triggering criminal charges. Most importantly he has grasped the foundation of the hoax – the PCR test not testing for the 'virus' – and Christian Drosten is therefore on his charge sheet along with Gates frontman Tedros at the World Health Organization. Major players must be not be allowed to inflict their horrors on the human race without being brought to book. A life sentence must follow for Bill Gates and the rest of them. A group of researchers has also indicted the government of Norway for crimes against humanity with copies sent to the police and the International Criminal Court. The lawsuit cites participation in an internationally-planned false pandemic and violation of international law and human rights, the European Commission's definition of human rights by coercive rules, Nuremberg and Hague rules on fundamental human rights, and the Norwegian constitution. We must take the initiative from hereon and not just complain, protest and react.

There are practical ways to support vital mass non-cooperation. Organising in numbers is one. Lockdown marches in London in the spring in 2021 were mass non-cooperation that the authorities could not stop. There were too many people. Hundreds of thousands walked the London streets in the centre of the road for mile after mile while the Face-Nappies could only look on. They were determined, but calm, and just *did it* with no histrionics and lots of smiles. The police were impotent. Others are organising group shopping without masks for mutual support and imagine if that was happening all over. Policing it would be impossible. If the store refuses to serve people in these circumstances they would be faced with a long line of trolleys full of goods standing on their own and everything would have to be returned to the shelves. How would they cope with that if it kept happening? I am talking here about moving on from complaining to being pro-active; from watching things happen to making things happen. I include in this our relationship with the police. The behaviour of many Face-Nappies has been disgraceful and anyone who thinks they would never find concentration camp guards in the 'enlightened' modern era have had that myth busted big-time. The period and setting may change – Wetikos never do. I watched film footage from a London march in which a police thug viciously kicked a protestor on the floor who had done nothing. His fellow Face-Nappies stood in a ring protecting him. What he did was a criminal assault and with a crowd far outnumbering the police this can no longer be allowed to happen unchallenged. I get it when people chant 'shame on you' in these circumstances, but that is no longer enough. They *have* no shame those who do this. Crowds needs to start making a citizen's arrest of the police who commit criminal offences and brutally attack innocent people and defenceless women. A citizen's arrest can be made under section 24A of the UK Police and Criminal Evidence (PACE) Act of 1984 and you will find something similar in other countries. I prefer to call it a Common Law arrest rather than citizen's for reasons I will come to shortly. Anyone can arrest a person committing an indictable offence or if they have reasonable grounds to suspect they are committing an indictable offence. On both counts the attack by the police thug would have fallen into this category. A citizen's arrest can be made to stop someone:

- Causing physical injury to himself or any other person
- Suffering physical injury
- Causing loss of or damage to property
- Making off before a constable can assume responsibility for him

A citizen's arrest may also be made to prevent a breach of the peace under Common Law and if they believe a breach of the peace will happen or anything related to harm likely to be done or already done in their presence. This is the way to go I think – the Common Law version. If police know that the crowd and members of the public will no longer be standing and watching while they commit their thuggery and crimes they will think twice about acting like Brownshirts and Blackshirts.

Common Law – common sense

Mention of Common Law is very important. Most people think the law is the law as in one law. This is not the case. There are two bodies of law, Common Law and Statute Law, and they are not the same. Common Law is founded on the simple premise of do no harm. It does not recognise victimless crimes in which no harm is done while Statute Law does. There is a Statute Law against almost everything. So what is Statute Law? Amazingly it's the law of the *sea* that was brought ashore by the Cult to override the law of the land which is Common Law. They had no right to do this and as always they did it anyway. They had to. They could not impose their will on the people through Common Law which only applies to do no harm. How could you stitch up the fine detail of people's lives with that? Instead they took the law of the sea, or Admiralty Law, and applied it to the population. Statute Law refers to all the laws spewing out of governments and their agencies including all the fascist laws and regulations relating to 'Covid'. The key point to make is that Statute Law is *contract law*. It only applies between *contracting* corporations. Most police officers don't even know this. They have to be kept in the dark, too. Long ago when merchants and their sailing ships began to trade with different countries a contractual law was developed called Admiralty Law and other names. Again it only applied to *contracts* agreed between *corporate* entities. If there is no agreed contract the law of the sea had no jurisdiction *and that still applies to its new alias of Statute Law*. The problem for the Cult when the law of the sea was brought ashore was an obvious one. People were not corporations and neither were government entities. To overcome the latter they made governments and all associated organisations corporations. All the institutions are *private corporations* and I mean governments and their agencies, local councils, police, courts, military, US states, the whole lot. Go to the

Dun and Bradstreet corporate listings website for confirmation that they are all corporations. You are arrested by a private corporation called the police by someone who is really a private security guard and they take you to court which is another private corporation. Neither have jurisdiction over you unless you consent and *contract* with them. This is why you hear the mantra about law enforcement policing by *consent* of the people. In truth the people 'consent' only in theory through monumental trickery.

Okay, the Cult overcame the corporate law problem by making governments and institutions corporate entities; but what about people? They are not corporations are they? Ah ... well in a sense, and *only* a sense, they are. Not people exactly – the illusion of people. The Cult creates a corporation in the name of everyone at the time that their birth certificate is issued. Note birth/ *berth* certificate and when you go to court under the law of the sea on land you stand in a *dock*. These are throwbacks to the origin. My Common Law name is David Vaughan Icke. The name of the corporation created by the government when I was born is called Mr David Vaughan Icke usually written in capitals as MR DAVID VAUGHAN ICKE. That is not me, the living, breathing man. It is a fictitious corporate entity. The trick is to make you think that David Vaughan Icke and MR DAVID VAUGHAN ICKE are the same thing. *They are not*. When police charge you and take you to court they are prosecuting the corporate entity and not the living, breathing, man or woman. They have to trick you into identifying as the corporate entity and contracting with them. Otherwise they have no jurisdiction. They do this through a language known as legalese. Lawful and legal are not the same either. Lawful relates to Common Law and legal relates to Statute Law. Legalese is the language of Statue Law which uses terms that mean one thing to the public and another in legalese. Notice that when a police officer tells someone why they are being charged he or she will say at the end: 'Do you understand?' To the public that means 'Do you comprehend?' In legalese it means 'Do you stand under me?' Do you stand under my authority? If you say yes to the question you are unknowingly agreeing to give them jurisdiction over you in a contract between two corporate entities.

This is a confidence trick in every way. Contracts have to be agreed between informed parties and if you don't know that David Vaughan Icke is agreeing to be the corporation MR DAVID VAUGHAN ICKE you cannot knowingly agree to contract. They are deceiving you and another way they do this is to ask for proof of identity. You usually show them a driving licence or other document on which your corporate name is written. In doing so you are accepting that you are that corporate entity when you are not. Referring to yourself as a 'person' or 'citizen' is also identifying with your corporate fiction which is why I made the Common Law point about the citizen's arrest. If you are approached by a police officer you identify yourself immediately as a living, breathing, man or woman and say 'I do not consent, I do not contract with you and I do not understand' or stand under their authority. I have a Common Law birth certificate as a living man and these are available at no charge from commonlawcourt.com. Businesses registered under the Statute Law system means that its laws apply. There are, however, ways to run a business under Common Law. Remember all 'Covid' laws and regulations are Statute Law – the law of *contracts* and you do not have to contract. This doesn't mean that you can kill someone and get away with it. Common Law says do no harm and that applies to physical harm, financial harm etc. Police are employees of private corporations and there needs to be a new system of non-corporate Common Law constables operating outside the Statute Law system. If you go to davidicke.com and put Common Law into the search engine you will find videos that explain Common Law in much greater detail. It is definitely a road we should walk.

With all my heart

I have heard people say that we are in a spiritual war. I don't like the term 'war' with its Wetiko dynamic, but I know what they mean. Sweep aside all the bodily forms and we are in a situation in which two states of consciousness are seeking very different realities.

Wetiko wants upheaval, chaos, fear, suffering, conflict and control. The other wants love, peace, harmony, fairness and freedom. That's where we are. We should not fall for the idea that Wetiko is all-powerful and there's nothing we can do. Wetiko is not all-powerful. It's a joke, pathetic. It doesn't have to be, but it has made that choice for now. A handful of times over the years when I have felt the presence of its frequency I have allowed it to attach briefly so I could consciously observe its nature. The experience is not pleasant, the energy is heavy and dark, but the ease with which you can kick it back out the door shows that its real power is in persuading us that it has power. It's all a con. Wetiko is a con. It's a trickster and not a power that can control us if we unleash our own. The con is founded on manipulating humanity to give its power to Wetiko which recycles it back to present the illusion that it has power when its power is *ours* that we gave away. This happens on an energetic level and plays out in the world of the seen as humanity giving its power to Wetiko authority which uses that power to control the population when the power is only the power the population has handed over. How could it be any other way for billions to be controlled by a relative few? I have had experiences with people possessed by Wetiko and again you can kick its arse if you do it with an open heart. Oh yes – the *heart* which can transform the world of perceived 'matter'.

We are receiver-transmitters and processors of information, but what information and where from? Information is processed into perception in three main areas – the brain, the heart and the belly. These relate to thinking, knowing, and emotion. Wetiko wants us to be head and belly people which means we think within the confines of the Matrix simulation and low-vibrational emotional reaction scrambles balance and perception. A few minutes on social media and you see how emotion is the dominant force. Woke is all emotion and is therefore thought-free and fact-free. Our heart is something different. It *knows* while the head *thinks* and has to try to work it out because it doesn't know. The human energy field has seven prime vortexes which connect us with wider reality (Fig 23). Chakra means ‘wheels of light’ in the Sanskrit language of ancient India. The main ones are: The crown chakra on top of the head; brow (or ‘third eye’) chakra in the centre of the forehead; throat chakra; heart chakra in the centre of the chest; solar plexus chakra below the sternum; sacral chakra beneath the navel; and base chakra at the bottom of the spine. Each one has a particular function or functions. We feel anxiety and nervousness in the belly where the sacral chakra is located and this processes emotion that can affect the colon to give people ‘the shits’ or make them ‘shit scared’ when they are nervous. Chakras all play an important role, but the Mr and Mrs Big is the heart chakra which sits at the centre of the seven, above the chakras that connect us to the ‘physical’ and below those that connect with higher realms (or at least should). Here in the heart chakra we feel love, empathy and compassion – ‘My heart goes out to you’. Those with closed hearts become literally ‘heart-less’ in their attitudes and behaviour (see Bill Gates). Native Americans portrayed Wetiko with what Paul Levy calls a ‘frigid, icy heart, devoid of mercy’ (see Bill Gates).

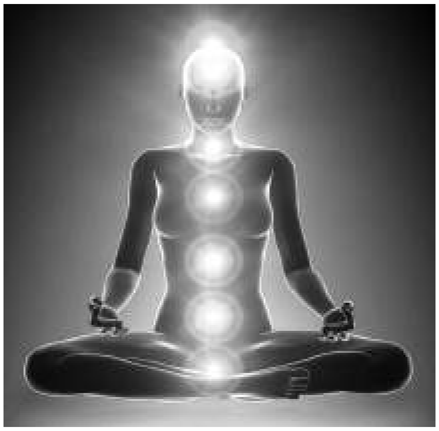


Figure 23: The chakra system which interpenetrates the human energy field. The heart chakra is the governor – or should be.

Wetiko trembles at the thought of heart energy which it cannot infiltrate. The frequency is too high. What it seeks to do instead is close the heart chakra vortex to block its perceptual and energetic influence. Psychopaths have ‘hearts of stone’ and emotionally-damaged people have ‘heartache’ and ‘broken hearts’. The astonishing amount of heart disease is related to heart chakra disruption with its fundamental connection to the 'physical' heart. Dr Tom Cowan has written an outstanding book challenging the belief that the heart is a pump and making the connection between the 'physical' and spiritual heart. Rudolph Steiner who was way ahead of his time said the same about the fallacy that the heart is a pump. *What*? The heart is not a pump? That's crazy, right? Everybody knows that. Read Cowan's *Human Heart, Cosmic Heart* and you will realise that the very idea of the heart as a pump is ridiculous when you see the evidence. How does blood in the feet so far from the heart get pumped horizontally up the body by the heart?? Cowan explains in the book the real reason why blood moves as it does. Our 'physical' heart is used to symbolise love when the source is really the heart vortex or spiritual heart which is our most powerful energetic connection to 'out there' expanded consciousness. That's why we feel *knowing* – intuitive knowing – in the centre of the chest. Knowing doesn't come from a process of thoughts leading to a conclusion. It is there in an instant all in one go. Our heart knows because of its connection to levels of awareness that *do* know. This is the meaning and source of intuition – intuitive *knowing*.

For the last more than 30 years of uncovering the global game and the nature of reality my heart has been my constant antenna for truth and accuracy. An American intelligence insider once said that I had quoted a disinformer in one of my books and yet I had only quoted the part that was true. He asked: 'How do you do that?' By using my heart antenna was the answer and anyone can do it. Heart-centred is how we are meant to be. With a closed heart chakra we withdraw into a closed mind and the bubble of five-sense reality. If you take a moment to focus your attention on the centre of your chest, picture a spinning wheel of light and see it opening and expanding. You will feel it happening, too, and perceptions of the heart like joy and love as the heart impacts on the mind as they interact. The more the chakra opens the more you will feel expressions of heart consciousness and as the process continues, and becomes part of you, insights and knowings will follow. An open heart is connected to that level of awareness that knows all is *One*. You will see from its perspective that the fault-lines that divide us are only illusions to control us. An open heart does not process the illusions of race, creed and sexuality except as brief experiences for a consciousness that is all. Our heart does not see division, only unity (Figs 24 and 25). There's something else, too. Our hearts love to laugh. Mark Twain's quote that says 'The human race has one really effective weapon, and that is laughter' is really a reference to the heart which loves to laugh with the joy of knowing the true nature of infinite reality and that all the madness of human society is an illusion of the mind. Twain also said: 'Against the assault of laughter nothing can stand.' This is so true of Wetiko and the Cult. Their insecurity demands that they be taken seriously and their power and authority acknowledged and feared. We should do nothing of the sort. We should not get aggressive or fearful which their insecurity so desires. We should laugh in their face. Even in their no-face as police come over in their face-nappies and expect to be taken seriously. They don't take themselves seriously looking like that so why should we? Laugh in the face of intimidation. Laugh in the face of tyranny. You will see by its reaction that you have pressed all of its buttons. Wetiko does not know what to do in the face of laughter or when its targets refuse to concede their joy to fear. We have seen many examples during the 'Covid' hoax when people have expressed their energetic power and the string puppets of Wetiko retreat with their tail limp between their knees. Laugh – the world is bloody mad after all and if it's a choice between laughter and tears I know which way I'm going.

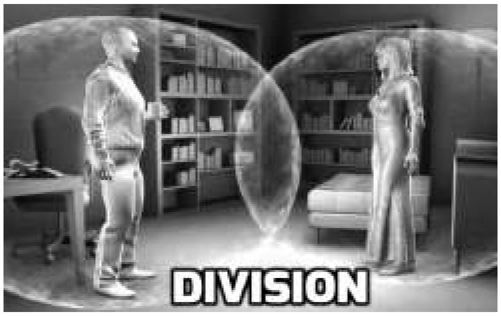


Figure 24: Head consciousness without the heart sees division and everything apart from everything else.

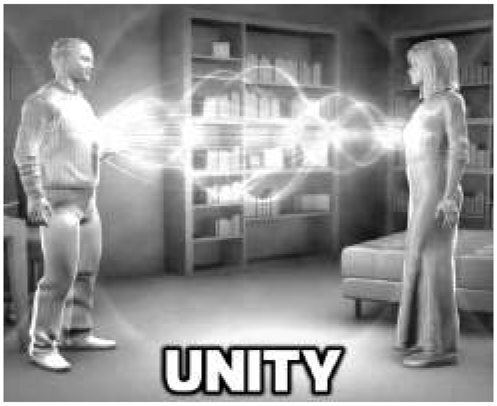


Figure 25: Heart consciousness sees everything as One.

'Vaccines' and the soul

The foundation of Wetiko/Archon control of humans is the separation of incarnate five-sense mind from the infinite 'I' and closing the heart chakra where the True 'I' lives during a human life. The goal has been to achieve complete separation in both cases. I was interested therefore to read an account by a French energetic healer of what she said she experienced with a patient who had been given the 'Covid' vaccine. Genuine energy healers can sense information and consciousness fields at different levels of being which are referred to as 'subtle bodies'. She described treating the patient who later returned after having, without the healer's knowledge, two doses of the 'Covid vaccine'. The healer said:

> I noticed immediately the change, very heavy energy emanating from [the] subtle bodies. The scariest thing was when I was working on the heart chakra, I connected with her soul: it was detached from the physical body, it had no contact and it was, as if it was floating in a state of total confusion: a damage to the consciousness that loses contact with the physical body, i.e. with our biological machine, there is no longer any communication between them.
>
> I continued the treatment by sending light to the heart chakra, the soul of the person, but it seemed that the soul could no longer receive any light, frequency or energy. It was a very powerful experience for me. Then I understood that this substance is indeed used to detach consciousness so that this consciousness can no longer interact through this body that it possesses in life, where there is no longer any contact, no frequency, no light, no more energetic balance or mind.

This would create a human that is rudderless and at the extreme almost zombie-like operating with a fractional state of consciousness at the mercy of Wetiko. I was especially intrigued by what the healer said in the light of the prediction by the highly-informed Rudolf Steiner more than a hundred years ago. He said:

> In the future, we will eliminate the soul with medicine. Under the pretext of a 'healthy point of view', there will be a vaccine by which the human body will be treated as soon as possible directly at birth, so that the human being cannot develop the thought of the existence of soul and Spirit. To materialistic doctors will be entrusted the task of removing the soul of humanity.
>
> As today, people are vaccinated against this disease or that disease, so in the future, children will be vaccinated with a substance that can be produced precisely in such a way that people, thanks to this vaccination, will be immune to being subjected to the 'madness' of spiritual life. He would be extremely smart, but he would not develop a conscience, and that is the true goal of some materialistic circles.

Steiner said the vaccine would detach the physical body from the etheric body (subtle bodies) and 'once the etheric body is detached the relationship between the universe and the etheric body would become extremely unstable, and man would become an automaton'. He said 'the physical body of man must be polished on this Earth by spiritual will – so the vaccine becomes a kind of arymanique (Wetiko) force' and 'man can no longer get rid of a given materialistic feeling'. Humans would then, he said, become 'materialistic of constitution and can no longer rise to the spiritual'. I have been writing for years about DNA being a receiver-transmitter of information that connects us to other levels of reality and these 'vaccines' changing DNA can be likened to changing an antenna and what it can transmit and receive. Such a disconnection would clearly lead to changes in personality and perception. Steiner further predicted the arrival of AI. Big Pharma 'Covid vaccine' makers, expressions of Wetiko, are testing their DNA-manipulating evil on children as I write with a view to giving the 'vaccine' to babies. If it's a soul-body disconnector – and I say that it is or can be – every child would be disconnected from 'soul' at birth and the 'vaccine' would create a closed system in which spiritual guidance from the greater self would play no part. This has been the ambition of Wetiko all along. A Pentagon video from 2005 was leaked of a presentation explaining the development of vaccines to change behaviour by their effect on the brain. Those that believe this is not happening with the 'Covid' genetically-modifying procedure masquerading as a 'vaccine' should make an urgent appointment with Naivety Anonymous. Klaus Schwab wrote in 2018:

> Neurotechnologies enable us to better influence consciousness and thought and to understand many activities of the brain. They include decoding what we are thinking in fine levels of detail through new chemicals and interventions that can influence our brains to correct for errors or enhance functionality.

The plan is clear and only the heart can stop it. With every heart that opens, every mind that awakens, Wetiko is weakened. Heart and love are far more powerful than head and hate and so nothing like a majority is needed to turn this around.

Beyond the Phantom

Our heart is the prime target of Wetiko and so it must be the answer to Wetiko. We *are* our heart which is part of one heart, the infinite heart. Our heart is where the true self lives in a human life behind firewalls of five-sense illusion when an imposter takes its place – *Phantom Self*; but our heart waits patiently to be set free any time we choose to see beyond the Phantom, beyond Wetiko. A Wetikoed Phantom Self can wreak mass death and destruction while the love of forever is locked away in its heart. The time is here to unleash its power and let it sweep away the fear and despair that is Wetiko. Heart consciousness does not seek manipulated, censored, advantage for its belief or religion, its activism and desires. As an expression of the One it treats all as One with the same rights to freedom and opinion. Our heart demands fairness for itself no more than for others. From this unity of heart we can come together in mutual support and transform this Wetikoed world into what reality is meant to be – a place of love, joy, happiness, fairness, justice and freedom. Wetiko has another agenda and that's why the world is as it is, but enough of this nonsense. Wetiko can't stay where hearts are open and it works so hard to keep them closed. Fear is its currency and its food source and love in its true sense has no fear. Why would love have fear when it knows it is *All That Is, Has Been, And Ever Can Be* on an eternal exploration of all possibility? Love in this true sense is not the physical attraction that passes for love. This can be an expression of it, yes, but Infinite Love, a love without condition, goes far deeper to the core of all being. It *is* the core of all being. Infinite realty was born from love beyond the illusions of the simulation. Love infinitely expressed is the knowing that all is One and the swiftly-passing experience of separation is a temporary hallucination. You cannot disconnect from Oneness; you can only *perceive* that you have and withdraw from its influence. This is the most important of all perception trickery by the mind parasite that is Wetiko and the foundation of all its potential for manipulation.

If we open our hearts, open the sluice gates of the mind, and redefine self-identity amazing things start to happen. Consciousness expands or contracts in accordance with self-identity. When true self is recognised as infinite awareness and label self – Phantom Self – is seen as only a series of brief experiences life is transformed. Consciousness expands to the extent that self-identity expands and everything changes. You see unity, not division, the picture, not the pixels. From this we can play the long game. No more is an experience something in and of itself, but a fleeting moment in the eternity of forever. Suddenly people in uniform and dark suits are no longer intimidating. Doing what your heart knows to be right is no longer intimidating and consequences for those actions take on the same nature of a brief experience that passes in the blink of an infinite eye. Intimidation is all in the mind. Beyond the mind there is no intimidation.

An open heart does not consider consequences for what it knows to be right. To do so would be to consider not doing what it knows to be right and for a heart in its power that is never an option. The Renegade Mind is really the Renegade Heart. Consideration of consequences will always provide a getaway car for the mind and the heart doesn’t want one. What is right in the light of what we face today is to stop cooperating with Wetiko in all its forms and to do it without fear or compromise. You cannot compromise with tyranny when tyranny always demands more until it has everything. Life is your perception and you are your destiny. Change your perception and you change your life. Change collective perception and we change the world.

Come on people … One human family, One heart, One goal … *FREEEEEEDOM*!

We must settle for nothing less.

Postscript

The big scare story as the book goes to press is the 'Indian' variant and the world is being deluged with propaganda about the 'Covid catastrophe' in India which mirrors in its lies and misrepresentations what happened in Italy before the first lockdown in 2020.

The *New York Post* published a picture of someone who had 'collapsed in the street from Covid' in India in April, 2021, which was actually taken during a gas leak in May, 2020. Same old, same old. Media articles in mid-February were asking why India had been so untouched by 'Covid' and then as their vaccine rollout gathered pace the alleged 'cases' began to rapidly increase. Indian 'Covid vaccine' maker Bharat Biotech was funded into existence by the Bill and Melinda Gates Foundation (the pair announced their divorce in May, 2021, which is a pity because they so deserve each other). The Indian 'Covid crisis' was ramped up by the media to terrify the world and prepare people for submission to still more restrictions. The scam that worked the first time was being repeated only with far more people seeing through the deceit. Davidicke.com and Ickonic.com have sought to tell the true story of what is happening by talking to people living through the Indian nightmare which has nothing to do with 'Covid'. We posted a letter from 'Alisha' in Pune who told a very different story to government and media mendacity. She said scenes of dying people and overwhelmed hospitals were designed to hide what was really happening – genocide and starvation. Alisha said that millions had already died of starvation during the ongoing lockdowns while government and media were lying and making it look like the 'virus':

> Restaurants, shops, gyms, theatres, basically everything is shut. The cities are ghost towns. Even so-called 'essential' businesses are only open till 11am in the morning. You basically have just an hour to buy food and then your time is up.
>
> Inter-state travel and even inter-district travel is banned. The cops wait at all major crossroads to question why you are traveling outdoors or to fine you if you are not wearing a mask.
>
> The medical community here is also complicit in genocide, lying about hospitals being full and turning away people with genuine illnesses, who need immediate care. They have even created a shortage of oxygen cylinders.

This is the classic Cult modus operandi played out in every country. Alisha said that people who would not have a PCR test not testing for the 'virus' were being denied hospital treatment. She said the people hit hardest were migrant workers and those in rural areas. Most businesses employed migrant workers and with everything closed there were no jobs, no income and no food. As a result millions were dying of starvation or malnutrition. All this was happening under Prime Minister Narendra Modi, a 100-percent asset of the Cult, and it emphasises yet again the scale of pure anti-human evil we are dealing with. Australia banned its people from returning home from India with penalties for trying to do so of up to five years in jail and a fine of £37,000. The manufactured 'Covid' crisis in India was being prepared to justify further fascism in the West. Obvious connections could be seen between the Indian 'vaccine' programme and increased 'cases' and this became a common theme. The Seychelles, the most per capita 'Covid vaccinated' population in the world, went back into lockdown after a 'surge of cases'.

Long ago the truly evil Monsanto agricultural biotechnology corporation with its big connections to Bill Gates devastated Indian farming with genetically-modified crops. Human rights activist Gurcharan Singh highlighted the efforts by the Indian government to complete the job by destroying the food supply to hundreds of millions with 'Covid' lockdowns. He said that 415 million people at the bottom of the disgusting caste system (still going whatever they say) were below the poverty line and struggled to feed themselves every year. Now the government was imposing lockdown at just the time to destroy the harvest. This deliberate policy was leading to mass starvation. People may reel back at the suggestion that a government would do that, but Wetiko-controlled ‘leaders’ are capable of any level of evil. In fact what is described in India is in the process of being instigated worldwide. The food chain and food supply are being targeted at every level to cause world hunger and thus control. Bill Gates is not the biggest owner of farmland in America for no reason and destroying access to food aids both the depopulation agenda and the plan for synthetic ‘food’ already being funded into existence by Gates. Add to this the coming hyper-inflation from the suicidal creation of fake ‘money’ in response to ‘Covid’ and the breakdown of container shipping systems and you have a cocktail that can only lead one way and is meant to. The Cult plan is to crash the entire system to ‘build back better’ with the Great Reset.

‘Vaccine’ transmission

Reports from all over the world continue to emerge of women suffering menstrual and fertility problems after having the fake ‘vaccine’ and of the non-’vaccinated’ having similar problems when interacting with the ‘vaccinated’. There are far too many for ‘coincidence’ to be credible. We’ve had menopausal women getting periods, others having periods stop or not stopping for weeks, passing clots, sometimes the lining of the uterus, breast irregularities, and miscarriages (which increased by 400 percent in parts of the United States). Non-‘vaccinated’ men and children have suffered blood clots and nose bleeding after interaction with the ‘vaccinated’. Babies have died from the effects of breast milk from a ‘vaccinated’ mother. Awake doctors – the small minority – speculated on the cause of non-’vaccinated’ suffering the same effects as the ‘vaccinated’. Was it nanotechnology in the synthetic substance transmitting frequencies or was it a straight chemical bioweapon that was being transmitted between people? I am not saying that some kind of chemical transmission is not one possible answer, but the foundation of all that the Cult does is frequency and this is fertile ground for understanding how transmission can happen. American doctor Carrie Madej, an internal medicine physician and osteopath, has been practicing for the last 20 years, teaching medical students, and she says attending different meetings where the agenda for humanity was discussed. Madej, who operates out of Georgia, did not dismiss other possible forms of transmission, but she focused on frequency in search of an explanation for transmission. She said the Moderna and Pfizer 'vaccines' contained nano-lipid particles as a key component. This was a brand new technology never before used on humanity. 'They're using a nanotechnology which is pretty much little tiny computer bits … nanobots or hydrogel.' Inside the 'vaccines' was 'this sci-fi kind of substance' which suppressed immune checkpoints to get into the cell. I referred to this earlier as the 'Trojan horse' technique that tricks the cell into opening a gateway for the self-replicating synthetic material and while the immune system is artificially suppressed the body has no defences. Madej said the substance served many purposes including an on-demand ability to 'deliver the payload' and using the nano 'computer bits' as biosensors in the body. 'It actually has the ability to accumulate data from your body, like your breathing, your respiration, thoughts, emotions, all kinds of things.'

She said the technology obviously has the ability to operate through Wi-Fi and transmit and receive energy, messages, frequencies or impulses. 'Just imagine you're getting this new substance in you and it can react to things all around you, the 5G, your smart device, your phones.' We had something completely foreign in the human body that had never been launched large scale at a time when we were seeing 5G going into schools and hospitals (plus the Musk satellites) and she believed the 'vaccine' transmission had something to do with this: '… if these people have this inside of them … it can act like an antenna and actually transmit it outwardly as well.' The synthetic substance produced its own voltage and so it could have that kind of effect. This fits with my own contention that the nano receiver-transmitters are designed to connect people to the

Smart Grid and break the receiver-transmitter connection to expanded consciousness. That would explain the French energy healer's experience of the disconnection of body from 'soul' with those who have had the 'vaccine'. The nanobots, self-replicating inside the body, would also transmit the synthetic frequency which could be picked up through close interaction by those who have not been 'vaccinated'. Madej speculated that perhaps it was 5G and increased levels of other radiation that was causing the symptoms directly although interestingly she said that non-'vaccinated' patients had shown improvement when they were away from the 'vaccinated' person they had interacted with. It must be remembered that you can control frequency and energy with your mind and you can consciously create energetic barriers or bubbles with the mind to stop damaging frequencies from penetrating your field. American paediatrician Dr Larry Palevsky said the 'vaccine' was not a 'vaccine' and was never designed to protect from a 'viral' infection. He called it 'a massive, brilliant propaganda of genocide' because they didn't have to inject everyone to get the result they wanted. He said the content of the jabs was able to infuse any material into the brain, heart, lungs, kidneys, liver, sperm and female productive system. 'This is genocide; this is a weapon of mass destruction.' At the same time American colleges were banning students from attending if they didn't have this life-changing and potentially life-ending 'vaccine'. Class action lawsuits must follow when the consequences of this college fascism come to light. As the book was going to press came reports about fertility effects on sperm in 'vaccinated' men which would absolutely fit with what I have been saying and hospitals continued to fill with 'vaccine' reactions. Another question is what about transmission via blood transfusions? The NHS has extended blood donation restrictions from seven days after a 'Covid vaccination' to 28 days after even a sore arm reaction.

I said in the spring of 2020 that the then touted 'Covid vaccine' would be ongoing each year like the flu jab. A year later Pfizer CEO, the appalling Albert Bourla, said people would 'likely' need a 'booster dose' of the 'vaccine' within 12 months of getting 'fully vaccinated' and then a yearly shot. 'Variants will play a key role', he said confirming the point. Johnson & Johnson CEO Alex Gorsky also took time out from his 'vaccine' disaster to say that people may need to be vaccinated against 'Covid-19' each year. UK Health Secretary, the psychopath Matt Hancock, said additional 'boosters' would be available in the autumn of 2021. This is the trap of the 'vaccine passport'. The public will have to accept every last 'vaccine' they introduce, including for the fake 'variants', or it would cease to be valid. The only other way in some cases would be continuous testing with a test not testing for the 'virus' and what is on the swabs constantly pushed up your noise towards the brain every time?

'Vaccines' changing behaviour

I mentioned in the body of the book how I believed we would see gathering behaviour changes in the 'vaccinated' and I am already hearing such comments from the non-'vaccinated' describing behaviour changes in friends, loved ones and work colleagues. This will only increase as the self-replicating synthetic material and nanoparticles expand in body and brain. An article in the *Guardian* in 2016 detailed research at the University of Virginia in Charlottesville which developed a new method for controlling brain circuits associated with complex animal behaviour. The method, dubbed 'magnetogenetics', involves genetically-engineering a protein called ferritin, which stores and releases iron, to create a magnetised substance – 'Magneto' – that can activate specific groups of nerve cells from a distance. This is claimed to be an advance on other methods of brain activity manipulation known as optogenetics and chemogenetics (the Cult has been developing methods of brain control for a long time). The ferritin technique is said to be non-invasive and able to activate neurons 'rapidly and reversibly'. In other words, human thought and perception. The article said that earlier studies revealed how nerve cell proteins 'activated by heat and mechanical pressure can be genetically engineered so that they become sensitive to radio waves and magnetic fields, by attaching them to an iron-storing protein called ferritin, or to inorganic paramagnetic particles'. Sensitive to radio waves and magnetic fields? You mean like 5G, 6G and 7G? This is the human-AI Smart Grid hive mind we are talking about. The *Guardian* article said:

> … the researchers injected Magneto into the striatum of freely behaving mice, a deep brain structure containing dopamine-producing neurons that are involved in reward and motivation, and then placed the animals into an apparatus split into magnetised and non-magnetised sections.
>
> Mice expressing Magneto spent far more time in the magnetised areas than mice that did not, because activation of the protein caused the striatal neurons expressing it to release dopamine, so that the mice found being in those areas rewarding. This shows that Magneto can remotely control the firing of neurons deep within the brain, and also control complex behaviours.

Make no mistake this basic methodology will be part of the 'Covid vaccine' cocktail and using magnetics to change brain function through electromagnetic field frequency activation. The Pentagon is developing a 'Covid vaccine' using ferritin. Magnetics would explain changes in behaviour and why videos are appearing across the Internet as I write showing how magnets stick to the skin at the point of the 'vaccine' shot. Once people take these 'vaccines' anything becomes possible in terms of brain function and illness which will be blamed on 'Covid-19' and 'variants'. Magnetic field manipulation would further explain why the non-'vaccinated' are reporting the same symptoms as the 'vaccinated' they interact with and why those symptoms are reported to decrease when not in their company. Interestingly 'Magneto', a 'mutant', is a character in the Marvel Comic *X-Men* stories with the ability to manipulate magnetic fields and he believes that mutants should fight back against their human oppressors by any means necessary. The character was born Erik Lehnsherr to a Jewish family in Germany.

Cult-controlled courts

The European Court of Human Rights opened the door for mandatory 'Covid-19 vaccines' across the continent when it ruled in a Czech Republic dispute over childhood immunisation that legally enforced vaccination could be 'necessary in a democratic society'. The 17 judges decided that compulsory vaccinations did not breach human rights law. On the face of it the judgement was so inverted you gasp for air. If not having a vaccine infused into your body is not a human right then what is? Ah, but they said human rights law which has been specifically written to delete all human rights at the behest of the state (the Cult). Article 8 of the European Convention on Human Rights relates to the right to a private life. The crucial word here is *'except'*:

> There shall be no interference by a public authority with the exercise of this right EXCEPT such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic wellbeing of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others [My emphasis].

No interference *except* in accordance with the law means there *are* no 'human rights' *except* what EU governments decide you can have at their behest. 'As is necessary in a democratic society' explains that reference in the judgement and 'in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others' gives the EU a coach and horses to ride through 'human rights' and scatter them in all directions. The judiciary is not a check and balance on government extremism; it is a vehicle to enforce it. This judgement was almost laughably predictable when the last thing the Cult wanted was a decision that went against mandatory vaccination. Judges rule over and over again to benefit the system of which they are a part. Vaccination disputes that come before them are invariably delivered in favour of doctors and authorities representing the view of the state which owns the judiciary. Oh, yes, and we have even had calls to stop putting 'Covid-19' on death certificates within 28 days of a 'positive test' because it is claimed the practice makes the 'vaccine' appear not to work. They are laughing at you.

The scale of madness, inhumanity and things to come was highlighted when those not ‘vaccinated’ for ‘Covid’ were refused evacuation from the Caribbean island of St Vincent during massive volcanic eruptions. Cruise ships taking residents to the safety of another island allowed only the ‘vaccinated’ to board and the rest were left to their fate. Even in life and death situations like this we see ‘Covid’ stripping people of their most basic human instincts and the insanity is even more extreme when you think that fake ‘vaccine’-makers are not even claiming their body-manipulating concoctions stop ‘infection’ and ‘transmission’ of a ‘virus’ that doesn’t exist. St Vincent Prime Minister Ralph Gonsalves said: ‘The chief medical officer will be identifying the persons already vaccinated so that we can get them on the ship.’ Note again the power of the chief medical officer who, like Whitty in the UK, will be answering to the World Health Organization. This is the Cult network structure that has overridden politicians who ‘follow the science’ which means doing what WHO-controlled ‘medical officers’ and ‘science advisers’ tell them. Gonsalves even said that residents who were ‘vaccinated’ after the order so they could board the ships would still be refused entry due to possible side effects such as ‘wooziness in the head’. The good news is that if they were woozy enough in the head they could qualify to be prime minister of St Vincent.

Microchipping freedom

The European judgement will be used at some point to justify moves to enforce the ‘Covid’ DNA-manipulating procedure. Sandra Ro, CEO of the Global Blockchain Business Council, told a World Economic Forum event that she hoped ‘vaccine passports’ would help to ‘drive forced consent and standardisation’ of global digital identity schemes: ‘I’m hoping with the desire and global demand for some sort of vaccine passport – so that people can get travelling and working again – [it] will drive forced consent, standardisation, and frankly, cooperation across the world.’ The lady is either not very bright, or thoroughly mendacious, to use the term ‘forced consent’.

You do not ‘consent’ if you are forced – you *submit*. She was describing what the plan has been all along and that’s to enforce a digital identity on every human without which they could not function. ‘Vaccine passports’ are opening the door and are far from the end goal. A digital identity would allow you to be tracked in everything you do in cyberspace and this is the same technique used by Cult-owned China to enforce its social credit system of total control. The ultimate ‘passport’ is planned to be a microchip as my books have warned for nearly 30 years. Those nice people at the Pentagon working for the Cult-controlled Defense Advanced Research Projects Agency (DARPA) claimed in April, 2021, they have developed a microchip inserted under the skin to detect ‘asymptomatic Covid-19 infection’ before it becomes an outbreak and a ‘revolutionary filter’ that can remove the ‘virus’ from the blood when attached to a dialysis machine. The only problems with this are that the ‘virus’ does not exist and people transmitting the ‘virus’ with no symptoms is brain-numbing bullshit. This is, of course, not a ruse to get people to be microchipped for very different reasons. DARPA also said it was producing a one-stop ‘vaccine’ for the ‘virus’ and all ‘variants’. One of the most sinister organisations on Planet Earth is doing this? Better have it then. These people are insane because Wetiko that possesses them is insane.

Researchers from the Salk Institute in California announced they have created an embryo that is part human and part monkey. My books going back to the 1990s have exposed experiments in top secret underground facilities in the United States where humans are being crossed with animal and non-human ‘extraterrestrial’ species. They are now easing that long-developed capability into the public arena and there is much more to come given we are dealing with psychiatric basket cases. Talking of which – Elon Musk’s scientists at Neuralink trained a monkey to play Pong and other puzzles on a computer screen using a joystick and when the monkey made the correct move a metal tube squirted banana smoothie into his mouth which is the basic technique for training humans into unquestioning compliance. Two Neuralink chips were in the monkey’s skull and more than 2,000 wires ‘fanned out’ into its brain. Eventually the monkey played a video game purely with its brain waves. Psychopathic narcissist Musk said the ‘breakthrough’ was a step towards putting Neuralink chips into human skulls and merging minds with artificial intelligence. *Exactly*. This man is so dark and Cult to his DNA.

World Economic Fascism (WEF)

The World Economic Forum is telling you the plan by the statements made at its many and various events. Cult-owned fascist YouTube CEO Susan Wojcicki spoke at the 2021 WEF Global Technology Governance Summit (see the name) in which 40 governments and 150 companies met to ensure ‘the responsible design and deployment of emerging technologies’. Orwellian translation: ‘Ensuring the design and deployment of long-planned technologies will advance the Cult agenda for control and censorship.’ Freedom-destroyer and Nuremberg-bound Wojcicki expressed support for tech platforms like hers to censor content that is ‘technically legal but could be harmful’. Who decides what is ‘harmful’? She does and they do. ‘Harmful’ will be whatever the Cult doesn’t want people to see and we have legislation proposed by the UK government that would censor content on the basis of ‘harm’ no matter if the information is fair, legal and provably true. Make that *especially* if it is fair, legal and provably true. Wojcicki called for a global coalition to be formed to enforce content moderation standards through automated censorship. This is a woman and mega-censor so self-deluded that she shamelessly accepted a ‘free expression’ award – *Wojcicki* – in an event sponsored by her own *YouTube*. They have no shame and no self-awareness.

You know that ‘Covid’ is a scam and Wojcicki a Cult operative when YouTube is censoring medical and scientific opinion purely on the grounds of whether it supports or opposes the Cult ‘Covid’ narrative. Florida governor Ron DeSantis compiled an expert panel with four professors of medicine from Harvard, Oxford, and Stanford Universities who spoke against forcing children and vaccinated people to wear masks. They also said there was no proof that lockdowns reduced spread or death rates of 'Covid-19'. Cult-gofer Wojcicki and her YouTube deleted the panel video 'because it included content that contradicts the consensus of local and global health authorities regarding the efficacy of masks to prevent the spread of Covid-19'. This 'consensus' refers to what the Cult tells the World Health Organization to say and the WHO tells 'local health authorities' to do. Wojcicki knows this, of course. The panellists pointed out that censorship of scientific debate was responsible for deaths from many causes, but Wojcicki couldn't care less. She would not dare go against what she is told and as a disgrace to humanity she wouldn't want to anyway. The UK government is seeking to pass a fascist 'Online Safety Bill' to specifically target with massive fines and other means non-censored video and social media platforms to make them censor 'lawful but harmful' content like the Cult-owned Facebook, Twitter, Google and YouTube. What is 'lawful but harmful' would be decided by the fascist Blair-created Ofcom.

Another WEF obsession is a cyber-attack on the financial system and this is clearly what the Cult has planned to take down the bank accounts of everyone – except theirs. Those that think they have enough money for the Cult agenda not to matter to them have got a big lesson coming if they continue to ignore what is staring them in the face. The World Economic Forum, funded by Gates and fronted by Klaus Schwab, announced it would be running a 'simulation' with the Russian government and global banks of just such an attack called Cyber Polygon 2021. What they simulate – as with the 'Covid' Event 201 – they plan to instigate. The WEF is involved in a project with the Cult-owned Carnegie Endowment for International Peace called the WEF-Carnegie Cyber Policy Initiative which seeks to merge Wall Street banks, 'regulators' (I love it) and intelligence agencies to 'prevent' (arrange and allow) a cyber-attack that would bring down the global financial system as long planned by those that control the WEF and the Carnegie operation. The Carnegie Endowment for International Peace sent an instruction to First World

War US President Woodrow Wilson not to let the war end before society had been irreversibly transformed.

The Wuhan lab diversion

As I close, the Cult-controlled authorities and lapdog media are systematically pushing 'the virus was released from the Wuhan lab' narrative. There are two versions – it happened by accident and it happened on purpose. Both are nonsense. The perceived existence of the never-shown-to-exist 'virus' is vital to sell the impression that there is actually an infective agent to deal with and to allow the endless potential for terrifying the population with 'variants' of a 'virus' that does not exist. The authorities at the time of writing are going with the 'by accident' while the alternative media is promoting the 'on purpose'. Cable news host Tucker Carlson who has questioned aspects of lockdown and 'vaccine' compulsion has bought the Wuhan lab story. 'Everyone now agrees' he said. Well, I don't and many others don't and the question is *why* does the system and its media suddenly 'agree'? When the media moves as one unit with a narrative it is always a lie – witness the hour by hour mendacity of the 'Covid' era. Why would this Cult-owned combination which has unleashed lies like machine gun fire suddenly 'agree' to tell the truth??

Much of the alternative media is buying the lie because it fits the conspiracy narrative, but it's the *wrong* conspiracy. The real conspiracy is that *there is no virus* and that is what the Cult is desperate to hide. The idea that the 'virus' was released by accident is ludicrous when the whole 'Covid' hoax was clearly long-planned and waiting to be played out as it was so fast in accordance with the Rockefeller document and Event 201. So they prepared everything in detail over decades and then sat around strumming their fingers waiting for an 'accidental' release from a bio-lab? *What??* It's crazy. Then there's the 'on purpose' claim. You want to circulate a 'deadly virus' and hide the fact that you've done so and you release it down the street from the highest-level bio-lab in China? I repeat – *What??*

You would release it far from that lab to stop any association being made. But, no, we'll do it in a place where the connection was certain to be made. Why would you need to scam 'cases' and 'deaths' and pay hospitals to diagnose 'Covid-19' if you had a real 'virus'? What are sections of the alternative media doing believing this crap? Where were all the mass deaths in Wuhan from a 'deadly pathogen' when the recovery to normal life after the initial propaganda was dramatic in speed? Why isn't the 'deadly pathogen' now circulating all over China with bodies in the street? Once again we have the technique of tell them what they want to hear and they will likely believe it. The alternative media has its 'conspiracy' and with Carlson it fits with his 'China is the danger' narrative over years. China *is* a danger as a global Cult operations centre, but not for this reason. The Wuhan lab story also has the potential to instigate conflict with China when at some stage the plan is to trigger a Problem-Reaction-Solution confrontation with the West. Question everything – *everything* – and especially when the media agrees on a common party line.

Third wave ... fourth wave ... fifth wave ...

As the book went into production the world was being set up for more lockdowns and a 'third wave' supported by invented 'variants' that were increasing all the time and will continue to do so in public statements and computer programs, but not in reality. India became the new Italy in the 'Covid' propaganda campaign and we were told to be frightened of the new 'Indian strain'. Somehow I couldn't find it within myself to do so. A document produced for the UK government entitled 'Summary of further modelling of easing of restrictions – Roadmap Step 2' declared that a third wave was inevitable (of course when it's in the script) and it would be the fault of children and those who refuse the health-destroying fake 'Covid vaccine'. One of the computer models involved came from the Cult-owned *Imperial College* and the other from Warwick University which I wouldn't trust to tell me the date in a calendar factory. The document states that both models presumed extremely high uptake of the ‘Covid vaccines’ and didn’t allow for ‘variants’. The document states: ‘The resurgence is a result of some people (mostly children) being ineligible for vaccination; others choosing not to receive the vaccine; and others being vaccinated but not perfectly protected.’ The mendacity takes the breath away. Okay, blame those with a brain who won’t take the DNA-modifying shots and put more pressure on children to have it as ‘trials’ were underway involving children as young as six months with parents who give insanity a bad name. Massive pressure is being put on the young to have the fake ‘vaccine’ and child age consent limits have been systematically lowered around the world to stop parents intervening. Most extraordinary about the document was its claim that the ‘third wave’ would be driven by ‘the resurgence in both hospitalisations and deaths … dominated by *those that have received two doses of the vaccine*, comprising around 60-70% of the wave respectively’. The predicted peak of the ‘third wave’ suggested 300 deaths per day with 250 of them *fully ‘vaccinated’ people*. How many more lies do acquiescers need to be told before they see the obvious? Those who took the jab to ‘protect themselves’ are projected to be those who mostly get sick and die? So what’s in the ‘vaccine’? The document went on:

> It is possible that a summer of low prevalence could be followed by substantial increases in incidence over the following autumn and winter. Low prevalence in late summer should not be taken as an indication that SARS-CoV-2 has retreated or that the population has high enough levels of immunity to prevent another wave.

They are telling you the script and while many British people believed ‘Covid’ restrictions would end in the summer of 2021 the government was preparing for them to be ongoing. Authorities were awarding contracts for ‘Covid marshals’ to police the restrictions with contracts starting in July, 2021, and going through to January 31st, 2022, and the government was advertising for ‘Media Buying Services’ to secure media propaganda slots worth a potential £320 million for ‘Covid-19 campaigns’ with a contract not ending until March, 2022. The recipient – via a list of other front companies – was reported to be American media marketing giant Omnicom Group

Inc. While money is no object for 'Covid' the UK waiting list for all other treatment – including life-threatening conditions – passed 4.5 million. Meantime the Cult is seeking to control all official 'inquiries' to block revelations about what has really been happening and why. It must not be allowed to – we need Nuremberg jury trials in every country. The cover-up doesn't get more obvious than appointing ultra-Zionist professor Philip Zelikow to oversee two dozen US virologists, public health officials, clinicians, former government officials and four American 'charitable foundations' to 'learn the lessons' of the 'Covid' debacle. The personnel will be those that created and perpetuated the 'Covid' lies while Zelikow is the former executive director of the 9/11 Commission who ensured that the truth about those attacks never came out and produced a report that must be among the most mendacious and manipulative documents ever written – see *The Trigger* for the detailed exposure of the almost unimaginable 9/11 story in which Sabbatians can be found at every level.

Passive no more

People are increasingly challenging the authorities with amazing numbers of people taking to the streets in London well beyond the ability of the Face-Nappies to stop them. Instead the Nappies choose situations away from the mass crowds to target, intimidate, and seek to promote the impression of 'violent protestors'. One such incident happened in London's Hyde Park. Hundreds of thousands walking through the streets in protest against 'Covid' fascism were ignored by the Cult-owned BBC and most of the rest of the mainstream media, but they delighted in reporting how police were injured in 'clashes with protestors'. The truth was that a group of people gathered in Hyde Park at the end of one march when most had gone home and they were peacefully having a good time with music and chat. Face-Nappies who couldn't deal with the full-march crowd then waded in with their batons and got more than they bargained for. Instead of just standing for this criminal brutality the crowd used their numerical superiority to push the Face-Nappies out of the park. Eventually the Nappies turned and ran. Unfortunately two or three idiots in the crowd threw drink cans striking two officers which gave the media and the government the image they wanted to discredit the 99.9999 percent who were peaceful. The idiots walked straight into the trap and we must always be aware of potential agent provocateurs used by the authorities to discredit their targets.

This response from the crowd – the can people apart – must be a turning point when the public no longer stand by while the innocent are arrested and brutally attacked by the Face-Nappies. That doesn't mean to be violent, that's the last thing we need. We'll leave the violence to the Face-Nappies and government. But it does mean that when the Face-Nappies use violence against peaceful people the numerical superiority is employed to stop them and make citizen's arrests or Common Law arrests for a breach of the peace. The time for being passive in the face of fascism is over.

We are the many, they are the few, and we need to make that count before there is no freedom left and our children and grandchildren face an ongoing fascist nightmare.

COME ON PEOPLE – IT'S TIME.

One final thought …

The power of love
A force from above
Cleaning my soul
Flame on burn desire
Love with tongues of fire
Purge the soul
Make love your goal

I’ll protect you from the hooded claw
Keep the vampires from your door
When the chips are down I’ll be around
With my undying, death-defying
Love for you

Envy will hurt itself
Let yourself be beautiful
Sparkling love, flowers
And pearls and pretty girls
Love is like an energy
Rushin’ rushin’ inside of me

This time we go sublime
Lovers entwine, divine, divine,
Love is danger, love is pleasure
Love is pure – the only treasure

I’m so in love with you
Purge the soul
Make love your goal

The power of love
A force from above
Cleaning my soul
The power of love
A force from above
A sky-scraping dove

Flame on burn desire
Love with tongues of fire
Purge the soul
Make love your goal

Frankie Goes To Hollywood

APPENDIX

Cowan-Kaufman-Morell Statement on Virus Isolation (SOVI)

Isolation: The action of isolating; the fact or condition of being isolated or standing alone; separation from other things or persons; solitariness
Oxford English Dictionary

The controversy over whether the SARS-CoV-2 virus has ever been isolated or purified continues. However, using the above definition, common sense, the laws of logic and the dictates of science, any unbiased person must come to the conclusion that the SARS-CoV-2 virus has never been isolated or purified. As a result, no confirmation of the virus' existence can be found. The logical, common sense, and scientific consequences of this fact are:

- the structure and composition of something not shown to exist can't be known, including the presence, structure, and function of any hypothetical spike or other proteins;
- the genetic sequence of something that has never been found can't be known;
- "variants" of something that hasn't been shown to exist can't be known;
- it's impossible to demonstrate that SARS-CoV-2 causes a disease called Covid-19.

In as concise terms as possible, here's the proper way to isolate, characterize and demonstrate a new virus. First, one takes samples (blood, sputum, secretions) from many people (e.g. 500) with symptoms which are unique and specific enough to characterize an illness. Without mixing these samples with ANY tissue or products that also contain genetic material, the virologist macerates, filters and ultracentrifuges i.e. *purifies* the specimen. This common virology technique, done for decades to isolate bacteriophages[1] and so-called giant viruses in every virology lab, then allows the virologist to demonstrate with electron microscopy thousands of identically sized and shaped particles. These particles are the isolated and purified virus.

These identical particles are then checked for uniformity by physical and/or microscopic techniques. Once the purity is determined, the particles may be further characterized. This would include examining the structure, morphology, and chemical composition of the particles. Next, their genetic makeup is characterized by extracting the genetic material directly from the purified particles and using genetic-sequencing techniques, such as Sanger sequencing, that have also been around for decades. Then one does an analysis to confirm that these uniform particles are exogenous (outside) in origin as a virus is conceptualized to be, and not the normal breakdown products of dead and dying tissues.[2] (As of May 2020, we know that virologists have no way to determine whether the particles they're seeing are viruses or just normal breakdown products of dead and dying tissues.)[3]

---

[1] Isolation, characterization and analysis of bacteriophages from the haloalkaline lake Elmenteita, KenyaJuliah Khayeli Akhwale et al, PLOS One, Published: April 25, 2019. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0215734 – accessed 2/15/21

---

[2] "Extracellular Vesicles Derived From Apoptotic Cells: An Essential Link Between Death and Regeneration," Maojiao Li1 et al, Frontiers in Cell and Developmental Biology, 2020 October 2. https://www.frontiersin.org/articles/10.3389/fcell.2020.573511/full – accessed 2/15/21

If we have come this far then we have fully isolated, characterized, and genetically sequenced an exogenous virus particle. However, we still have to show it is causally related to a disease. This is carried out by exposing a group of healthy subjects (animals are usually used) to this isolated, purified virus in the manner in which the disease is thought to be transmitted. If the animals get sick with the same disease, as confirmed by clinical and autopsy findings, one has now shown that the virus actually causes a disease. This demonstrates infectivity and transmission of an infectious agent.

None of these steps has even been attempted with the SARS-CoV-2 virus, nor have all these steps been successfully performed for any so-called pathogenic virus. Our research indicates that a single study showing these steps does not exist in the medical literature.

Instead, since 1954, virologists have taken unpurified samples from a relatively few people, often less than ten, with a similar disease. They then minimally process this sample and inoculate this unpurified sample onto tissue culture containing usually four to six other types of material – all of which contain identical genetic material as to what is called a "virus." The tissue culture is starved and poisoned and naturally disintegrates into many types of particles, some of which contain genetic material. Against all common sense, logic, use of the English language and scientific integrity, this process is called "virus isolation." This brew containing fragments of genetic material from many sources is then subjected to genetic analysis, which then creates in a computer-simulation process the alleged sequence of the alleged virus, a so called in silico genome. At no time is an actual virus confirmed by electron microscopy. At no time is a genome extracted and sequenced from an actual virus. This is scientific fraud.

---

3 "The Role of Extraellular Vesicles as Allies of HIV, HCV and SARS Viruses," Flavia Giannessi, et al, Viruses, 2020 May

The observation that the unpurified specimen — inoculated onto tissue culture along with toxic antibiotics, bovine fetal tissue, amniotic fluid and other tissues — destroys the kidney tissue onto which it is inoculated is given as evidence of the virus' existence and pathogenicity. This is scientific fraud.

From now on, when anyone gives you a paper that suggests the SARS-CoV-2 virus has been isolated, please check the methods sections. If the researchers used Vero cells or any other culture method, you know that their process was not isolation. You will hear the following excuses for why actual isolation isn't done:

1. There were not enough virus particles found in samples from patients to analyze.
2. Viruses are intracellular parasites; they can't be found outside the cell in this manner.

If No. 1 is correct, and we can't find the virus in the sputum of sick people, then on what evidence do we think the virus is dangerous or even lethal? If No. 2 is correct, then how is the virus spread from person to person? We are told it emerges from the cell to infect others. Then why isn't it possible to find it?

Finally, questioning these virology techniques and conclusions is not some distraction or divisive issue. Shining the light on this truth is essential to stop this terrible fraud that humanity is confronting. For, as we now know, if the virus has never been isolated, sequenced or shown to cause illness, if the virus is imaginary, then why are we wearing masks, social distancing and putting the whole world into prison?

Finally, if pathogenic viruses don't exist, then what is going into those injectable devices erroneously called "vaccines," and what is their purpose? This scientific question is the most urgent and relevant one of our time.

We are correct. The SARS-CoV2 virus does not exist.

Sally Fallon Morell, MA
Dr. Thomas Cowan, MD
Dr. Andrew Kaufman, MD

Bibliography

Alinsky, Saul: *Rules for Radicals* (Vintage, 1989)

Antelman, Rabbi Marvin: *To Eliminate the Opiate* (Zahavia, 1974)

Bastardi, Joe: *The Climate Chronicles* (Relentless Thunder Press, 2018)

Cowan, Tom: *Human Heart, Cosmic Heart* (Chelsea Green Publishing, 2016)

Cowan, Tom, and Fallon Morell, Sally: *The Contagion Myth* (Skyhorse Publishing, 2020)

Forbes, Jack D: *Columbus And Other Cannibals – The Wetiko Disease of Exploitation, Imperialism, and Terrorism* (Seven Stories Press, 2008 – originally published in 1979)

Gates, Bill: *How to Avoid a Climate Disaster: The Solutions We Have and the Breakthroughs We Need* (Allen Lane, 2021)

Huxley, Aldous: *Brave New World* (Chatto & Windus, 1932)

Köhnlein, Dr Claus, and Engelbrecht, Torsten: *Virus Mania* (emu-Vertag, Lahnstein, 2020)

Lanza, Robert, and Berman, Bob: *Biocentrism* (BenBella Books, 2010)

Lash, John Lamb: *Not In His Image* (Chelsea Green Publishing, 2006)

Lester, Dawn, and Parker, David: *What Really Makes You Ill – Why everything you thought you knew about disease is wrong* (Independently Published, 2019)

Levy, Paul: *Dispelling Wetiko, Breaking the Spell of Evil* (North Atlantic Books, 2013)

Marx, Karl: *A World Without Jews* (Philosophical Library, first edition, 1959)

Mullis, Kary: *Dancing Naked in the Mine Field* (Bloomsbury, 1999)

O'Brien, Cathy: *Trance-Formation of America* (Reality Marketing, 1995)

Scholem, Gershon: *The Messianic Idea in Judaism* (Schocken Books, 1994)

Schwab, Klaus, and Davis, Nicholas: *Shaping the Future of the Fourth Industrial Revolution: A guide to building a better world* (Penguin Books, 2018)

Schwab, Klaus: *The Great Reset* (Agentur Schweiz, 2020)

Sunstein, Cass and Thaler, Richard: *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Penguin, 2009)

Swan, Shanna: *Count Down: How Our Modern World Is Threatening Sperm Counts, Altering Male and Female Reproductive Development and Imperiling the Future of the Human Race* (Scribner, 2021)

Tegmark, Max: *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality* (Penguin, 2015)

Velikovsky, Immanuel: *Worlds in Collision* (Paradigma, 2009)

Wilton, Robert: *The Last Days of the Romanovs* (Blurb, 2018, first published 1920)

Index

A

abusive relationships
  blaming themselves, abused as ref1
  children ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
  conspiracy theories ref1
  domestic abuse ref1, ref2
  economic abuse and dependency ref1
  isolation ref1
  physical abuse ref1
  psychological abuse ref1
  signs of abuse ref1
addiction
  alcoholism ref1
  frequencies ref1
  substance abuse ref1, ref2
  technology ref1, ref2, ref3
Adelson, Sheldon ref1, ref2, ref3
Agenda 21/Agenda 2030 (UN) ref1, ref2, ref3, ref4
AIDs/HIV ref1
  causal link between HIV and AIDs ref1, ref2
retroviruses ref1
  testing ref1, ref2
  trial-run for Covid-19, as ref1, ref2
aliens/extraterrestrials ref1, ref2
aluminium ref1
Amazon ref1, ref2, ref3

amplification cycles ref1, ref2
anaphylactic shock ref1, ref2, ref3, ref4
animals ref1, ref2, ref3
antibodies ref1, ref2, ref3, ref4, ref5
Antifa ref1, ref2, ref3, ref4
antigens ref1, ref2
anti-Semitism ref1, ref2, ref3
Archons ref1, ref2
  consciousness ref1, ref2, ref3
  energy ref1, ref2, ref3
  ennoia ref1
  genetic manipulation ref1, ref2
  inversion ref1, ref2, ref3
  lockdowns ref1
  money ref1
  radiation ref1
  religion ref1, ref2
  technology ref1, ref2, ref3
  Wetiko factor ref1, ref2, ref3, ref4
artificial intelligence (AI) ref1
army made up of robots ref1, ref2
  Human 2.0 ref1, ref2
  Internet ref1
  MHRA ref1
  Morgellons fibres ref1, ref2
  Smart Grid ref1
  Wetiko factor ref1
asymptomatic, Covid-19 as ref1, ref2, ref3
aviation industry ref1

B

banking, finance and money ref1, ref2, ref3
  2008 crisis ref1, ref2
  boom and bust ref1
  cashless digital money systems ref1
  central banks ref1
  credit ref1
  digital currency ref1
  fractional reserve lending ref1
  Great Reset ref1
  guaranteed income ref1, ref2, ref3
  Human 2.0 ref1
  incomes, destruction of ref1, ref2
  interest ref1
  one per cent ref1, ref2
  scams ref1
BBC ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
Becker-Phelps, Leslie ref1
Behavioural Insights Team (BIT) (Nudge Unit) ref1, ref2, ref3
behavioural scientists *and* **psychologists, advice from** ref1, ref2
Bezos, Jeff ref1, ref2, ref3, ref4
Biden, Hunter ref1
Biden, Joe ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11, ref12, ref13, ref14, ref15, ref16, ref17
Big Pharma
  cholesterol ref1
  health professionals ref1, ref2
  immunity from prosecution in US ref1
  vaccines ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
  Wetiko factor ref1, ref2
  WHO ref1, ref2, ref3
Bill and Melinda Gates Foundation ref1, ref2, ref3, ref4, ref5, ref6, ref7

billionaires ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9 ref10, ref11
bird flu (H5N1) ref1
Black Lives Matter (BLM) ref1, ref2, ref3, ref4, ref5
Blair, Tony ref1, ref2, ref3, ref4, ref5, ref6, ref7
Brin, Sergei ref1, ref2, ref3, ref4, ref5, ref6, ref7
British Empire ref1
Bush, George HW ref1, ref2
Bush, George W ref1, ref2, ref3, ref4
Byrd, Robert ref1

C

Canada
  Global Cult ref1
  hate speech ref1
  internment ref1
  masks ref1
  old people ref1
  SARS-COV-2 ref1
  satellites ref1
  vaccines ref1
  wearable technology ref1
Capitol Hill riot ref1, ref2
  agents provocateur ref1
  Antifa ref1
  Black Lives Matter (BLM) ref1, ref2
  QAnon ref1
  security precautions, lack of ref1, ref2, ref3
carbon dioxide ref1, ref2
care homes, deaths in ref1, ref2
cashless digital money systems ref1
censorship ref1, ref2, ref3, ref4, ref5

  fact-checkers ref1
  masks ref1
  media ref1, ref2
  private messages ref1
  social media ref1, ref2, ref3, ref4, ref5, ref6
  transgender persons ref1
  vaccines ref1, ref2, ref3
  Wokeness ref1

Centers for Disease Control (CDC) (United States) ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11, ref12, ref13

centralisation ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8

chakras ref1

change agents ref1, ref2, ref3

chemtrails ref1, ref2, ref3

chief medical officers and scientific advisers ref1, ref2, ref3, ref4, ref5, ref6

children *see also* **young people**
  abuse ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
  care, taken into ref1, ref2, ref3
  education ref1, ref2, ref3, ref4
  energy ref1
  family courts ref1
  hand sanitisers ref1
  human sacrifice ref1
  lockdowns ref1, ref2, ref3

masks ref1, ref2, ref3, ref4, ref5
  mental health ref1
  old people ref1
  parents, replacement of ref1, ref2
  Psyop (psychological operation), Covid as a ref1, ref2
  reframing ref1
  smartphone addiction ref1

  - social distancing and isolation ref1
  - social media ref1
  - transgender persons ref1, ref2
  - United States ref1
  - vaccines ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
  - Wetiko factor ref1

China ref1, ref2, ref3, ref4
  - anal swab tests ref1
  - **Chinese Revolution** ref1, ref2, ref3
  - digital currency ref1
  - Global Cult ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9
  - guaranteed income ref1
  - Imperial College ref1
  - Israel ref1
  - lockdown ref1, ref2
  - masculinity crisis ref1
  - masks ref1
  - media ref1
  - origins of virus in China ref1, ref2, ref3, ref4, ref5
  - pollution causing respiratory diseases ref1
  - Sabbatians ref1, ref2
  - Smart Grid ref1, ref2
  - social credit system ref1
  - testing ref1, ref2
  - United States ref1, ref2
  - vaccines ref1, ref2
  - Wetiko factor ref1
  - wet market conspiracy ref1
  - Wuhan ref1, ref2, ref3, ref4, ref5, ref6, ref7

cholesterol ref1, ref2

Christianity ref1, ref2, ref3, ref4, ref5
  - criticism ref1
  - cross, inversion of the ref1

  Nag Hammadi texts ref1, ref2, ref3
  Roman Catholic Church ref1, ref2
  Sabbatians ref1, ref2
  Satan ref1, ref2, ref3, ref4
  Wokeness ref1
class ref1, ref2
climate change hoax ref1, ref2, ref3, ref4, ref5
  Agenda 21/Agenda 2030 ref1, ref2, ref3
  carbon dioxide ref1, ref2
  Club of Rome ref1, ref2, ref3, ref4, ref5
  fear ref1
  funding ref1
  Global Cult ref1
  green new deals ref1
  green parties ref1
  inversion ref1
  perception, control of ref1
  PICC ref1
  reframing ref1
  temperature, increases in ref1
  United Nations ref1, ref2
  Wikipedia ref1
  Wokeness ref1, ref2
Clinton, Bill ref1, ref2, ref3, ref4, ref5, ref6
Clinton, Hillary ref1, ref2, ref3
the cloud ref1, ref2, ref3, ref4, ref5, ref6, ref7
Club of Rome and climate change hoax ref1, ref2, ref3, ref4, ref5
cognitive therapy ref1
Cohn, Roy ref1
Common Law ref1
  Admiralty Law ref1
  arrests ref1, ref2

  contractual law, Statute Law as ref1
  corporate entities, people as ref1
  legalese ref1
  sea, law of the ref1
  Statute Law ref1

Common Purpose leadership programme ref1, ref2

communism ref1, ref2

co-morbidities ref1

computer-generated virus,

Covid-19 as ref1, ref2, ref3

computer models ref1, ref2, ref3, ref4, ref5

connections ref1, ref2, ref3, ref4

consciousness ref1, ref2, ref3, ref4
  Archons ref1, ref2, ref3
  expanded ref1, ref2, ref3, ref4, ref5, ref6, ref7
  experience ref1
  heart ref1
  infinity ref1, ref2
  religion ref1, ref2
  self-identity ref1
  simulation thesis ref1
  vaccines ref1
  Wetiko factor ref1, ref2

conspiracy theorists ref1, ref2, ref3, ref4, ref5

contradictory rules ref1

contrails ref1

Corman-Drosten test ref1, ref2, ref3, ref4

countermimicry ref1, ref2, ref3

Covid-19 vaccines *see* vaccines

Covidiots ref1, ref2

Cowan, Tom ref1, ref2, ref3, ref4

crimes against humanity ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8

cyber-operations ref1
cyberwarfare ref1

D

DARPA (Defense Advanced Research Projects Agency) ref1
deaths
  care homes ref1
  certificates ref1, ref2, ref3, ref4
  mortality rate ref1
  post-mortems/autopsies ref1
  recording ref1, ref2, ref3, ref4, ref5, ref6, ref7
  vaccines ref1, ref2, ref3, ref4, ref5
deceit
  pyramid of deceit ref1, ref2
  sequence of deceit ref1
decoding ref1, ref2, ref3
dehumanisation ref1, ref2, ref3
Delphi technique ref1
democracy ref1
dependency ref1, ref2, ref3, ref4, ref5
Descartes, René ref1
DNA
  numbers ref1
  vaccines ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
DNR (do not resuscitate)
  orders ref1
domestic abuse ref1, ref2
downgrading of Covid-19 ref1
Drosten, Christian ref1, ref2, ref3, ref4, ref5, ref6, ref7
Duesberg, Peter ref1, ref2

E

economic abuse ref1
Edmunds, John ref1, ref2
education ref1, ref2, ref3, ref4
electromagnetic spectrum ref1, ref2
Enders, John ref1
energy
  Archons ref1, ref2, ref3
  children and young people ref1
  consciousness ref1
  decoding ref1
  frequencies ref1, ref2, ref3, ref4
  heart ref1
  human energy field ref1
  source, humans as an energy ref1, ref2
  vaccines ref1
  viruses ref1
ennoia ref1
Epstein, Jeffrey ref1, ref2
eternal 'I' ref1, ref2
ethylene oxide ref1
European Union ref1, ref2, ref3, ref4
Event **ref1** *and* **Bill Gates** ref2
exosomes, Covid-19 as natural defence mechanism called ref1
experience ref1, ref2
Extinction Rebellion ref1, ref2

F

Facebook
  addiction ref1, 448–50
  Facebook

Archons ref1
censorship ref1, ref2, ref3
hate speech ref1
monopoly, as ref1
private messages, censorship of ref1
Sabbatians ref1
United States election fraud ref1
vaccines ref1
Wetiko factor ref1
fact-checkers ref1
Fauci, Anthony ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11, ref12
fear ref1, ref2, ref3, ref4
climate change ref1
computer models ref1
conspiracy theories ref1
empty hospitals ref1
Italy ref1, ref2, ref3
lockdowns ref1, ref2, ref3, ref4
masks ref1, ref2
media ref1, ref2
medical staff ref1
Psyop (psychological operation), Covid as a ref1
Wetiko factor ref1, ref2
female infertility ref1
Fermi Paradox ref1
Ferguson, Neil ref1, ref2, ref3, ref4, ref5, ref6, ref7
fertility, decline in ref1
The Field ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
finance *see* **banking, finance and money**
five-senses ref1, ref2
Archons ref1, ref2, ref3

censorship ref1
consciousness, expansion of ref1, ref2, ref3, ref4, ref5, ref6
decoding ref1
education ref1, ref2
the Field ref1, ref2
God, personification of ref1
infinity ref1, ref2
media ref1
paranormal ref1
perceptual programming ref1, ref2
Phantom Self ref1
pneuma not nous, using ref1
reincarnation ref1
self-identity ref1
Wetiko factor ref1, ref2, ref3, ref4, ref5, ref6
5G ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
Floyd, George and protests, killing of ref1
flu, re-labelling of ref1, ref2, ref3
food and water, control of ref1, ref2
Freemasons ref1, ref2, ref3, ref4, ref5, ref6
Frei, Rosemary ref1
frequencies
addictions ref1
Archons ref1, ref2, ref3
awareness ref1
chanting and mantras ref1
consciousness ref1
decoding ref1, ref2
education ref1
electromagnetic (EMF) frequencies ref1
energy ref1, ref2, ref3, ref4
fear ref1

  the Field ref1, ref2 5G ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
  five-senses ref1, ref2
  ghosts ref1
  Gnostics ref1
  hive-minds ref1
  human, meaning of ref1
  light ref1, ref2
  love ref1, ref2
  magnetism ref1
  perception ref1
  reality ref1, ref2, ref3
  simulation ref1
  terror ref1
  vaccines ref1
  Wetiko ref1, ref2, ref3
Fuellmich, Reiner ref1, ref2, ref3
furlough/rescue payments ref1

G

Gallo, Robert ref1, ref2, ref3
Gates, Bill
  Archons ref1, ref2, ref3
  climate change ref1, ref2, ref3, ref4
  Daily Pass tracking system ref1
  Epstein ref1
  fascism ref1
  five senses ref1
  GAVI ref1
  Great Reset ref1
  GSK ref1
  Imperial College ref1, ref2
  Johns Hopkins University ref1, ref2, ref3

  lockdowns ref1, ref2
  masks ref1
  Nuremberg trial, proposal for ref1, ref2
  Rockefellers ref1, ref2
  social distancing and isolation ref1
  Sun, dimming the ref1
  synthetic meat ref1, ref2
  vaccines ref1, ref2, ref3, ref4, ref5, ref6, ref7
  Wellcome Trust ref1
  Wetiko factor ref1, ref2, ref3
  WHO ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
  Wokeness ref1
  World Economic Forum ref1, ref2, ref3, ref4
Gates, Melinda ref1, ref2, ref3
GAVI vaccine alliance ref1
genetics, manipulation of ref1, ref2, ref3
Germany ref1, ref2, ref3, ref4, ref5, ref6 *see also* **Nazi Germany**
Global Cult ref1, ref2, ref3, ref4, ref5
  anti-human, why Global Cult is ref1
  Black Lives Matter (BLM) ref1, ref2, ref3, ref4
  China ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9
  climate change hoax ref1
  contradictory rules ref1
  Covid-19 ref1, ref2, ref3
  fascism ref1
  geographical origins ref1
  immigration ref1
  Internet ref1
  mainstream media ref1, ref2
  masks ref1, ref2
  monarchy ref1
  non-human dimension ref1

  - perception ref1
  - political parties ref1, ref2
  - pyramidal hierarchy ref1, ref2, ref3
  - reframing ref1
  - Sabbantian-Frankism ref1, ref2
  - science, manipulation of ref1
  - spider and the web ref1
  - transgender persons ref1
  - vaccines ref1
  - who controls the Cult ref1
  - Wokeness ref1, ref2, ref3, ref4

globalisation ref1, ref2

Gnostics ref1, ref2, ref3, ref4, ref5

Google ref1, ref2, ref3, ref4

government
  - behavioural scientists and psychologists, advice from ref1, ref2

definition ref1

Joint Biosecurity Centre (JBC) ref1

people, abusive relationship with ref1

Great Reset ref1, ref2, ref3, ref4, ref5, ref6
  - fascism ref1, ref2, ref3
  - financial system ref1
  - Human 2.0 ref1
  - water and food, control of ref1

green parties ref1

Griesz-Brisson, Margarite ref1

guaranteed income ref1, ref2, ref3

H

Hancock, Matt ref1, ref2, ref3, ref4, ref5

hand sanitisers ref1

heart ref1, ref2

hive-minds/groupthink ref1, ref2, ref3
holographs ref1, ref2, ref3, ref4
hospitals, empty ref1
human, meaning of ref1
Human 2.0 ref1
  addiction to technology ref1
  artificial intelligence (AI) ref1, ref2
  elimination of Human 1.0 ref1
  fertility, decline in ref1
  Great Reset ref1
  implantables ref1
  money ref1
  mRNA ref1
  nanotechnology ref1
  parents, replacement of ref1, ref2
  Smart Grid, connection to ref1, ref2
  synthetic biology ref1, ref2, ref3, ref4
  testosterone levels, decrease in ref1
  transgender = transhumanism ref1, ref2, ref3
  vaccines ref1, ref2, ref3, ref4
human sacrifice ref1, ref2, ref3
Hunger Games Society ref1, ref2, ref3, ref4, ref5, ref6, ref7
Huxley, Aldous ref1, ref2, ref3

I

identity politics ref1, ref2, ref3
Illuminati ref1, ref2
illusory physical reality ref1
immigration ref1, ref2, ref3, ref4
Imperial College ref1, ref2, ref3, ref4, ref5, ref6
implantables ref1, ref2

incomes, destruction of ref1, ref2
Infinite Awareness ref1, ref2, ref3, ref4
Internet ref1, ref2 *see also* social media
  artificial intelligence (AI) ref1
  independent journalism, lack of ref1
  Internet of Bodies (IoB) ref1
Internet of Everything (IoE) ref1, ref2
Internet of Things (IoT) ref1, ref2
lockdowns ref1
Psyop (psychological operation), Covid as a ref1
  trolls ref1
intersectionality ref1
inversion
  Archons ref1, ref2, ref3
  climate change hoax ref1
  energy ref1
  Judaism ref1, ref2, ref3
  symbolism ref1
  Wetiko factor ref1
  Wokeness ref1, ref2, ref3
Islam
  Archons ref1
  crypto-Jews ref1
  Islamic State ref1, ref2
  Jinn and Djinn ref1, ref2, ref3
  Ottoman Empire ref1
  Wahhabism ref1
isolation *see* **social distancing** *and* **isolation**
Israel
  China ref1
  Cyber Intelligence Unit Beersheba complex ref1
  expansion of illegal settlements ref1

- formation ref1
- Global Cult ref1
- Judaism ref1, ref2, ref3, ref4, ref5
- medical experiments, consent for ref1
- Mossad ref1, ref2, ref3, ref4
- Palestine-Israel conflict ref1, ref2, ref3
- parents, replacement of ref1
- Sabbatians ref1, ref2, ref3, ref4, ref5
- September 11, 2001, terrorist attacks on United States ref1
- Silicon Valley ref1
- Smart Grid ref1, ref2
- United States ref1, ref2
- vaccines ref1
- Wetiko factor ref1

Italy
- fear ref1, ref2, ref3
- Lombardy ref1, ref2, ref3
- vaccines ref1

J

Johns Hopkins University ref1, ref2, ref3, ref4, ref5, ref6, ref7

Johnson, Boris ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8

Joint Biosecurity Centre (JBC) ref1

Judaism
- anti-Semitism ref1, ref2, ref3
- Archons ref1, ref2
- crypto-Jews ref1
- inversion ref1, ref2, ref3
- Israel ref1, ref2, ref3, ref4, ref5
- Labour Party ref1
- Nazi Germany ref1, ref2, ref3, ref4
- Sabbatians ref1, ref2, ref3, ref4, ref5

Silicon Valley ref1
Torah ref1
United States ref1, ref2
Zionists ref1, ref2, ref3

K

Kaufman, Andrew ref1, ref2, ref3, ref4
knowledge ref1, ref2, ref3, ref4, ref5, ref6
Koch's postulates ref1
Kurzweil, Ray ref1, ref2, ref3, ref4, ref5, ref6, ref7
Kushner, Jared **ref1, ref2**

L

Labour Party ref1, ref2
Lanka, Stefan ref1, ref2
Lateral Flow Device (LFD) ref1
Levy, Paul ref1, ref2, ref3
Life Program ref1
lockdowns ref1, ref2, ref3
amplification tampering ref1
Archons ref1
Behavioural Insights Team ref1
Black Lives Matter (BLM) ref1
care homes, deaths in ref1
children
abuse ref1, ref2
mental health ref1
China ref1, ref2
computer models ref1
consequences ref1, ref2
dependency ref1, ref2, ref3

  domestic abuse ref1
  fall in cases ref1
  fear ref1, ref2, ref3, ref4
  guaranteed income ref1
  Hunger Games Society ref1, ref2, ref3
  interaction, destroying ref1
  Internet ref1, ref2
  overdoses ref1
  perception ref1
police-military state ref1, ref2
protests ref1, ref2, ref3, ref4, ref5
psychopathic personality ref1, ref2, ref3
reporting/snitching, encouragement of ref1, ref2
  testing ref1
  vaccines ref1
  Wetiko factor ref1
  WHO ref1

love ref1, ref2, ref3

Lucifer ref1, ref2, ref3

M

Madej, Carrie ref1, ref2

Magufuli, John ref1, ref2

mainstream media ref1
  BBC ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
  censorship ref1, ref2
  China ref1
  climate change hoax ref1
  fear ref1, ref2
  Global Cult ref1, ref2
  independent journalism, lack of ref1
  Ofcom ref1, ref2, ref3

  - perception ref1, ref2
  - Psyop (psychological operation), Covid as a ref1
  - Sabbatians ref1, ref2
  - social disapproval ref1
  - social distancing and isolation ref1
  - United States ref1, ref2
  - vaccines ref1, ref2, ref3, ref4, ref5

Mao Zedong ref1, ref2, ref3

Marx and Marxism ref1, ref2, ref3, ref4, ref5, ref6

masculinity ref1

masks/face coverings ref1, ref2, ref3
  - censorship ref1
  - children ref1, ref2, ref3, ref4, ref5
  - China, made in ref1
  - dehumanisation ref1, ref2, ref3
  - fear ref1, ref2
  - flu ref1
  - health professionals ref1, ref2, ref3, ref4
  - isolation ref1
  - laughter ref1

mass non-cooperation ref1

microplastics, risk of ref1

mind control ref1

multiple masks ref1

oxygen deficiency ref1, ref2, ref3

police ref1, ref2, ref3, ref4, ref5

pollution, as cause of plastic ref1

Psyop (psychological operation), Covid as a ref1

reframing ref1, ref2

risk assessments, lack of ref1, ref2

self-respect ref1

surgeons ref1

United States ref1
vaccines ref1, ref2, ref3, ref4, ref5
Wetiko factor ref1
'worms' ref1
The Matrix movies ref1, ref2, ref3
measles ref1, ref2
media see mainstream media
Medicines and Healthcare products Regulatory Agency (MHRA) ref1, ref2, ref3, ref4
Mesopotamia ref1
messaging ref1
military-police state ref1, ref2, ref3
mind control ref1, ref2, ref3, ref4, ref5, ref6 *see also* **MKUltra**
MKUltra ref1, ref2, ref3
monarchy **ref1**
money *see* **banking, finance and money**
Montagnier, Luc ref1, ref2, ref3
Mooney, Bel ref1
Morgellons disease ref1, ref2
mortality rate ref1
Mullis, Kary ref1, ref2, ref3
Musk, Elon ref1

N
Nag Hammadi texts ref1, ref2, ref3
nanotechnology ref1, ref2, ref3
narcissism ref1
Nazi Germany ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
near-death experiences ref1, ref2
Neocons ref1, ref2, ref3

Neuro-Linguistic Programming (NLP) and the Delphi technique **ref1**
NHS (National Health Service)
  amplification cycles ref1
  Common Purpose ref1, ref2
  mind control ref1
  NHS England ref1
  saving the NHS ref1, ref2
  vaccines ref1, ref2, ref3, ref4, ref5
  whistle-blowers ref1, ref2, ref3
No-Problem-Reaction-Solution ref1, ref2, ref3, ref4
non-human dimension of Global Cult ref1
nous ref1
numbers, reality as ref1
Nuremberg Codes ref1, ref2, ref3
Nuremberg-like tribunal, proposal for ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11, ref12

O

Obama, Barack ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
O'Brien, Cathy ref1, ref2, ref3, ref4
Ochel, Evita ref1
Ofcom ref1, ref2, ref3
old people ref1, ref2, ref3, ref4, ref5
Oneness ref1, ref2, ref3
Open Society Foundations (Soros) ref1, ref2, ref3
oxygen 406, 528–34

P

paedophilia ref1, ref2
Page, Larry ref1, ref2, ref3, ref4, ref5, ref6, ref7

Palestine-Israel conflict ref1, ref2, ref3
pandemic, definition of ref1
pandemic and health crisis scenarios/simulations ref1, ref2, ref3, ref4
paranormal ref1
PCR tests ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
Pearl Harbor attacks, prior knowledge of ref1
Pelosi, Nancy ref1, ref2, ref3
perception ref1, ref2, ref3, ref4
  climate change hoax ref1
  control ref1, ref2, ref3
  decoding ref1, ref2
  enslavement ref1
  externally-delivered perceptions ref1
  five senses ref1
  human labels ref1
  media ref1, ref2
  political parties ref1, ref2
  Psyop (psychological operation), Covid as a ref1
  sale of perception ref1
  self-identity ref1, ref2
  Wokeness ref1
Phantom Self ref1, ref2, ref3
pharmaceutical industry *see* **Big Pharma**
phthalates ref1
Plato's Allegory of the Cave ref1, ref2
pneuma ref1
police
  Black Lives Matter (BLM) ref1
  brutality ref1
  citizen's arrests ref1, ref2
  common law arrests ref1, ref2

  Common Purpose ref1
  defunding ref1
  lockdowns ref1, ref2
  masks ref1, ref2, ref3, ref4
  police-military state ref1, ref2, ref3
  psychopathic personality ref1, ref2, ref3, ref4
  reframing ref1
  United States ref1, ref2, ref3, ref4
  Wokeness ref1
polio ref1
political correctness ref1, ref2, ref3, ref4
political parties ref1, ref2, ref3, ref4
political puppets ref1
pollution ref1, ref2, ref3
post-mortems/autopsies ref1
Postage Stamp Consensus ref1, ref2
pre-emptive programming ref1
Problem-Reaction-Solution ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
Project for the New American Century ref1, ref2, ref3, ref4
psychopathic personality ref1
  Archons ref1
  heart energy ref1
  lockdowns ref1, ref2, ref3
  police ref1, ref2, ref3, ref4
  recruitment ref1, ref2
  vaccines ref1
  wealth ref1
  Wetiko ref1, ref2
Psyop (psychological operation), Covid as a ref1, ref2, ref3, ref4, ref5
Pushbackers ref1, ref2, ref3, ref4
pyramid structure ref1, ref2, ref3, ref4

Q

QAnon Psyop ref1, ref2, ref3

R

racism *see also* **Black Lives**
  Matter (BLM)
  anti-racism industry ref1
  class ref1
  critical race theory ref1
  culture ref1
  intersectionality ref1
  reverse racism ref1
white privilege ref1, ref2
white supremacy ref1, ref2, ref3, ref4, ref5
  Wokeness ref1, ref2, ref3
radiation ref1, ref2
randomness, illusion of ref1, ref2, ref3
reality ref1, ref2, ref3
reframing ref1, ref2
  change agents ref1, ref2
  children ref1
  climate change ref1
  Common Purpose leadership programme ref1, ref2
  contradictory rules ref1
  enforcers ref1
  masks ref1, ref2
  NLP and the Delphi technique ref1
  police ref1
  Wetiko factor ref1
  Wokeness ref1, ref2
religion *see also* particular religions
  alien invasions ref1

  Archons ref1, ref2
  consciousness ref1, ref2
  control, system of ref1, ref2, ref3
  criticism, prohibition on ref1
  five senses ref1
  good and evil, war between ref1
  hidden non-human forces ref1, ref2
  Sabbatians ref1
  save me syndrome ref1
  Wetiko ref1
  Wokeness ref1
repetition and mind control ref1, ref2, ref3
reporting/snitching, encouragement of ref1, ref2
Reptilians/Grey entities ref1
rewiring the mind ref1
Rivers, Thomas Milton ref1, ref2
Rockefeller family ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9
Rockefeller Foundation documents ref1, ref2, ref3, ref4
Roman Empire ref1
Rothschild family ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9
RT-PCR tests ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
Russia
  collusion inquiry in US ref1
Russian Revolution ref1, ref2
Sabbatians ref1

S

Sabbantian-Frankism ref1, ref2
  anti-Semitism ref1, ref2
  banking and finance ref1, ref2, ref3
  China ref1, ref2
  Israel ref1, ref2, ref3, ref4, ref5

Judaism ref1, ref2, ref3, ref4, ref5
Lucifer ref1
media ref1, ref2
Nazis ref1, ref2
QAnon ref1
Rothschilds ref1, ref2, ref3, ref4, ref5, ref6
Russia ref1
Saudi Arabia ref1
Silicon Valley ref1
Sumer ref1
United States ref1, ref2, ref3
Wetiko factor ref1
Wokeness ref1, ref2, ref3

SAGE (Scientific Advisory Group for Emergencies) ref1, ref2, ref3, ref4

SARS-1 ref1

SARs-CoV-2 ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8

Satan/Satanism ref1, ref2, ref3, ref4, ref5, ref6, ref7

satellites in low-orbit ref1

Saudi Arabia ref1

Save Me Syndrome ref1

scapegoating ref1

Schwab, Klaus ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11, ref12

science, manipulation of ref1

self-identity ref1, ref2, ref3, ref4

self-respect, attacks on ref1

September 11, 2001, terrorist attacks on United States ref1, ref2, ref3, ref4

77th Brigade of UK military ref1, ref2, ref3

Silicon Valley/tech giants ref1, ref2, ref3, ref4, ref5, ref6 *see also* **Facebook**

Israel ref1
Sabbatians ref1
technocracy ref1
Wetiko factor ref1
Wokeness ref1
simulation hypothesis ref1, ref2, ref3, ref4, ref5
Smart Grid ref1, ref2, ref3
artificial intelligence (AI) ref1
China ref1, ref2
control centres ref1
the Field ref1
Great Reset ref1
Human 2.0 ref1, ref2
Israel ref1, ref2
vaccines ref1
Wetiko factor ref1
social disapproval ref1
social distancing *and* **isolation** ref1, ref2, ref3
abusive relationships ref1, ref2
children ref1
flats and apartments ref1
heart issues ref1
hugs ref1
Internet ref1
masks ref1
media ref1
older people ref1, ref2
one-metre (three feet) rule ref1
rewiring the mind ref1
simulation, universe as a ref1
SPI-B ref1
substance abuse ref1

  suicide and self-harm ref1, ref2, ref3, ref4, ref5
  technology ref1
  torture, as ref1, ref2
  two-metre (six feet) rule ref1
  women ref1

social justice ref1, ref2, ref3, ref4

social media *see also* **Facebook bans on alternative views** ref1
  censorship ref1, ref2, ref3, ref4, ref5, ref6
  children ref1
  emotion ref1
  perception ref1
  private messages ref1
  Twitter ref1, ref2, ref3, ref4, ref5, ref6, ref7
  Wetiko factor ref1
  YouTube ref1, ref2, ref3, ref4, ref5

Soros, George ref1, ref2, ref3, ref4, ref5, ref6

Spain ref1

SPI-B (Scientific Pandemic Insights Group on Behaviours) ref1, ref2, ref3, ref4

spider and the web ref1, ref2, ref3, ref4

Starmer, Keir ref1

Statute Law ref1

Steiner, Rudolf ref1, ref2, ref3

Stockholm syndrome ref1

streptomycin ref1

suicide and self-harm ref1, ref2, ref3, ref4, ref5

Sumer ref1, ref2

Sunstein, Cass ref1, ref2, ref3

swine flu (H1N1) ref1, ref2, ref3

synchronicity ref1

synthetic biology ref1, ref2, ref3, ref4

synthetic meat ref1, ref2

T

technology *see also* **artificial intelligence (AI); Internet;**
  social media addiction ref1, ref2, ref3, ref4
  Archons ref1, ref2
  the cloud ref1, ref2, ref3, ref4, ref5, ref6, ref7
cyber-operations ref1
cyberwarfare ref1
  radiation ref1, ref2
  social distancing and isolation ref1
  technocracy ref1
Tedros Adhanom Ghebreyesus ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11, ref12, ref13
telepathy ref1
Tenpenny, Sherri ref1
Tesla, Nikola ref1
testosterone levels, decrease in ref1
testing for Covid-19 ref1, ref2
  anal swab tests ref1
  cancer ref1
  China ref1, ref2, ref3
  Corman-Drosten test ref1, ref2, ref3, ref4
  death certificates ref1, ref2
  fraudulent testing ref1
  genetic material, amplification of ref1
  Lateral Flow Device (LFD) ref1
  PCR tests ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
  vaccines ref1, ref2, ref3
Thunberg, Greta ref1, ref2, ref3
Totalitarian Tiptoe ref1, ref2, ref3, ref4
transgender persons
  activism ref1
  artificial wombs ref1

  censorship ref1
  child abuse ref1, ref2
  Human 2.0 ref1, ref2, ref3
  Wokeness ref1, ref2, ref3, ref4, ref5
  women, deletion of rights and status of ref1, ref2
  young persons ref1

travel restrictions ref1

Trudeau, Justin ref1, ref2, ref3

Trump, Donald ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11

Twitter ref1, ref2, ref3, ref4, ref5, ref6, ref7

U

UKColumn ref1, ref2

United Nations (UN) ref1, ref2, ref3, ref4, ref5 *see also* **Agenda 21/Agenda 2030 (UN)**

United States ref1, ref2
  American Revolution ref1
  borders ref1, ref2
  Capitol Hill riot ref1, ref2
  children ref1
  China ref1, ref2
  CIA ref1, ref2
  Daily Pass tracking system ref1
  demographics by immigration, changes in ref1
  Democrats ref1, ref2, ref3, ref4, ref5, ref6, ref7
  election fraud ref1
  far-right domestic terrorists, pushbackers as ref1
  Federal Reserve ref1
  flu/respiratory diseases statistics ref1
  Global Cult ref1, ref2
  hand sanitisers, FDA warnings on ref1

  - immigration, effects of illegal ref1
  - impeachment ref1
  - Israel ref1, ref2
  - Judaism ref1, ref2, ref3
  - lockdown ref1
  - masks ref1
  - mass media ref1, ref2
  - nursing homes ref1
  - Pentagon ref1, ref2, ref3, ref4
  - police ref1, ref2, ref3, ref4
  - pushbackers ref1
  - Republicans ref1, ref2
  - borders ref1, ref2

Democrats ref1, ref2, ref3, ref4, ref5

  - Russia, inquiry into collusion with ref1
  - Sabbatians ref1, ref2, ref3
  - September 11, 2001, terrorist attacks ref1, ref2, ref3, ref4
  - UFO sightings, release of information on ref1
  - vaccines ref1
  - white supremacy ref1, ref2, ref3, ref4
  - Woke Democrats ref1, ref2

V

vaccines ref1, ref2, ref3

  - adverse reactions ref1, ref2, ref3, ref4, ref5
  - Africa ref1
  - anaphylactic shock ref1, ref2, ref3, ref4
  - animals ref1, ref2
  - anti-vax movement ref1, ref2, ref3, ref4, ref5
  - AstraZeneca/Oxford ref1, ref2, ref3, ref4
  - autoimmune diseases, rise in ref1, ref2
  - Big Pharma ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8

bioweapon, as real ref1, ref2
black and ethnic minority communities ref1
blood clots ref1, ref2
Brain Computer Interface (BCI) ref1
care homes, deaths in ref1
censorship ref1, ref2, ref3
chief medical officers and scientific advisers, financial interests of ref1, ref2
children ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
China ref1, ref2
clinical trials ref1, ref2, ref3, ref4, ref5, ref6
compensation ref1
compulsory vaccinations ref1, ref2, ref3
computer programs ref1
consciousness ref1
cover-ups ref1
creation before Covid ref1
cytokine storm ref1
deaths and illnesses caused by vaccines ref1, ref2, ref3, ref4, ref5
definition ref1
developing countries ref1
digital tattoos ref1
DNA-manipulation ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
emergency approval ref1, ref2, ref3, ref4, ref5
female infertility ref1
funding ref1
genetic suicide ref1
Global Cult ref1
heart chakras ref1
hesitancy ref1
Human 2.0 ref1, ref2, ref3, ref4
immunity from prosecution ref1, ref2, ref3

  implantable technology ref1
  Israel ref1
  Johnson & Johnson ref1, ref2, ref3, ref4
  lockdowns ref1
  long-term effects ref1
  mainstream media ref1, ref2, ref3, ref4, ref5
  masks ref1, ref2, ref3, ref4, ref5
  Medicines and Healthcare products Regulatory Agency (MHRA) ref1, ref2
  messaging ref1
  Moderna ref1, ref2, ref3, ref4, ref5, ref6
  mRNA vaccines ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9
  nanotechnology ref1, ref2
  NHS ref1, ref2, ref3, ref4, ref5
  older people ref1, ref2
  operating system ref1
  passports ref1, ref2, ref3, ref4
  Pfizer/BioNTech ref1, ref2, ref3, ref4, ref5, ref6, ref7
polyethylene glycol ref1
pregnant women ref1
psychopathic personality ref1
races, targeting different ref1
  reverse transcription ref1
  Smart Grid ref1
  social distancing ref1
  social media ref1
  sterility ref1
  synthetic material, introduction of ref1
  tests ref1, ref2, ref3
  travel restrictions ref1
  variants ref1, ref2
  viruses, existence of ref1
  whistle-blowing ref1

  WHO ref1, ref2, ref3, ref4
  Wokeness ref1
  working, vaccine as ref1
  young people ref1

Vallance, Patrick ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9

variants ref1, ref2, ref3

vegans ref1

ventilators ref1, ref2

virology ref1, ref2

virtual reality ref1, ref2, ref3

viruses, existence of ref1

visual reality ref1, ref2

vitamin D ref1, ref2

von Braun, Wernher ref1, ref2

W

war-zone hospital myths ref1

waveforms ref1, ref2

wealth ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9 ref10, ref11

wet market conspiracy ref1

Wetiko factor ref1
  alcoholism and drug addiction ref1
  anti-human, why Global Cult is ref1
  Archons ref1, ref2, ref3, ref4
  artificial intelligence (AI) ref1
  Big Pharma ref1, ref2
  children ref1
  China ref1
  consciousness ref1, ref2
  education ref1
  Facebook ref1

  fear ref1, ref2
  frequency ref1, ref2
  Gates ref1, ref2
  Global Cult ref1, ref2
  heart ref1, ref2
  lockdowns ref1
  masks ref1
  Native American concept ref1
  psychopathic personality ref1, ref2
  reframing/retraining programmes ref1
  religion ref1
  Silicon Valley ref1
  Smart Grid ref1
  smartphone addiction ref1, ref2
  social media ref1
  war ref1, ref2
  WHO ref1
  Wokeness ref1, ref2, ref3
  Yaldabaoth ref1, ref2, ref3, ref4
whistle-blowing ref1, ref2, ref3, ref4, ref5, ref6, ref7
white privilege ref1, ref2
white supremacy ref1, ref2, ref3, ref4, ref5
Whitty, Christopher ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
‘who benefits’ ref1
Wi-Fi ref1, ref2, ref3, ref4
Wikipedia ref1, ref2
Wojcicki, Susan ref1, ref2, ref3, ref4, ref5, ref6, ref7
Wokeness
  Antifa ref1, ref2, ref3, ref4
  anti-Semitism ref1
  billionaire social justice warriors ref1, ref2, ref3

  - Capitol Hill riot ref1, ref2
  - censorship ref1
  - Christianity ref1
  - climate change hoax ref1, ref2
  - culture ref1
  - education, control of ref1
  - emotion ref1
  - facts ref1
  - fascism ref1, ref2, ref3
  - Global Cult ref1, ref2, ref3, ref4
  - group-think ref1
  - immigration ref1
  - indigenous people, solidarity with ref1
  - inversion ref1, ref2, ref3
  - left, hijacking the ref1, ref2
- Marxism ref1, ref2, ref3
- mind control ref1
- New Woke ref1
- Old Woke ref1
- Oneness ref1
- perceptual programming ref1
  - Phantom Self ref1
- police ref1
- defunding the ref1
- reframing ref1
- public institutions ref1
- Pushbackers ref1, ref2, ref3
- racism ref1, ref2, ref3
- reframing ref1, ref2
- religion, as ref1
- Sabbatians ref1, ref2, ref3
- Silicon Valley ref1
- social justice ref1, ref2, ref3, ref4

transgender ref1, ref2, ref3, ref4, ref5
United States ref1, ref2
  vaccines ref1
  Wetiko factor ref1, ref2, ref3
young people ref1, ref2, ref3
women, deletion of rights and status of ref1, ref2
World Economic Forum (WEF) ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9
World Health Organization (WHO) ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9
  AIDs/HIV ref1
  amplification cycles ref1
  Big Pharma ref1, ref2, ref3
  cooperation in health emergencies ref1
  creation ref1, ref2
  fatality rate ref1
  funding ref1, ref2, ref3
  Gates ref1
  Internet ref1
  lockdown ref1
  vaccines ref1, ref2, ref3, ref4
  Wetiko factor ref1
world number 1 (masses) ref1, ref2
world number 2 ref1
Wuhan ref1, ref2, ref3, ref4, ref5, ref6, ref7 ref8

Y

Yaldabaoth ref1, ref2, ref3, ref4, ref5, ref6
Yeadon, Michael ref1, ref2, ref3, ref4
young people *see also* **children** addiction to technology ref1
  Human 2.0 ref1
  vaccines ref1, ref2

  Wokeness ref1, ref2, ref3
YouTube ref1, ref2, ref3, ref4, ref5
  WHO 548

Z

Zaks, Tal ref1
Zionism ref1, ref2, ref3
Zuckerberg, Mark ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11, ref12
Zulus ref1

ICKONIC
THE ALTERNATIVE

Ickonic is something that has been a dream of mine for the last 5 years, growing up around alternative information I have always had a natural interest in what is going on in the World and what could I do to make it better.

Across the range of subjects and positions of influence occupied mainly by people who don't strive to make things better it's the Media that I have always found the most frustrating and fascinating. Mainly because if the Media did their Jobs properly then so much of the negative things happening in the World simply would not be able to happen, because they would be exposed within a heartbeat.

Free Press and the Opportunities that the internet could have given would mean that the Media are able to expose things like never before and hold people to account for their actions. As we all know there are 'Untouchables' that walk among us, people the Media simply won't touch, expose or investigate and that leads to the dark underworlds that infest the establishment the World over.

Well I say enough, it's time for something different, a different kind of Media, where no one is off limits from exposing and investigating. All we're interested in at Ickonic is the truth of what is really going on in the World on whichever subject we're covering.

We hope you enjoy what we have created and take something away from the platform, we aim to deliver information that's informative and most importantly self-empowering, you're not a little person, you're part of something much bigger than that and its time we as a collective race began to understand that and look to the future as ours to take.

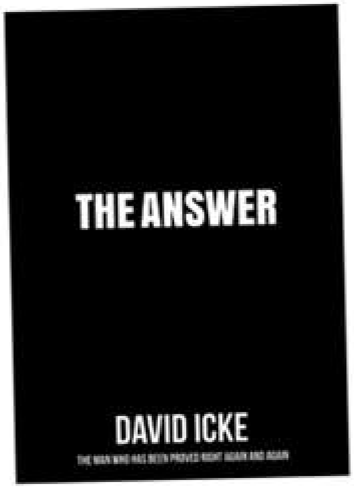
It's time...

Jaymie Icke - Founder Ickonic Alternative Media.

SIGN UP NOW AT ICKONIC.COM

DAVID ICKE

THE ANSWER



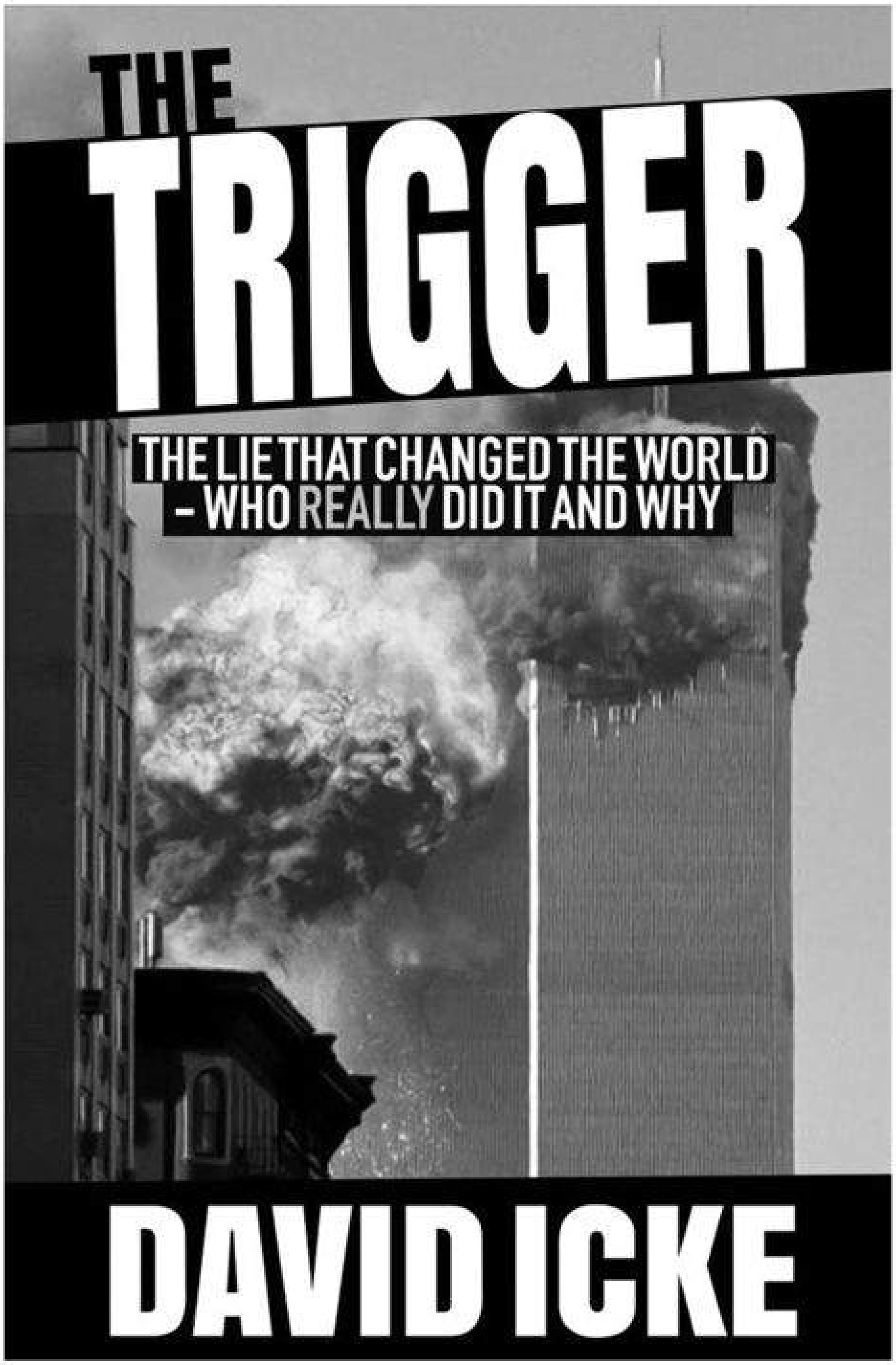
THE ANSWER

DAVID ICKE

We live in extraordinary times with billions bewildered and seeking answers for what is happening. David Icke, the man who has been proved right again and again, has spent 30 years uncovering the truth behind world affairs and in a stream of previous books he predicted current events.

The Answer will change your every perception of life and the world and set you free of the illusions that control human society. There is nothing more vital for our collective freedom than humanity becoming aware of what is in this book.

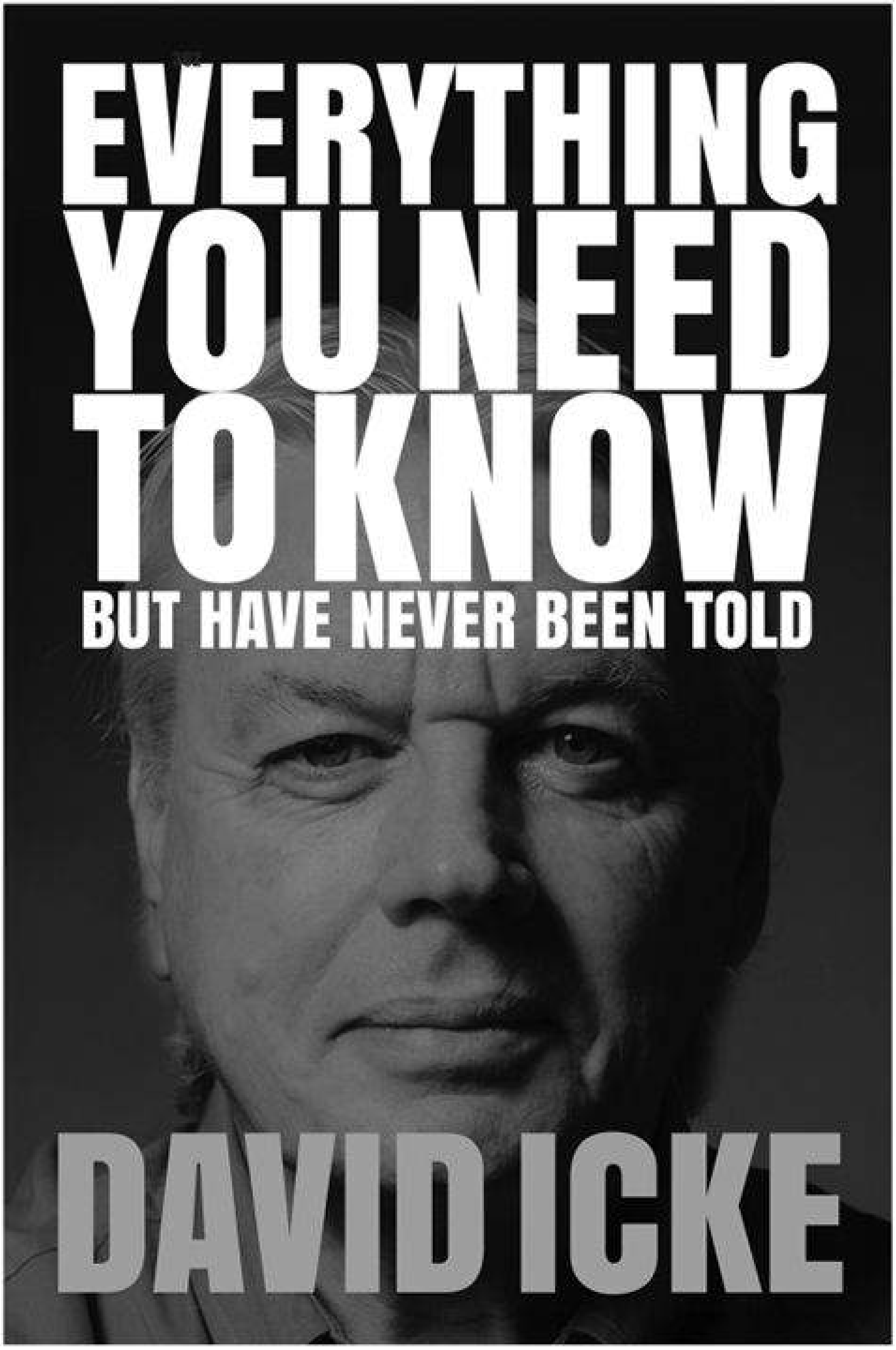
Available now at davidicke.com.



THE TRIGGER

THE LIE THAT CHANGED THE WORLD
– WHO REALLY DID IT AND WHY

DAVID ICKE



EVERYTHING YOU NEED TO KNOW

BUT HAVE NEVER BEEN TOLD

DAVID ICKE

DAVIDICKE.COM

DAVID ICKE STORE
LATEST NEWS ARTICLES
DAVID ICKE VIDEOS
WEEKLY DOT-CONNECTOR PODCASTS
LIVE EVENTS
WWW.DAVIDICKE.COM

THE LIFE STORY OF DAVID ICKE

RENEGADE

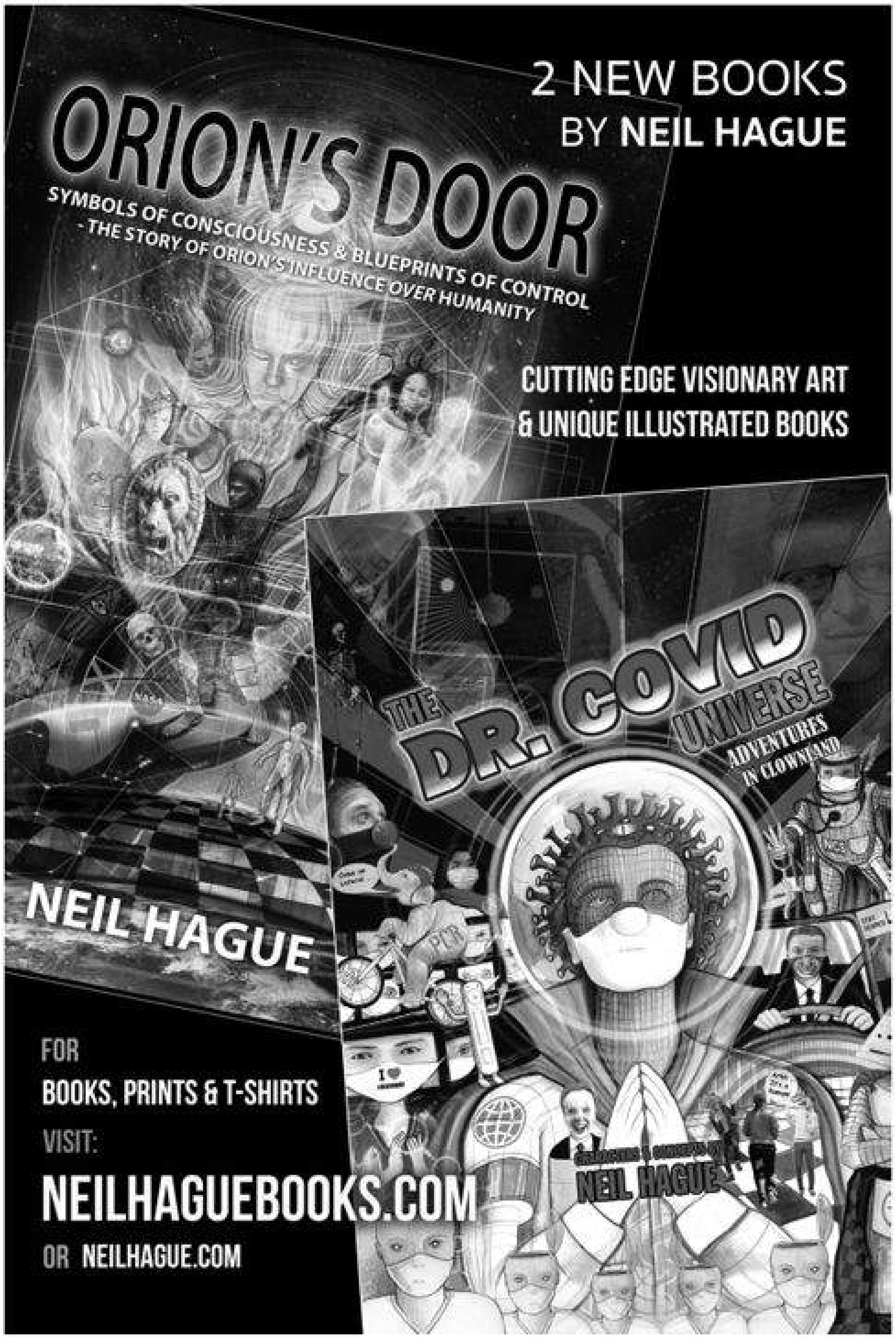
THE FEATURE LENGTH FILM

/ˈren·ɪˌɡeɪd/

noun

A person who behaves in a rebelliously unconventional manner.

AVAILABLE NOW AT DAVIDICKE.COM


ORION'S DOOR
SYMBOLS OF CONSCIOUSNESS & BLUEPRINTS OF CONTROL
- THE STORY OF ORION'S INFLUENCE OVER HUMANITY
NEIL HAGUE
THE DR. COVID UNIVERSE
ADVENTURES IN CLOWNLAND
NEIL HAGUE

2 NEW BOOKS
BY **NEIL HAGUE**

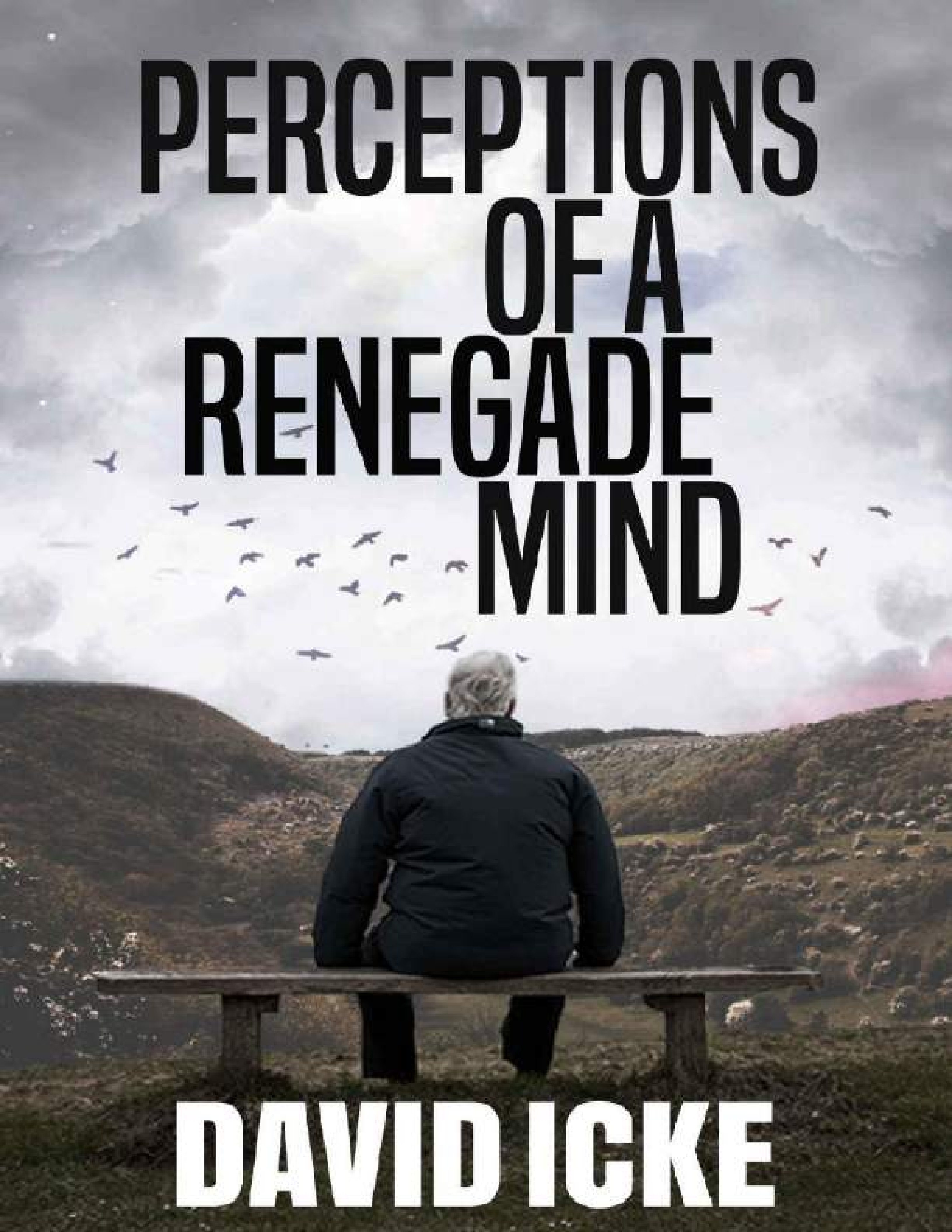
CUTTING EDGE VISIONARY ART
& UNIQUE ILLUSTRATED BOOKS

FOR
BOOKS, PRINTS & T-SHIRTS
VISIT:
NEILHAGUEBOOKS.COM
OR NEILHAGUE.COM

Before you go ...

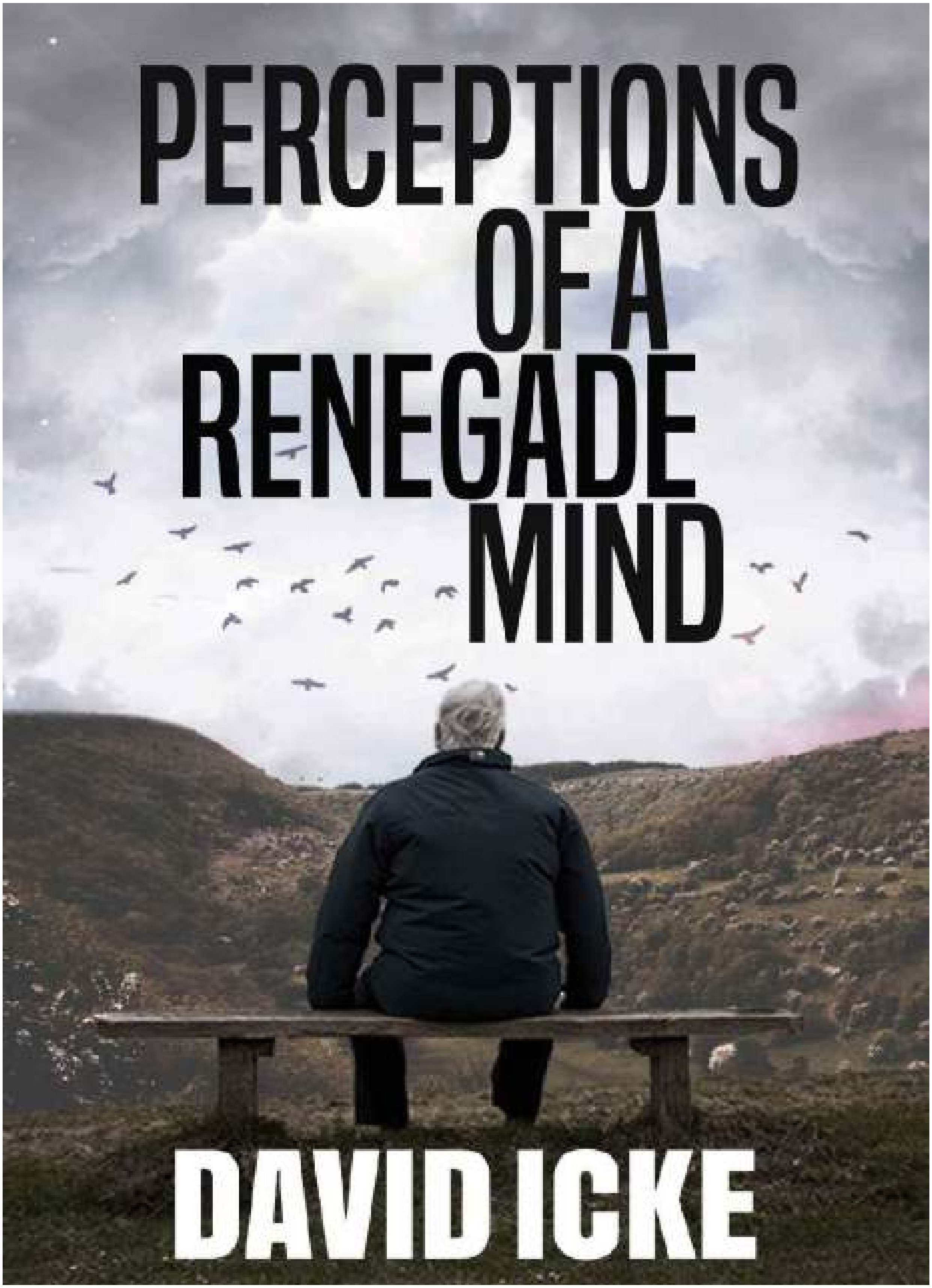
For more detail, background and evidence about the subjects in *Perceptions of a Renegade Mind* – and so much more – see my others books including *And The Truth Shall Set You Free; The Biggest Secret; Children of the Matrix; The David Icke Guide to the Global Conspiracy; Tales from the Time Loop; The Perception Deception; Remember Who You Are; Human Race Get Off Your Knees; Phantom Self; Everything You Need To Know But Have Never Been Told, The Trigger and The Answer*.

You can subscribe to the fantastic new Ickonic media platform where there are many hundreds of hours of cutting-edge information in videos, documentaries and series across a whole range of subjects which are added to every week. This includes my 90 minute breakdown of the week's news every Friday to explain *why* events are happening and to what end.



PERCEPTIONS OF A RENEGADE MIND

DAVID ICKE



PERCEPTIONS OF A RENEGADE MIND

DAVID ICKE

PERCEPTIONS
OF A
RENEGADE
MIND

ickonic
publishing

First published in July 2021.

ickonic
publishing

New Enterprise House
St Helens Street
Derby
DE1 3GY
UK

email: gareth.icke@davidicke.com

Copyright © 2021 David Icke

No part of this book may be reproduced in any form without permission from the Publisher, except for the quotation of brief passages in criticism

Cover Design: Gareth Icke
Book Design: Neil Hague

British Library Cataloguing-in Publication Data
A catalogue record for this book is available from the British Library

eISBN 978-18384153-1-0

PERCEPTIONS
OF A
RENEGADE
MIND

DAVID ICKE

Dedication:

To *Freeeeeedom!*

ICKONIC
THE ALTERNATIVE

NEW. DIFFERENT. REVOLUTIONARY

HUNDREDS OF CUTTING EDGE DOCUMENTARIES,
FEATURE FILMS, SERIES & PODCASTS.

SIGN UP NOW AT ICKONIC.COM

THE LIFE STORY OF DAVID ICKE

RENEGADE

THE FEATURE LENGTH FILM

AVAILABLE NOW AT DAVIDICKE.COM

Renegade:
Adjective
‘Having rejected tradition: Unconventional.’
Merriam-Webster Dictionary

Acquiescence to tyranny is the death of the spirit

You may be 38 years old, as I happen to be. And one day, some great opportunity stands before you and calls you to stand up for some great principle, some great issue, some great cause. And you refuse to do it because you are afraid … You refuse to do it because you want to live longer … You're afraid that you will lose your job, or you are afraid that you will be criticised or that you will lose your popularity, or you're afraid that somebody will stab you, or shoot at you or bomb your house; so you refuse to take the stand.

Well, you may go on and live until you are 90, but you're just as dead at 38 as you would be at 90. And the cessation of breathing in your life is but the belated announcement of an earlier death of the spirit.

Martin Luther King

How the few control the many and always have – the many do whatever they're told

'Forward, the Light Brigade!'
Was there a man dismayed?
Not though the soldier knew
Someone had blundered.
Theirs not to make reply,
Theirs not to reason why,
Theirs but to do and die.
Into the valley of Death
Rode the six hundred.

Cannon to right of them,
Cannon to left of them,
Cannon in front of them
Volleyed and thundered;
Stormed at with shot and shell,
Boldly they rode and well,
Into the jaws of Death,
Into the mouth of hell
Rode the six hundred

Alfred Lord Tennyson (1809-1892)

The mist is lifting slowly
I can see the way ahead
And I've left behind the empty streets
That once inspired my life
And the strength of the emotion
Is like thunder in the air
'Cos the promise that we made each other
Haunts me to the end

The secret of your beauty
And the mystery of your soul
I've been searching for in everyone I meet
And the times I've been mistaken
It's impossible to say
And the grass is growing
Underneath our feet

The words that I remember
From my childhood still are true
That there's none so blind
As those who will not see
And to those who lack the courage
And say it's dangerous to try
Well they just don't know
That love eternal will not be denied

I know you're out there somewhere
Somewhere, somewhere
I know you're out there somewhere

Somewhere you can hear my voice
I know I'll find you somehow
Somehow, somehow
I know I'll find you somehow
And somehow I'll return again to you

The Moody Blues

Are you a gutless wonder – or a Renegade Mind?

Monuments put from pen to paper,
Turns me into a gutless wonder,
And if you tolerate this,
Then your children will be next.
Gravity keeps my head down,
Or is it maybe shame ...

Manic Street Preachers

Rise like lions after slumber
In unvanquishable number.
Shake your chains to earth like dew
Which in sleep have fallen on you.
Ye are many – they are few.

Percy Shelley

Contents

CHAPTER 1 'I'm thinking' – Oh, but *are* you?

CHAPTER 2 Renegade perception

CHAPTER 3 The Pushbacker sting

CHAPTER 4 'Covid': The calculated catastrophe

CHAPTER 5 There *is no* 'virus'

CHAPTER 6 Sequence of deceit

CHAPTER 7 War on your mind

CHAPTER 8 'Reframing' insanity

CHAPTER 9 We must have it? So what is it?

CHAPTER 10 Human 2.0

CHAPTER 11 Who controls the Cult?

CHAPTER 12 Escaping Wetiko

POSTSCRIPT

APPENDIX Cowan-Kaufman-Morell Statement on Virus Isolation

BIBLIOGRAPHY

INDEX

CHAPTER ONE

I'm thinking' – Oh, but *are* you?

Think for yourself and let others enjoy the privilege of doing so too
Voltaire

French-born philosopher, mathematician and scientist René Descartes became famous for his statement in Latin in the 17th century which translates into English as: 'I think, therefore I am.'

On the face of it that is true. Thought reflects perception and perception leads to both behaviour and self-identity. In that sense 'we' are what we think. But who or what is doing the thinking and is thinking the only route to perception? Clearly, as we shall see, 'we' are not always the source of 'our' perception, indeed with regard to humanity as a whole this is rarely the case; and thinking is far from the only means of perception. Thought is the village idiot compared with other expressions of consciousness that we all have the potential to access and tap into. This has to be true when we *are* those other expressions of consciousness which are infinite in nature. We have forgotten this, or, more to the point, been manipulated to forget.

These are not just the esoteric musings of the navel. The whole foundation of human control and oppression is control of perception. Once perception is hijacked then so is behaviour which is dictated by perception. Collective perception becomes collective behaviour and collective behaviour is what we call human society. Perception is all and those behind human control know that which is why perception is the target 24/7 of the psychopathic manipulators that I call the Global Cult. They know that if they dictate perception they will dictate behaviour and collectively dictate the nature of human society. They are further aware that perception is formed from information received and if they control the circulation of information they will to a vast extent direct human behaviour. Censorship of information and opinion has become globally Nazi-like in recent years and never more blatantly than since the illusory 'virus pandemic' was triggered out of China in 2019 and across the world in 2020. Why have billions submitted to house arrest and accepted fascistic societies in a way they would have never believed possible? Those controlling the information spewing from government, mainstream media and Silicon Valley (all controlled by the same Global Cult networks) told them they were in danger from a 'deadly virus' and only by submitting to house arrest and conceding their most basic of freedoms could they and their families be protected. This monumental and provable lie became the *perception* of the billions and therefore the *behaviour* of the billions. In those few words you have the whole structure and modus operandi of human control. Fear is a perception – **F**alse **E**motion **A**ppearing **R**eal – and fear is the currency of control. In short … get them by the balls (or give them the impression that you have) and their hearts and minds will follow. Nothing grips the dangly bits and freezes the rear-end more comprehensively than fear.

World number 1

There are two 'worlds' in what appears to be one 'world' and the prime difference between them is knowledge. First we have the mass of human society in which the population is maintained in coldly-calculated ignorance through control of information and the 'education' (indoctrination) system. That's all you really need to control to enslave billions in a perceptual delusion in which what are perceived to be *their* thoughts and opinions are ever-repeated mantras that the system has been downloading all their lives through 'education', media, science, medicine, politics and academia in which the personnel and advocates are themselves overwhelmingly the perceptual products of the same repetition. Teachers and academics in general are processed by the same programming machine as everyone else, but unlike the great majority they never leave the ‘education’ program. It gripped them as students and continues to grip them as programmers of subsequent generations of students. The programmed become the programmers – the programmed programmers. The same can largely be said for scientists, doctors and politicians and not least because as the American writer Upton Sinclair said: ‘It is difficult to get a man to understand something when his salary depends upon his not understanding it.’ If your career and income depend on thinking the way the system demands then you will – bar a few free-minded exceptions – concede your mind to the Perceptual Mainframe that I call the Postage Stamp Consensus. This is a tiny band of perceived knowledge and possibility ‘taught’ (downloaded) in the schools and universities, pounded out by the mainstream media and on which all government policy is founded. Try thinking, and especially speaking and acting, outside of the ‘box’ of consensus and see what that does for your career in the Mainstream Everything which bullies, harasses, intimidates and ridicules the population into compliance. Here we have the simple structure which enslaves most of humanity in a perceptual prison cell for an entire lifetime and I’ll go deeper into this process shortly. Most of what humanity is taught as fact is nothing more than programmed belief. American science fiction author Frank Herbert was right when he said: ‘Belief can be manipulated. Only knowledge is dangerous.’ In the ‘Covid’ age belief is promoted and knowledge is censored. It was always so, but never to the extreme of today.

World number 2

A ‘number 2’ is slang for ‘doing a poo’ and how appropriate that is when this other ‘world’ is doing just that on humanity every minute of every day. World number 2 is a global network of secret societies and semi-secret groups dictating the direction of society via governments, corporations and authorities of every kind. I have spent more than 30 years uncovering and exposing this network that I call the Global Cult and knowing its agenda is what has made my books so accurate in predicting current and past events. Secret societies are secret for a reason. They want to keep their hoarded knowledge to themselves and their chosen initiates and to hide it from the population which they seek through ignorance to control and subdue. The whole foundation of the division between World 1 and World 2 is *knowledge*. What number 1 knows number 2 must not. Knowledge they have worked so hard to keep secret includes (a) the agenda to enslave humanity in a centrally-controlled global dictatorship, and (b) the nature of reality and life itself. The latter (b) must be suppressed to allow the former (a) to prevail as I shall be explaining. The way the Cult manipulates and interacts with the population can be likened to a spider's web. The 'spider' sits at the centre in the shadows and imposes its will through the web with each strand represented in World number 2 by a secret society, satanic or semi-secret group, and in World number 1 – the world of the seen – by governments, agencies of government, law enforcement, corporations, the banking system, media conglomerates and Silicon Valley (Fig 1 overleaf). The spider and the web connect and coordinate all these organisations to pursue the same global outcome while the population sees them as individual entities working randomly and independently. At the level of the web governments *are* the banking system *are* the corporations *are* the media *are* Silicon Valley *are* the World Health Organization working from their inner cores as one unit. Apparently unconnected countries, corporations, institutions, organisations and people are on the *same team* pursuing the same global outcome. Strands in the web immediately around the spider are the most secretive and exclusive secret societies and their membership is emphatically restricted to the Cult inner-circle emerging through the generations from particular bloodlines for reasons I will come to. At the core of the core you would get them in a single room. That's how many people are dictating the direction of human society and its transformation through the ‘Covid’ hoax and other means. As the web expands out from the spider we meet the secret societies that many people will be aware of – the Freemasons, Knights Templar, Knights of Malta, Opus Dei, the inner sanctum of the Jesuit Order, and such like. Note how many are connected to the Church of Rome and there is a reason for that. The Roman Church was established as a revamp, a rebranding, of the relocated ‘Church’ of Babylon and the Cult imposing global tyranny today can be tracked back to Babylon and Sumer in what is now Iraq.

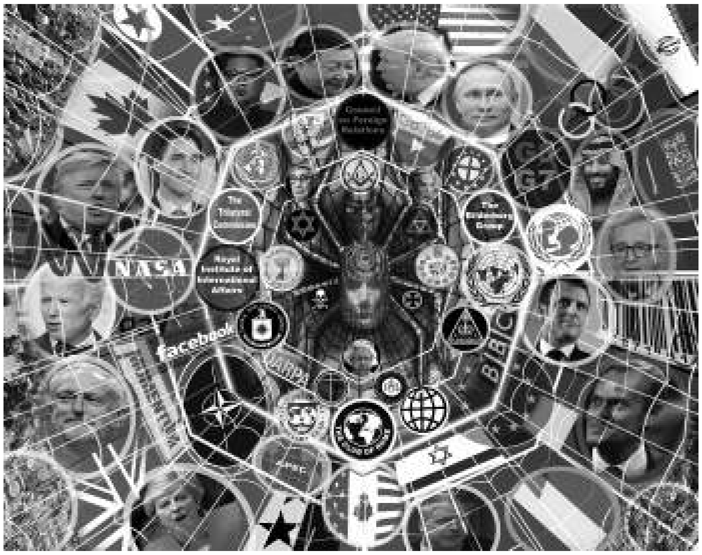

Council on Foreign Relations
The Trilateral Commission
The Bilderberg Group
NASA
Royal Institute of International Affairs
facebook
BBC

Figure 1: The global web through which the few control the many. (Image Neil Hague.)

Inner levels of the web operate in the unseen away from the public eye and then we have what I call the cusp organisations located at the point where the hidden meets the seen. They include a series of satellite organisations answering to a secret society founded in London in the late 19th century called the Round Table and among them are the Royal Institute of International Affairs (UK, founded in 1920); Council on Foreign Relations (US, 1921); Bilderberg Group (worldwide, 1954); Trilateral Commission (US/worldwide, 1972); and the Club of Rome (worldwide, 1968) which was created to exploit environmental concerns to justify the centralisation of global power to ‘save the planet’. The Club of Rome instigated with others the human-caused climate change hoax which has led to all the ‘green new deals' demanding that very centralisation of control. Cusp organisations, which include endless 'think tanks' all over the world, are designed to coordinate a single global policy between political and business leaders, intelligence personnel, media organisations and anyone who can influence the direction of policy in their own sphere of operation. Major players and regular attenders will know what is happening – or some of it – while others come and go and are kept overwhelmingly in the dark about the big picture. I refer to these cusp groupings as semi-secret in that they can be publicly identified, but what goes on at the inner-core is kept very much 'in house' even from most of their members and participants through a fiercely-imposed system of compartmentalisation. Only let them know what they need to know to serve your interests and no more. The structure of secret societies serves as a perfect example of this principle. Most Freemasons never get higher than the bottom three levels of 'degree' (degree of knowledge) when there are 33 official degrees of the Scottish Rite. Initiates only qualify for the next higher 'compartment' or degree if those at that level choose to allow them. Knowledge can be carefully assigned only to those considered 'safe'. I went to my local Freemason's lodge a few years ago when they were having an 'open day' to show how cuddly they were and when I chatted to some of them I was astonished at how little the rank and file knew even about the most ubiquitous symbols they use. The mushroom technique – keep them in the dark and feed them bullshit – applies to most people in the web as well as the population as a whole. Sub-divisions of the web mirror in theme and structure transnational corporations which have a headquarters somewhere in the world dictating to all their subsidiaries in different countries. Subsidiaries operate in their methodology and branding to the same centrally-dictated plan and policy in pursuit of particular ends. The Cult web functions in the same way. Each country has its own web as a subsidiary of the global one. They consist of networks of secret societies, semi-secret groups and bloodline families and their job is to impose the will of the spider and the global web in their particular country. Subsidiary networks control and manipulate the national political system, finance, corporations, media, medicine, etc. to ensure that they follow the globally-dictated Cult agenda. These networks were the means through which the 'Covid' hoax could be played out with almost every country responding in the same way.

The 'Yessir' pyramid

Compartmentalisation is the key to understanding how a tiny few can dictate the lives of billions when combined with a top-down sequence of imposition and acquiescence. The inner core of the Cult sits at the peak of the pyramidal hierarchy of human society (Fig 2 overleaf). It imposes its will – its agenda for the world – on the level immediately below which acquiesces to that imposition. This level then imposes the Cult will on the level below them which acquiesces and imposes on the next level. Very quickly we meet levels in the hierarchy that have no idea there even is a Cult, but the sequence of imposition and acquiescence continues down the pyramid in just the same way. 'I don't know why we are doing this but the order came from "on-high" and so we better just do it.' Alfred Lord Tennyson said of the cannon fodder levels in his poem *The Charge of the Light Brigade*: 'Theirs not to reason why; theirs but to do and die.' The next line says that 'into the valley of death rode the six hundred' and they died because they obeyed without question what their perceived 'superiors' told them to do. In the same way the population capitulated to 'Covid'. The whole hierarchical pyramid functions like this to allow the very few to direct the enormous many. Eventually imposition-acquiescence-imposition-acquiescence comes down to the mass of the population at the foot of the pyramid. If they acquiesce to those levels of the hierarchy imposing on them (governments/law enforcement/doctors/media) a circuit is completed between the population and the handful of super-psychopaths in the Cult inner core at the top of the pyramid. Without a circuit-breaking refusal to obey, the sequence of imposition and acquiescence allows a staggeringly few people to impose their will upon the entirety of humankind. We are looking at the very sequence that has subjugated billions since the start of 2020. Our freedom has not been taken from us. Humanity has given it away. Fascists do not impose fascism because there are not enough of them. Fascism is imposed by the population acquiescing to fascism. Put another way allowing their perceptions to be programmed to the extent that leads to the population giving their freedom away by giving their perceptions – their mind – away. If this circuit is not broken by humanity ceasing to cooperate with their own enslavement then nothing can change. For that to happen people have to critically think and see through the lies and window dressing and then summon the backbone to act upon what they see. The Cult spends its days working to stop either happening and its methodology is systematic and highly detailed, but it can be overcome and that is what this book is all about.

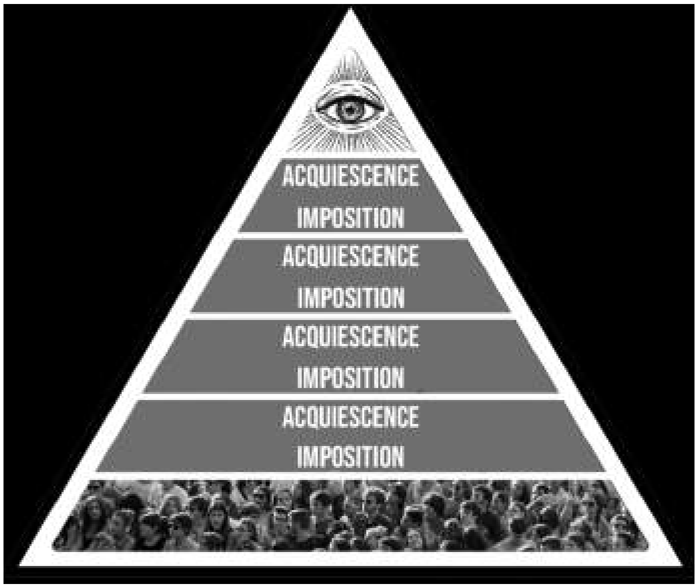


Figure 2: The simple sequence of imposition and compliance that allows a handful of people at the peak of the pyramid to dictate the lives of billions.

The Life Program

Okay, back to world number 1 or the world of the ‘masses’. Observe the process of what we call ‘life’ and it is a perceptual download from cradle to grave. The Cult has created a global structure in which perception can be programmed and the program continually topped-up with what appears to be constant confirmation that the program is indeed true reality. The important word here is ‘appears’.

This is the structure, the fly-trap, the Postage Stamp Consensus or Perceptual Mainframe, which represents that incredibly narrow band of perceived possibility delivered by the 'education' system, mainstream media, science and medicine. From the earliest age the download begins with parents who have themselves succumbed to the very programming their children are about to go through. Most parents don't do this out of malevolence and mostly it is quite the opposite. They do what they believe is best for their children and that is what the program has told them is best. Within three or four years comes the major transition from parental programming to full-blown state (Cult) programming in school, college and university where perceptually-programmed teachers and academics pass on their programming to the next generations. Teachers who resist are soon marginalised and their careers ended while children who resist are called a problem child for whom Ritalin may need to be prescribed. A few years after entering the 'world' children are under the control of authority figures representing the state telling them when they have to be there, when they can leave and when they can speak, eat, even go to the toilet. This is calculated preparation for a lifetime of obeying authority in all its forms. Reflex-action fear of authority is instilled by authority from the start. Children soon learn the carrot and stick consequences of obeying or defying authority which is underpinned daily for the rest of their life. Fortunately I daydreamed through this crap and never obeyed authority simply because it told me to. This approach to my alleged 'betters' continues to this day. There can be consequences of pursuing open-minded freedom in a world of closed-minded conformity. I spent a lot of time in school corridors after being ejected from the classroom for not taking some of it seriously and now I spend a lot of time being ejected from Facebook, YouTube and Twitter. But I can tell you that being true to yourself and not compromising your self-respect is far more exhilarating than bowing to authority for authority's sake. You don't have to be a sheep to the shepherd (authority) and the sheep dog (fear of not obeying authority).

The perceptual download continues throughout the formative years in school, college and university while script-reading 'teachers', 'academics' 'scientists', 'doctors' and 'journalists' insist that ongoing generations must be as programmed as they are. Accept the program or you will not pass your 'exams' which confirm your 'degree' of programming. It is tragic to think that many parents pressure their offspring to work hard at school to download the program and qualify for the next stage at college and university. The late, great, American comedian George Carlin said: 'Here's a bumper sticker I'd like to see: We are proud parents of a child who has resisted his teachers' attempts to break his spirit and bend him to the will of his corporate masters.' Well, the best of luck finding many of those, George. Then comes the moment to leave the formal programming years in academia and enter the 'adult' world of work. There you meet others in your chosen or prescribed arena who went through the same Postage Stamp Consensus program before you did. There is therefore overwhelming agreement between almost everyone on the basic foundations of Postage Stamp reality and the rejection, even contempt, of the few who have a mind of their own and are prepared to use it. This has two major effects. Firstly, the consensus confirms to the programmed that their download is really how things are. I mean, everyone knows that, right? Secondly, the arrogance and ignorance of Postage Stamp adherents ensure that anyone questioning the program will have unpleasant consequences for seeking their own truth and not picking their perceptions from the shelf marked: 'Things you must believe without question and if you don't you're a dangerous lunatic conspiracy theorist and a harebrained nutter'.

Every government, agency and corporation is founded on the same Postage Stamp prison cell and you can see why so many people believe the same thing while calling it their own 'opinion'. Fusion of governments and corporations in pursuit of the same agenda was the definition of fascism described by Italian dictator Benito Mussolini. The pressure to conform to perceptual norms downloaded for a lifetime is incessant and infiltrates society right down to family groups that become censors and condemners of their own 'black sheep' for not, ironically, being sheep. We have seen an explosion of that in the 'Covid' era. Cult-owned global media unleashes its propaganda all day every day in support of the Postage Stamp and targets with abuse and ridicule anyone in the public eye who won't bend their mind to the will of the tyranny. Any response to this is denied (certainly in my case). They don't want to give a platform to expose official lies. Cult-owned-and-created Internet giants like Facebook, Google, YouTube and Twitter delete you for having an unapproved opinion. Facebook boasts that its AI censors delete 97-percent of 'hate speech' before anyone even reports it. Much of that 'hate speech' will simply be an opinion that Facebook and its masters don't want people to see. Such perceptual oppression is widely known as fascism. Even Facebook executive Benny Thomas, a 'CEO Global Planning Lead', said in comments secretly recorded by investigative journalism operation Project Veritas that Facebook is 'too powerful' and should be broken up:

> I mean, no king in history has been the ruler of two billion people, but Mark Zuckerberg is … And he's 36. That's too much for a 36-year-old ... You should not have power over two billion people. I just think that's wrong.

Thomas said Facebook-owned platforms like Instagram, Oculus, and WhatsApp needed to be separate companies. 'It's too much power when they're all one together'. That's the way the Cult likes it, however. We have an executive of a Cult organisation in Benny Thomas that doesn't know there is a Cult such is the compartmentalisation. Thomas said that Facebook and Google 'are no longer companies, they're countries'. Actually they are more powerful than countries on the basis that if you control information you control perception and control human society.

I love my oppressor

Another expression of this psychological trickery is for those who realise they are being pressured into compliance to eventually convince themselves to believe the official narratives to protect their self-respect from accepting the truth that they have succumbed to meek and subservient compliance. Such people become some of the most vehement defenders of the system. You can see them everywhere screaming abuse at those who prefer to think for themselves and by doing so reminding the compliers of their own capitulation to conformity. ‘You are talking dangerous nonsense you Covidiot!!’ Are you trying to convince me or yourself? It is a potent form of Stockholm syndrome which is defined as: ‘A psychological condition that occurs when a victim of abuse identifies and attaches, or bonds, positively with their abuser.’ An example is hostages bonding and even ‘falling in love’ with their kidnappers. The syndrome has been observed in domestic violence, abused children, concentration camp inmates, prisoners of war and many and various Satanic cults. These are some traits of Stockholm syndrome listed at goodtherapy.org:

- Positive regard towards perpetrators of abuse or captor [see ‘Covid’].
- Failure to cooperate with police and other government authorities when it comes to holding perpetrators of abuse or kidnapping accountable [or in the case of ‘Covid’ cooperating with the police to enforce and defend their captors’ demands].
- Little or no effort to escape [see ‘Covid’].
- Belief in the goodness of the perpetrators or kidnappers [see ‘Covid’].
- Appeasement of captors. This is a manipulative strategy for maintaining one’s safety. As victims get rewarded – perhaps with less abuse or even with life itself – their appeasing behaviours are reinforced [see ‘Covid’].
- Learned helplessness. This can be akin to ‘if you can’t beat ‘em, join ‘em’. As the victims fail to escape the abuse or captivity, they may start giving up and soon realize it’s just easier for everyone if they acquiesce all their power to their captors [see ‘Covid’].

- Feelings of pity toward the abusers, believing they are actually victims themselves. Because of this, victims may go on a crusade or mission to 'save' [protect] their abuser [see the venom unleashed on those challenging the official 'Covid' narrative].
- Unwillingness to learn to detach from their perpetrators and heal. In essence, victims may tend to be less loyal to themselves than to their abuser [*definitely* see 'Covid'].

Ponder on those traits and compare them with the behaviour of great swathes of the global population who have defended governments and authorities which have spent every minute destroying their lives and livelihoods and those of their children and grandchildren since early 2020 with fascistic lockdowns, house arrest and employment deletion to 'protect' them from a 'deadly virus' that their abusers' perceptually created to bring about this very outcome. We are looking at mass Stockholm syndrome. All those that agree to concede their freedom will believe those perceptions are originating in their own independent 'mind' when in fact by conceding their reality to Stockholm syndrome they have by definition conceded any independence of mind. Listen to the 'opinions' of the acquiescing masses in this 'Covid' era and what gushes forth is the repetition of the official version of everything delivered unprocessed, unfiltered and unquestioned. The whole programming dynamic works this way. I must be free because I'm told that I am and so I think that I am.

You can see what I mean with the chapter theme of 'I'm thinking – Oh, but *are* you?' The great majority are not thinking, let alone for themselves. They are repeating what authority has told them to believe which allows them to be controlled. Weaving through this mentality is the fear that the 'conspiracy theorists' are right and this again explains the often hysterical abuse that ensues when you dare to contest the official narrative of anything. Denial is the mechanism of hiding from yourself what you don't want to be true. Telling people what they want to hear is easy, but it's an infinitely greater challenge to tell them what they would rather not be happening.

One is akin to pushing against an open door while the other is met with vehement resistance no matter what the scale of evidence. I don't want it to be true so I'll convince myself that it's not. Examples are everywhere from the denial that a partner is cheating despite all the signs to the reflex-action rejection of any idea that world events in which country after country act in exactly the same way are centrally coordinated. To accept the latter is to accept that a force of unspeakable evil is working to destroy your life and the lives of your children with nothing too horrific to achieve that end. Who the heck wants that to be true? But if we don't face reality the end is duly achieved and the consequences are far worse and ongoing than breaking through the walls of denial today with the courage to make a stand against tyranny.

Connect the dots – but how?

A crucial aspect of perceptual programming is to portray a world in which everything is random and almost nothing is connected to anything else. Randomness cannot be coordinated by its very nature and once you perceive events as random the idea they could be connected is waved away as the rantings of the tinfoil-hat brigade. You can't plan and coordinate random you idiot! No, you can't, but you can hide the coldly-calculated and long-planned behind the *illusion* of randomness. A foundation manifestation of the Renegade Mind is to scan reality for patterns that connect the apparently random and turn pixels and dots into pictures. This is the way I work and have done so for more than 30 years. You look for similarities in people, modus operandi and desired outcomes and slowly, then ever quicker, the picture forms. For instance: There would seem to be no connection between the 'Covid pandemic' hoax and the human-caused global-warming hoax and yet they are masks (appropriately) on the same face seeking the same outcome. Those pushing the global warming myth through the Club of Rome and other Cult agencies are driving the lies about 'Covid' – Bill Gates is an obvious one, but they are endless. Why would the same people be involved in both when they are clearly not connected? Oh, but they *are*. Common themes with personnel are matched by common goals. The 'solutions' to both 'problems' are centralisation of global power to impose the will of the few on the many to 'save' humanity from 'Covid' and save the planet from an 'existential threat' (we need 'zero Covid' and 'zero carbon emissions'). These, in turn, connect with the 'dot' of globalisation which was coined to describe the centralisation of global power in every area of life through incessant political and corporate expansion, trading blocks and superstates like the European Union. If you are the few and you want to control the many you have to centralise power and decision-making. The more you centralise power the more power the few at the centre will have over the many; and the more that power is centralised the more power those at the centre have to centralise even quicker. The momentum of centralisation gets faster and faster which is exactly the process we have witnessed. In this way the hoaxed 'pandemic' and the fakery of human-caused global warming serve the interests of globalisation and the seizure of global power in the hands of the Cult inner-circle which is behind 'Covid', 'climate change' *and* globalisation. At this point random 'dots' become a clear and obvious picture or pattern.

Klaus Schwab, the classic Bond villain who founded the Cult's Gates-funded World Economic Forum, published a book in 2020, *The Great Reset*, in which he used the 'problem' of 'Covid' to justify a total transformation of human society to 'save' humanity from 'climate change'. Schwab said: 'The pandemic represents a rare but narrow window of opportunity to reflect, reimagine, and reset our world.' What he didn't mention is that the Cult he serves is behind both hoaxes as I show in my book *The Answer*. He and the Cult don't have to reimagine the world. They know precisely what they want and that's why they destroyed human society with 'Covid' to 'build back better' in their grand design. Their job is not to imagine, but to get humanity to imagine and agree with their plans while believing it's all random. It must be pure coincidence that 'The Great Reset' has long been the Cult's code name for the global imposition of fascism and replaced previous code-names of the 'New World

Order' used by Cult frontmen like Father George Bush and the 'New Order of the Ages' which emerged from Freemasonry and much older secret societies. New Order of the Ages appears on the reverse of the Great Seal of the United States as 'Novus ordo seclorum' underneath the Cult symbol used since way back of the pyramid and all seeing-eye (Fig 3). The pyramid is the hierarchy of human control headed by the illuminated eye that symbolises the force behind the Cult which I will expose in later chapters. The term 'Annuit Coeptis' translates as 'He favours our undertaking'. We are told the 'He' is the Christian god, but 'He' is not as I will be explaining.

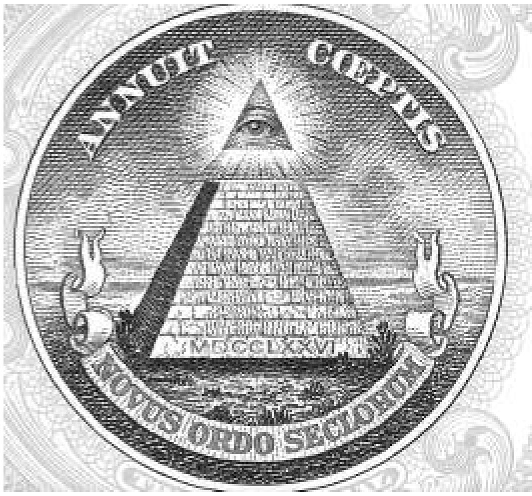


Figure 3: The all-seeing eye of the Cult 'god' on the Freemason-designed Great Seal of the United States and also on the dollar bill.

Having you on

Two major Cult techniques of perceptual manipulation that relate to all this are what I have called since the 1990s Problem-Reaction-Solution (PRS) and the Totalitarian Tiptoe (TT). They can be uncovered by the inquiring mind with a simple question: Who benefits? The answer usually identifies the perpetrators of a given action or happening through the concept of 'he who most benefits from a crime is the one most likely to have committed it'. The Latin 'Cue bono?' – Who benefits? – is widely attributed to the Roman orator and statesman Marcus Tullius Cicero. No wonder it goes back so far when the concept has been relevant to human behaviour since history was recorded. Problem-Reaction-Solution is the technique used to manipulate us every day by covertly creating a problem (or the illusion of one) and offering the solution to the problem (or the illusion of one). In the first phase you create the problem and blame someone or something else for why it has happened. This may relate to a financial collapse, terrorist attack, war, global warming or pandemic, anything in fact that will allow you to impose the 'solution' to change society in the way you desire at that time. The 'problem' doesn't have to be real. PRS is manipulation of perception and all you need is the population to believe the problem is real. Human-caused global warming and the 'Covid pandemic' only have to be *perceived* to be real for the population to accept the 'solutions' of authority. I refer to this technique as NO-Problem-Reaction-Solution. Billions did not meekly accept house arrest from early 2020 because there was a real deadly 'Covid pandemic' but because they perceived – believed – that to be the case. The antidote to Problem-Reaction-Solution is to ask who benefits from the proposed solution. Invariably it will be anyone who wants to justify more control through deletion of freedom and centralisation of power and decision-making.

The two world wars were Problem-Reaction-Solutions that transformed and realigned global society. Both were manipulated into being by the Cult as I have detailed in books since the mid-1990s. They dramatically centralised global power, especially World War Two, which led to the United Nations and other global bodies thanks to the overt and covert manipulations of the Rockefeller family and other Cult bloodlines like the Rothschilds. The UN is a stalking horse for full-blown world government that I will come to shortly. The land on which the UN building stands in New York was donated by the Rockefellers and the same Cult family was behind Big Pharma scalpel and drug 'medicine' and the creation of the World Health Organization as part of the UN. They have been stalwarts of the eugenics movement and funded Hitler's race-purity expert' Ernst Rudin. The human-caused global warming hoax has been orchestrated by the Club of Rome through the UN which is manufacturing both the ‘problem’ through its Intergovernmental Panel on Climate Change and imposing the ‘solution’ through its Agenda 21 and Agenda 2030 which demand the total centralisation of global power to ‘save the world’ from a climate hoax the United Nations is itself perpetrating. What a small world the Cult can be seen to be particularly among the inner circles. The bedfellow of Problem-Reaction-Solution is the Totalitarian Tiptoe which became the Totalitarian Sprint in 2020. The technique is fashioned to hide the carefully-coordinated behind the cover of apparently random events. You start the sequence at ‘A’ and you know you are heading for ‘Z’. You don’t want people to know that and each step on the journey is presented as a random happening while all the steps strung together lead in the same direction. The speed may have quickened dramatically in recent times, but you can still see the incremental approach of the Tiptoe in the case of ‘Covid’ as each new imposition takes us deeper into fascism. Tell people they have to do this or that to get back to ‘normal’, then this and this and this. With each new demand adding to the ones that went before the population’s freedom is deleted until it disappears. The spider wraps its web around the flies more comprehensively with each new diktat. I’ll highlight this in more detail when I get to the ‘Covid’ hoax and how it has been pulled off. Another prime example of the Totalitarian Tiptoe is how the Cult-created European Union went from a ‘free-trade zone’ to a centralised bureaucratic dictatorship through the Tiptoe of incremental centralisation of power until nations became mere administrative units for Cult-owned dark suits in Brussels.

The antidote to ignorance is knowledge which the Cult seeks vehemently to deny us, but despite the systematic censorship to that end the Renegade Mind can overcome this by vociferously seeking out the facts no matter the impediments put in the way. There is also a method of thinking and perceiving – *knowing* – that doesn’t even need names, dates, place-type facts to identify the patterns that reveal the story. I’ll get to that in the final chapter. All you need to know about the manipulation of human society and to what end is still out there – *at the time of writing* – in the form of books, videos and websites for those that really want to breach the walls of programmed perception. To access this knowledge requires the abandonment of the mainstream media as a source of information in the awareness that this is owned and controlled by the Cult and therefore promotes mass perceptions that suit the Cult. Mainstream media lies all day, every day. That is its function and very reason for being. Where it does tell the truth, here and there, is only because the truth and the Cult agenda very occasionally coincide. If you look for fact and insight to the BBC, CNN and virtually all the rest of them you are asking to be conned and perceptually programmed.

Know the outcome and you'll see the journey

Events seem random when you have no idea where the world is being taken. Once you do the random becomes the carefully planned. Know the outcome and you'll see the journey is a phrase I have been using for a long time to give context to daily happenings that appear unconnected. Does a problem, or illusion of a problem, trigger a proposed 'solution' that further drives society in the direction of the outcome? Invariably the answer will be yes and the random – *abracadabra* – becomes the clearly coordinated. So what is this outcome that unlocks the door to a massively expanded understanding of daily events? I will summarise its major aspects – the fine detail is in my other books – and those new to this information will see that the world they thought they were living in is a very different place. The foundation of the Cult agenda is the incessant centralisation of power and all such centralisation is ultimately in pursuit of Cult control on a global level. I have described for a long time the planned world structure of top-down dictatorship as the Hunger Games Society. The term obviously comes from the movie series which portrayed a world in which a few living in military-protected hi-tech luxury were the overlords of a population condemned to abject poverty in isolated 'sectors' that were not allowed to interact. 'Covid' lockdowns and travel bans anyone? The 'Hunger Games' pyramid of structural control has the inner circle of the Cult at the top with pretty much the entire population at the bottom under their control through dependency for survival on the Cult. The whole structure is planned to be protected and enforced by a military-police state (Fig 4).

Here you have the reason for the global lockdowns of the fake pandemic to coldly destroy independent incomes and livelihoods and make everyone dependent on the 'state' (the Cult that controls the 'states'). I have warned in my books for many years about the plan to introduce a 'guaranteed income' – a barely survivable pittance – designed to impose dependency when employment was destroyed by AI technology and now even more comprehensively at great speed by the 'Covid' scam. Once the pandemic was played and lockdown consequences began to delete independent income the authorities began to talk right on cue about the need for a guaranteed income and a 'Great Reset'. Guaranteed income will be presented as benevolent governments seeking to help a desperate people – desperate as a direct result of actions of the same governments. The truth is that such payments are a trap. You will only get them if you do exactly what the authorities demand including mass vaccination (genetic manipulation). We have seen this theme already in Australia where those dependent on government benefits have them reduced if parents don't agree to have their children vaccinated according to an insane health-destroying government-dictated schedule. Calculated economic collapse applies to governments as well as people. The Cult wants rid of countries through the creation of a world state with countries broken up into regions ruled by a world government and super states like the European Union. Countries must be bankrupted, too, to this end and it's being achieved by the trillions in 'rescue packages' and furlough payments, trillions in lost taxation, and money-no-object spending on 'Covid' including constant all-medium advertising (programming) which has made the media dependent on government for much of its income. The day of reckoning is coming – as planned – for government spending and given that it has been made possible by printing money and not by production/taxation there is inflation on the way that has the potential to wipe out monetary value. In that case there will be no need for the Cult to steal your money. It just won't be worth anything (see the German Weimar Republic before the Nazis took over). Many have been okay with lockdowns while getting a percentage of their income from so-called furlough payments without having to work. Those payments are dependent, however, on people having at least a theoretical job with a business considered non-essential and ordered to close. As these business go under because they are closed by lockdown after lockdown the furlough stops and it will for everyone eventually. Then what? The 'then what?' is precisely the idea.

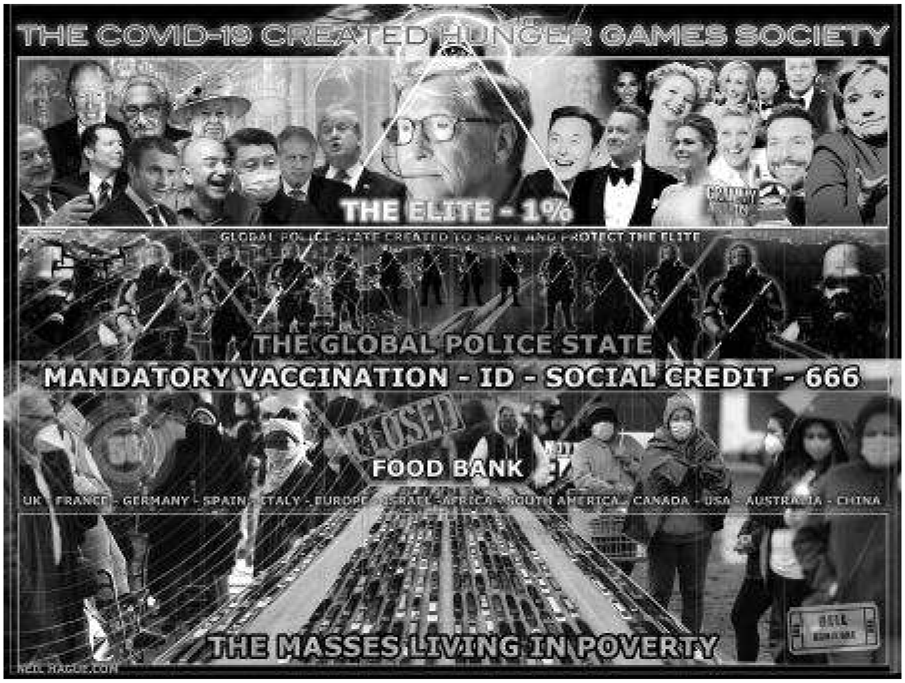

THE COVID-19 CREATED HUNGER GAMES SOCIETY
THE ELITE - 1%
GLOBAL POLICE STATE CREATED TO SERVE AND PROTECT THE ELITE
THE GLOBAL POLICE STATE
MANDATORY VACCINATION - ID - SOCIAL CREDIT - 666
CLOSED
FOOD BANK
UK - FRANCE - GERMANY - SPAIN - ITALY - EUROPE - ISRAEL - AFRICA - SOUTH AMERICA - CANADA - USA - AUSTRALIA - CHINA
THE MASSES LIVING IN POVERTY
NEIL HAGUE.COM

Figure 4: The Hunger Games Society structure I have long warned was planned and now the 'Covid' hoax has made it possible. This is the real reason for lockdowns.

Hired hands

Between the Hunger Games Cult elite and the dependent population is planned to be a vicious military-police state (a fusion of the two into one force). This has been in the making for a long time with police looking ever more like the military and carrying weapons to match. The pandemic scam has seen this process accelerate so fast as lockdown house arrest is brutally enforced by carefully recruited fascist minds and gormless system-servers. The police and military are planned to merge into a centrally-directed world army in a global structure headed by a world government which wouldn't be elected even by the election fixes now in place. The world army is not planned even to be human and instead wars would be fought, primarily against the population, using robot technology controlled by artificial intelligence. I have been warning about this for decades and now militaries around the world are being transformed by this very AI technology. The global regime that I describe is a particular form of fascism known as a technocracy in which decisions are not made by clueless and co-opted politicians but by unelected technocrats – scientists, engineers, technologists and bureaucrats. Cult-owned-and-controlled Silicon Valley giants are examples of technocracy and they already have far more power to direct world events than governments. They are with their censorship *selecting* governments. I know that some are calling the 'Great Reset' a Marxist communist takeover, but fascism and Marxism are different labels for the same tyranny. Tell those who lived in fascist Germany and Stalinist Russia that there was a difference in the way their freedom was deleted and their lives controlled. I could call it a fascist technocracy or a Marxist technocracy and they would be equally accurate. The Hunger Games society with its world government structure would oversee a world army, world central bank and single world cashless currency imposing its will on a microchipped population (Fig 5). Scan its different elements and see how the illusory pandemic is forcing society in this very direction at great speed. Leaders of 23 countries and the World Health Organization (WHO) backed the idea in March, 2021, of a global treaty for 'international cooperation' in 'health emergencies' and nations should 'come together as a global community for peaceful cooperation that extends beyond this crisis'. Cut the Orwellian bullshit and this means another step towards global government. The plan includes a cashless digital money system that I first warned about in 1993. Right at the start of 'Covid' the deeply corrupt Tedros

Adhanom Ghebreyesus, the crooked and merely gofer 'head' of the World Health Organization, said it was possible to catch the 'virus' by touching cash and it was better to use cashless means. The claim was ridiculous nonsense and like the whole 'Covid' mind-trick it was nothing to do with 'health' and everything to do with pushing every aspect of the Cult agenda. As a result of the Tedros lie the use of cash has plummeted. The Cult script involves a single world digital currency that would eventually be technologically embedded in the body. China is a massive global centre for the Cult and if you watch what is happening there you will know what is planned for everywhere. The Chinese government is developing a digital currency which would allow fines to be deducted immediately via AI for anyone caught on camera breaking its fantastic list of laws and the money is going to be programmable with an expiry date to ensure that no one can accrue wealth except the Cult and its operatives.

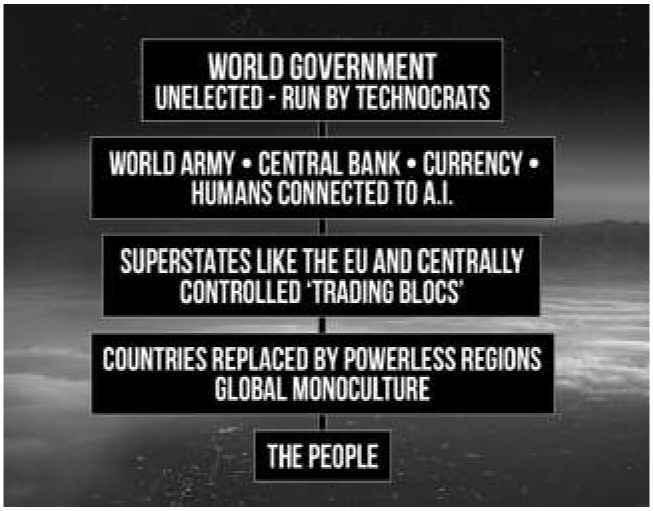

WORLD GOVERNMENT
UNELECTED - RUN BY TECHNOCRATS
WORLD ARMY • CENTRAL BANK • CURRENCY • HUMANS CONNECTED TO A.I.
SUPERSTATES LIKE THE EU AND CENTRALLY CONTROLLED 'TRADING BLOCS'
COUNTRIES REPLACED BY POWERLESS REGIONS
GLOBAL MONOCULTURE
THE PEOPLE

Figure 5: The structure of global control the Cult has been working towards for so long and this has been enormously advanced by the 'Covid' illusion.

Serfdom is so smart

The Cult plan is far wider, extreme, and more comprehensive than even most conspiracy researchers appreciate and I will come to the true depths of deceit and control in the chapters 'Who controls the

Cult?' and 'Escaping Wetiko'. Even the world that we know is crazy enough. We are being deluged with ever more sophisticated and controlling technology under the heading of 'smart'. We have smart televisions, smart meters, smart cards, smart cars, smart driving, smart roads, smart pills, smart patches, smart watches, smart skin, smart borders, smart pavements, smart streets, smart cities, smart communities, smart environments, smart growth, smart planet ... smart *everything* around us. Smart technologies and methods of operation are designed to interlock to create a global Smart Grid connecting the entirety of human society including human minds to create a centrally-dictated 'hive' mind. 'Smart cities' is code for densely-occupied megacities of total surveillance and control through AI. Ever more destructive frequency communication systems like 5G have been rolled out without any official testing for health and psychological effects (colossal). 5G/6G/7G systems are needed to run the Smart Grid and each one becomes more destructive of body and mind. Deleting independent income is crucial to forcing people into these AI-policed prisons by ending private property ownership (except for the Cult elite). The Cult's Great Reset now openly foresees a global society in which no one will own any possessions and everything will be rented while the Cult would own literally everything under the guise of government and corporations. The aim has been to use the lockdowns to destroy sources of income on a mass scale and when the people are destitute and in unrepayable amounts of debt (problem) Cult assets come forward with the pledge to write-off debt in return for handing over all property and possessions (solution). Everything – literally everything including people – would be connected to the Internet via AI. I was warning years ago about the coming Internet of Things (IoT) in which all devices and technology from your car to your fridge would be plugged into the Internet and controlled by AI. Now we are already there with much more to come. The next stage is the Internet of Everything (IoE) which is planned to include the connection of AI to the human brain and body to replace the human mind with a centrally-controlled AI mind. Instead of perceptions being manipulated through control of information and censorship those perceptions would come direct from the Cult through AI. What do you think? You think whatever AI decides that you think. In human terms there would be no individual 'think' any longer. Too incredible? The ravings of a lunatic? Not at all. Cult-owned crazies in Silicon Valley have been telling us the plan for years without explaining the real motivation and calculated implications. These include Google executive and 'futurist' Ray Kurzweil who highlights the year 2030 for when this would be underway. He said:

> Our thinking ... will be a hybrid of biological and non-biological thinking ... humans will be able to extend their limitations and 'think in the cloud' ... We're going to put gateways to the cloud in our brains ... We're going to gradually merge and enhance ourselves ... In my view, that's the nature of being human – we transcend our limitations.
>
> As the technology becomes vastly superior to what we are then the small proportion that is still human gets smaller and smaller and smaller until it's just utterly negligible.

The sales-pitch of Kurzweil and Cult-owned Silicon Valley is that this would make us 'super-human' when the real aim is to make us post-human and no longer 'human' in the sense that we have come to know. The entire global population would be connected to AI and become the centrally-controlled 'hive-mind' of externally-delivered perceptions. The Smart Grid being installed to impose the Cult's will on the world is being constructed to allow particular locations – even one location – to control the whole global system. From these prime control centres, which absolutely include China and Israel, anything connected to the Internet would be switched on or off and manipulated at will. Energy systems could be cut, communication via the Internet taken down, computer-controlled driverless autonomous vehicles driven off the road, medical devices switched off, the potential is limitless given how much AI and Internet connections now run human society. We have seen nothing yet if we allow this to continue. Autonomous vehicle makers are working with law enforcement to produce cars designed to automatically pull over if they detect a police or emergency vehicle flashing from up to 100 feet away. At a police stop the car would be unlocked and the window rolled down automatically. Vehicles would only take you where the computer (the state) allowed. The end of petrol vehicles and speed limiters on all new cars in the UK and EU from 2022 are steps leading to electric computerised transport over which ultimately you have no control. The picture is far bigger even than the Cult global network or web and that will become clear when I get to the nature of the 'spider'. There is a connection between all these happenings and the instigation of DNA-manipulating 'vaccines' (which aren't 'vaccines') justified by the 'Covid' hoax. That connection is the unfolding plan to transform the human body from a biological to a synthetic biological state and this is why synthetic biology is such a fast-emerging discipline of mainstream science. 'Covid vaccines' are infusing self-replicating synthetic genetic material into the cells to cumulatively take us on the Totalitarian Tiptoe from Human 1.0 to the synthetic biological Human 2.0 which will be physically and perceptually attached to the Smart Grid to one hundred percent control every thought, perception and deed. Humanity needs to wake up and *fast*.

This is the barest explanation of where the 'outcome' is planned to go but it's enough to see the journey happening all around us. Those new to this information will already see 'Covid' in a whole new context. I will add much more detail as we go along, but for the minutiae evidence see my mega-works, *The Answer, The Trigger* and *Everything You Need to Know But Have Never Been Told*.

Now – how does a Renegade Mind see the 'world'?

CHAPTER TWO

Renegade Perception

It is one thing to be clever and another to be wise
George R.R. Martin

A simple definition of the difference between a programmed mind and a Renegade Mind would be that one sees only dots while the other connects them to see the picture. Reading reality with accuracy requires the observer to (a) know the planned outcome and (b) realise that everything, but *everything*, is connected.

The entirety of infinite reality is connected – that's its very nature – and with human society an expression of infinite reality the same must apply. Simple cause and effect is a connection. The effect is triggered by the cause and the effect then becomes the cause of another effect. Nothing happens in isolation because it *can't*. Life in whatever reality is simple choice and consequence. We make choices and these lead to consequences. If we don't like the consequences we can make different choices and get different consequences which lead to other choices and consequences. The choice and the consequence are not only connected they are indivisible. You can't have one without the other as an old song goes. A few cannot control the world unless those being controlled allow that to happen – cause and effect, choice and consequence. Control – who has it and who doesn't – is a two-way process, a symbiotic relationship, involving the controller and controlled. 'They took my freedom away!!' Well, yes, but you also gave it to them. Humanity is subjected to mass control because humanity has acquiesced to that control. This is all cause and effect and literally a case of give and take. In the same way world events of every kind are connected and the Cult works incessantly to sell the illusion of the random and coincidental to maintain the essential (to them) perception of dots that hide the picture. Renegade Minds know this and constantly scan the world for patterns of connection. This is absolutely pivotal in understanding the happenings in the world and without that perspective clarity is impossible. First you know the planned outcome and then you identify the steps on the journey – the day-by-day apparently random which, when connected in relation to the outcome, no longer appear as individual events, but as the proverbial *chain* of events leading in the same direction. I'll give you some examples:

Political puppet show

We are told to believe that politics is 'adversarial' in that different parties with different beliefs engage in an endless tussle for power. There may have been some truth in that up to a point – and only a point – but today divisions between 'different' parties are rhetorical not ideological. Even the rhetorical is fusing into one-speak as the parties eject any remaining free thinkers while others succumb to the ever-gathering intimidation of anyone with the 'wrong' opinion. The Cult is not a new phenomenon and can be traced back thousands of years as my books have documented. Its intergenerational initiates have been manipulating events with increasing effect the more that global power has been centralised. In ancient times the Cult secured control through the system of monarchy in which 'special' bloodlines (of which more later) demanded the right to rule as kings and queens simply by birthright and by vanquishing others who claimed the same birthright. There came a time, however, when people had matured enough to see the unfairness of such tyranny and demanded a say in who governed them. Note the word – *governed* them. Not served them – *governed* them, hence government defined as 'the political direction and control exercised over the actions of the members, citizens, or inhabitants of communities, societies, and states; direction of the affairs of a state, community, etc.' Governments exercise control over rather than serve just like the monarchies before them. Bizarrely there are still countries like the United Kingdom which are ruled by a monarch *and* a government that officially answers to the monarch. The UK head of state and that of Commonwealth countries such as Canada, Australia and New Zealand is 'selected' by who in a *single family* had unprotected sex with whom and in what order. Pinch me it can't be true. Ouch! Shit, it is. The demise of monarchies in most countries offered a potential vacuum in which some form of free and fair society could arise and the Cult had that base covered. Monarchies had served its interests but they couldn't continue in the face of such widespread opposition and, anyway, replacing a 'royal' dictatorship that people could see with a dictatorship 'of the people' hiding behind the concept of 'democracy' presented far greater manipulative possibilities and ways of hiding coordinated tyranny behind the illusion of 'freedom'.

Democracy is quite wrongly defined as government selected by the population. This is not the case at all. It is government selected by *some* of the population (and then only in theory). This 'some' doesn't even have to be the majority as we have seen so often in first-past-the-post elections in which the so-called majority party wins fewer votes than the 'losing' parties combined. Democracy can give total power to a party in government from a minority of the votes cast. It's a sleight of hand to sell tyranny as freedom. Seventy-four million Trump-supporting Americans didn't vote for the 'Democratic' Party of Joe Biden in the distinctly dodgy election in 2020 and yet far from acknowledging the wishes and feelings of that great percentage of American society the Cult-owned Biden government set out from day one to destroy them and their right to a voice and opinion. Empty shell Biden and his Cult handlers said they were doing this to 'protect democracy'. Such is the level of lunacy and sickness to which politics has descended. Connect the dots and relate them to the desired outcome – a world government run by self-appointed technocrats and no longer even elected politicians. While operating through its political agents in government the Cult is at the same time encouraging public distain for politicians by putting idiots and incompetents in theoretical power on the road to deleting them. The idea is to instil a public reaction that says of the technocrats: 'Well, they couldn't do any worse than the pathetic politicians.' It's all about controlling perception and Renegade Minds can see through that while programmed minds cannot when they are ignorant of both the planned outcome and the manipulation techniques employed to secure that end. This knowledge can be learned, however, and fast if people choose to get informed.

Politics may at first sight appear very difficult to control from a central point. I mean look at the 'different' parties and how would you be able to oversee them all and their constituent parts? In truth, it's very straightforward because of their structure. We are back to the pyramid of imposition and acquiescence. Organisations are structured in the same way as the system as a whole. Political parties are not open forums of free expression. They are hierarchies. I was a national spokesman for the British Green Party which claimed to be a different kind of politics in which influence and power was devolved; but I can tell you from direct experience – and it's far worse now – that Green parties are run as hierarchies like all the others however much they may try to hide that fact or kid themselves that it's not true. A very few at the top of all political parties are directing policy and personnel. They decide if you are elevated in the party or serve as a government minister and to do that you have to be a yes man or woman. Look at all the maverick political thinkers who never ascended the greasy pole. If you want to progress within the party or reach 'high-office' you need to fall into line and conform. Exceptions to this are rare indeed. Should you want to run for parliament or Congress you have to persuade the local or state level of the party to select you and for that you need to play the game as dictated by the hierarchy. If you secure election and wish to progress within the greater structure you need to go on conforming to what is acceptable to those running the hierarchy from the peak of the pyramid. Political parties are perceptual gulags and the very fact that there are party 'Whips' appointed to 'whip' politicians into voting the way the hierarchy demands exposes the ridiculous idea that politicians are elected to serve the people they are supposed to represent. Cult operatives and manipulation has long seized control of major parties that have any chance of forming a government and at least most of those that haven't. A new party forms and the Cult goes to work to infiltrate and direct. This has reached such a level today that you see video compilations of 'leaders' of all parties whether Democrats, Republicans, Conservative, Labour and Green parroting the same Cult mantra of 'Build Back Better' and the 'Great Reset' which are straight off the Cult song-sheet to describe the transformation of global society in response to the Cult-instigated hoaxes of the 'Covid pandemic' and human-caused 'climate change'. To see Caroline Lucas, the Green Party MP that I knew when I was in the party in the 1980s, speaking in support of plans proposed by Cult operative Klaus Schwab representing the billionaire global elite is a real head-shaker.

Many parties – one master

The party system is another mind-trick and was instigated to change the nature of the dictatorship by swapping 'royalty' for dark suits that people believed – though now ever less so – represented their interests. Understanding this trick is to realise that a single force (the Cult) controls all parties either directly in terms of the major ones or through manipulation of perception and ideology with others. You don't need to manipulate Green parties to demand your transformation of society in the name of 'climate change' when they are obsessed with the lie that this is essential to 'save the planet'. You just give them a platform and away they go serving your interests while believing they are being environmentally virtuous. America's political structure is a perfect blueprint for how the two or multi-party system is really a one-party state. The Republican Party is controlled from one step back in the shadows by a group made up of billionaires and their gofers known as neoconservatives or Neocons.

I have exposed them in fine detail in my books and they were the driving force behind the policies of the imbecilic presidency of Boy George Bush which included 9/11 (see *The Trigger* for a comprehensive demolition of the official story), the subsequent 'war on terror' (war *of* terror) and the invasions of Afghanistan and Iraq. The latter was a No-Problem-Reaction-Solution based on claims by Cult operatives, including Bush and British Prime Minister Tony Blair, about Saddam Hussein's 'weapons of mass destruction' which did not exist as war criminals Bush and Blair well knew.

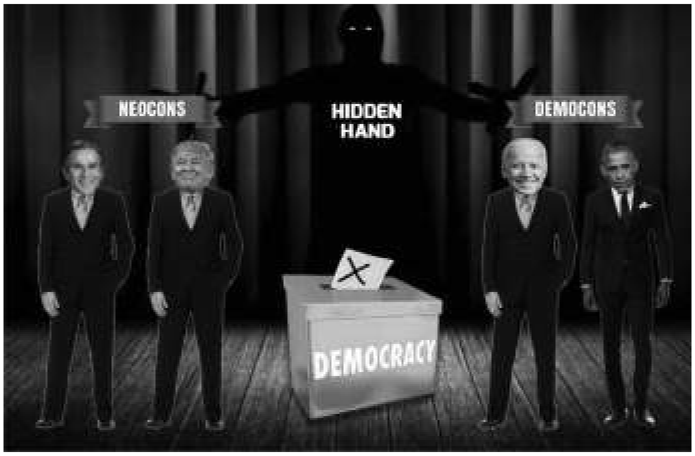


Figure 6: Different front people, different parties – same control system.

The Democratic Party has its own 'Neocon' group controlling from the background which I call the 'Democons' and here's the penny-drop – the Neocons and Democons answer to the same masters one step further back into the shadows (Fig 6). At that level of the Cult the Republican and Democrat parties are controlled by the same people and no matter which is in power the Cult is in power. This is how it works in almost every country and certainly in Britain with Conservative, Labour, Liberal Democrat and Green parties now all on the same page whatever the rhetoric may be in their feeble attempts to appear different. Neocons operated at the time of Bush through a think tank called The Project for the New American Century which in September, 2000, published a document entitled *Rebuilding America's Defenses: Strategies, Forces, and Resources*

For a New Century demanding that America fight 'multiple, simultaneous major theatre wars' as a 'core mission' to force regime-change in countries including Iraq, Libya and Syria. Neocons arranged for Bush ('Republican') and Blair ('Labour Party') to front-up the invasion of Iraq and when they departed the Democons orchestrated the targeting of Libya and Syria through Barack Obama ('Democrat') and British Prime Minister David Cameron ('Conservative Party'). We have 'different' parties and 'different' people, but the same unfolding script. The more the Cult has seized the reigns of parties and personnel the more their policies have transparently pursued the same agenda to the point where the fascist 'Covid' impositions of the Conservative junta of Jackboot Johnson in Britain were opposed by the Labour Party because they were not fascist enough. The Labour Party is likened to the US Democrats while the Conservative Party is akin to a British version of the Republicans and on both sides of the Atlantic they all speak the same language and support the direction demanded by the Cult although some more enthusiastically than others. It's a similar story in country after country because it's all centrally controlled. Oh, but what about Trump? I'll come to him shortly. Political 'choice' in the 'party' system goes like this: You vote for Party A and they get into government. You don't like what they do so next time you vote for Party B and they get into government. You don't like what they do when it's pretty much the same as Party A and why wouldn't that be with both controlled by the same force? Given that only two, sometimes three, parties have any chance of forming a government to get rid of Party B that you don't like you have to vote again for Party A which … you don't like. This, ladies and gentlemen, is what they call 'democracy' which we are told – wrongly – is a term interchangeable with 'freedom'.

The cult of cults

At this point I need to introduce a major expression of the Global Cult known as Sabbatian-Frankism. Sabbatian is also spelt as Sabbatean. I will summarise here. I have published major exposés and detailed background in other works. Sabbatian-Frankism combines the names of two frauds posing as 'Jewish' men, Sabbatai Zevi (1626-1676), a rabbi, black magician and occultist who proclaimed he was the Jewish messiah; and Jacob Frank (1726-1791), the Polish 'Jew', black magician and occultist who said he was the reincarnation of 'messiah' Zevi and biblical patriarch Jacob. They worked across two centuries to establish the Sabbatian-Frankist cult that plays a major, indeed central, role in the manipulation of human society by the Global Cult which has its origins much further back in history than Sabbatai Zevi. I should emphasise two points here in response to the shrill voices that will scream 'anti-Semitism': (1) Sabbatian-Frankists are NOT Jewish and only pose as such to hide their cult behind a Jewish façade; and (2) my information about this cult has come from Jewish sources who have long realised that their society and community has been infiltrated and taken over by interloper Sabbatian-Frankists. Infiltration has been the foundation technique of Sabbatian-Frankism from its official origin in the 17th century. Zevi's Sabbatian sect attracted a massive following described as the biggest messianic movement in Jewish history, spreading as far as Africa and Asia, and he promised a return for the Jews to the 'Promised Land' of Israel. Sabbatianism was not Judaism but an inversion of everything that mainstream Judaism stood for. So much so that this sinister cult would have a feast day when Judaism had a fast day and whatever was forbidden in Judaism the Sabbatians were encouraged and even commanded to do. This included incest and what would be today called Satanism. Members were forbidden to marry outside the sect and there was a system of keeping their children ignorant of what they were part of until they were old enough to be trusted not to unknowingly reveal anything to outsiders. The same system is employed to this day by the Global Cult in general which Sabbatian-Frankism has enormously influenced and now largely controls.

Zevi and his Sabbatians suffered a setback with the intervention by the Sultan of the Islamic Ottoman Empire in the Middle East and what is now the Republic of Turkey where Zevi was located. The

Sultan gave him the choice of proving his ‘divinity’, converting to Islam or facing torture and death. Funnily enough Zevi chose to convert or at least appear to. Some of his supporters were disillusioned and drifted away, but many did not with 300 families also converting – only in theory – to Islam. They continued behind this Islamic smokescreen to follow the goals, rules and rituals of Sabbatianism and became known as ‘crypto-Jews’ or the ‘Dönmeh’ which means ‘to turn’. This is rather ironic because they didn’t ‘turn’ and instead hid behind a fake Islamic persona. The process of appearing to be one thing while being very much another would become the calling card of Sabbatianism especially after Zevi’s death and the arrival of the Satanist Jacob Frank in the 18th century when the cult became Sabbatian-Frankism and plumbed still new depths of depravity and infiltration which included – still includes – human sacrifice and sex with children. Wherever Sabbatians go paedophilia and Satanism follow and is it really a surprise that Hollywood is so infested with child abuse and Satanism when it was established by Sabbatian-Frankists and is still controlled by them? Hollywood has been one of the prime vehicles for global perceptual programming and manipulation. How many believe the version of ‘history’ portrayed in movies when it is a travesty and inversion (again) of the truth? Rabbi Marvin Antelman describes Frankism in his book, *To Eliminate the Opiate*, as ‘a movement of complete evil’ while Jewish professor Gershom Scholem said of Frank in *The Messianic Idea in Judaism*: ‘In all his actions [he was] a truly corrupt and degenerate individual ... one of the most frightening phenomena in the whole of Jewish history.’ Frank was excommunicated by traditional rabbis, as was Zevi, but Frank was undeterred and enjoyed vital support from the House of Rothschild, the infamous banking dynasty whose inner-core are Sabbatian-Frankists and not Jews. Infiltration of the Roman Church and Vatican was instigated by Frank with many Dönmeh ‘turning’ again to convert to Roman Catholicism with a view to hijacking the reins of power. This was the ever-repeating modus operandi and continues to be so. Pose as an advocate of the religion, culture or country that you want to control and then manipulate your people into the positions of authority and influence largely as advisers, administrators and Svengalis for those that appear to be in power. They did this with Judaism, Christianity (Christian Zionism is part of this), Islam and other religions and nations until Sabbatian-Frankism spanned the world as it does today.

Sabbatian Saudis and the terror network

One expression of the Sabbatian-Frankist Dönmeh within Islam is the ruling family of Saudi Arabia, the House of Saud, through which came the vile distortion of Islam known as Wahhabism. This is the violent creed followed by terrorist groups like Al-Qaeda and ISIS or Islamic State. Wahhabism is the hand-chopping, head-chopping 'religion' of Saudi Arabia which is used to keep the people in a constant state of fear so the interloper House of Saud can continue to rule. Al-Qaeda and Islamic State were lavishly funded by the House of Saud while being created and directed by the Sabbatian-Frankist network in the United States that operates through the Pentagon, CIA and the government in general of whichever 'party'. The front man for the establishment of Wahhabism in the middle of the 18th century was a Sabbatian-Frankist 'crypto-Jew' posing as Islamic called Muhammad ibn Abd al-Wahhab. His daughter would marry the son of Muhammad bin Saud who established the first Saudi state before his death in 1765 with support from the British Empire. Bin Saud's successors would establish modern Saudi Arabia in league with the British and Americans in 1932 which allowed them to seize control of Islam's major shrines in Mecca and Medina. They have dictated the direction of Sunni Islam ever since while Iran is the major centre of the Shiite version and here we have the source of at least the public conflict between them. The Sabbatian network has used its Wahhabi extremists to carry out Problem-Reaction-Solution terrorist attacks in the name of 'Al-Qaeda' and 'Islamic State' to justify a devastating 'war on terror', ever-increasing surveillance of the population and to terrify people into compliance. Another insight of the Renegade Mind is the streetwise understanding that just because a country, location or people are attacked doesn't mean that those apparently representing that country, location or people are not behind the attackers. Often they are *orchestrating* the attacks because of the societal changes that can be then justified in the name of 'saving the population from terrorists'.

I show in great detail in *The Trigger* how Sabbatian-Frankists were the real perpetrators of 9/11 and not '19 Arab hijackers' who were blamed for what happened. Observe what was justified in the name of 9/11 alone in terms of Middle East invasions, mass surveillance and control that fulfilled the demands of the Project for the New American Century document published by the Sabbatian Neocons. What appear to be enemies are on the deep inside players on the same Sabbatian team. Israel and Arab 'royal' dictatorships are all ruled by Sabbatians and the recent peace agreements between Israel and Saudi Arabia, the United Arab Emirates (UAE) and others are only making formal what has always been the case behind the scenes. Palestinians who have been subjected to grotesque tyranny since Israel was bombed and terrorised into existence in 1948 have never stood a chance. Sabbatian-Frankists have controlled Israel (so the constant theme of violence and war which Sabbatians love) and they have controlled the Arab countries that Palestinians have looked to for real support that never comes. 'Royal families' of the Arab world in Saudi Arabia, Bahrain, UAE, etc., are all Sabbatians with allegiance to the aims of the cult and not what is best for their Arabic populations. They have stolen the oil and financial resources from their people by false claims to be 'royal dynasties' with a genetic right to rule and by employing vicious militaries to impose their will.

Satanic 'illumination'

The Satanist Jacob Frank formed an alliance in 1773 with two other Sabbatians, Mayer Amschel Rothschild (1744-1812), founder of the Rothschild banking dynasty, and Jesuit-educated fraudulent Jew, Adam Weishaupt, and this led to the formation of the Bavarian Illuminati, firstly under another name, in 1776. The Illuminati would be the manipulating force behind the French Revolution (1789-1799) and was also involved in the American Revolution (1775-1783) before and after the Illuminati's official creation. Weishaupt would later become (in public) a Protestant Christian in archetypal Sabbatian style. I read that his name can be decoded as Adam-Weishaupt or 'the first man to lead those who know'. He wasn't a leader in the sense that he was a subordinate, but he did lead those below him in a crusade of transforming human society that still continues today. The theme was confirmed as early as 1785 when a horseman courier called Lanz was reported to be struck by lighting and extensive Illuminati documents were found in his saddlebags. They made the link to Weishaupt and detailed the plan for world takeover. Current events with 'Covid' fascism have been in the making for a very long time. Jacob Frank was jailed for 13 years by the Catholic Inquisition after his arrest in 1760 and on his release he headed for Frankfurt, Germany, home city and headquarters of the House of Rothschild where the alliance was struck with Mayer Amschel Rothschild and Weishaupt. Rothschild arranged for Frank to be given the title of Baron and he became a wealthy nobleman with a big following of Jews in Germany, the Austro-Hungarian Empire and other European countries. Most of them would have believed he was on their side.

The name 'Illuminati' came from the Zohar which is a body of works in the Jewish mystical 'bible' called the Kabbalah. 'Zohar' is the foundation of Sabbatian-Frankist belief and in Hebrew 'Zohar' means 'splendour', 'radiance', 'illuminated', and so we have 'Illuminati'. They claim to be the 'Illuminated Ones' from their knowledge systematically hidden from the human population and passed on through generations of carefully-chosen initiates in the global secret society network or Cult. Hidden knowledge includes an awareness of the Cult agenda for the world and the nature of our collective reality that I will explore later. Cult 'illumination' is symbolised by the torch held by the Statue of Liberty which was gifted to New York by French Freemasons in Paris who knew exactly what it represents. 'Liberty' symbolises the goddess worshipped in

Babylon as Queen Semiramis or Ishtar. The significance of this will become clear. Notice again the ubiquitous theme of inversion with the Statue of 'Liberty' really symbolising mass control (Fig 7). A mirror-image statute stands on an island in the River Seine in Paris from where New York Liberty originated (Fig 8). A large replica of the Liberty flame stands on top of the Pont de l'Alma tunnel in Paris where Princess Diana died in a Cult ritual described in *The Biggest Secret*. Lucifer 'the light bringer' is related to all this (and much more as we'll see) and 'Lucifer' is a central figure in Sabbatian-Frankism and its associated Satanism. Sabbatians reject the Jewish Torah, or Pentateuch, the 'five books of Moses' in the Old Testament known as Genesis, Exodus, Leviticus, Numbers, and Deuteronomy which are claimed by Judaism and Christianity to have been dictated by 'God' to Moses on Mount Sinai. Sabbatians say these do not apply to them and they seek to replace them with the Zohar to absorb Judaism and its followers into their inversion which is an expression of a much greater global inversion. They want to delete all religions and force humanity to worship a one-world religion – Sabbatian Satanism that also includes worship of the Earth goddess. Satanic themes are being more and more introduced into mainstream society and while Christianity is currently the foremost target for destruction the others are planned to follow.

Figure 7: The Cult goddess of Babylon disguised as the Statue of Liberty holding the flame of Lucifer the 'light bringer'.

Radio France

Figure 8: Liberty's mirror image in Paris where the New York version originated.

Marx brothers

Rabbi Marvin Antelman connects the Illuminati to the Jacobins in *To Eliminate the Opiate* and Jacobins were the force behind the French Revolution. He links both to the Bund der Gerechten, or League of the Just, which was the network that inflicted communism/Marxism on the world. Antelman wrote:

> The original inner circle of the Bund der Gerechten consisted of born Catholics, Protestants and Jews [Sabbatian-Frankist infiltrators], and those representatives of respective subdivisions formulated schemes for the ultimate destruction of their faiths. The heretical Catholics laid plans which they felt would take a century or more for the ultimate destruction of the church; the apostate Jews for the ultimate destruction of the Jewish religion.

Sabbatian-created communism connects into this anti-religion agenda in that communism does not allow for the free practice of religion. The Sabbatian 'Bund' became the International Communist Party and Communist League and in 1848 'Marxism' was born with the Communist Manifesto of Sabbatian assets Karl Marx and Friedrich Engels. It is absolutely no coincidence that Marxism, just a different name for fascist and other centrally-controlled tyrannies, is being imposed worldwide as a result of the 'Covid' hoax and nor that Marxist/fascist China was the place where the hoax originated. The reason for this will become very clear in the chapter 'Covid: The calculated catastrophe'. The so-called 'Woke' mentality has hijacked traditional beliefs of the political left and replaced them with far-right make-believe 'social justice' better known as Marxism. Woke will, however, be swallowed by its own perceived 'revolution' which is really the work of billionaires and billionaire corporations feigning being 'Woke'. Marxism is being touted by Wokers as a replacement for 'capitalism' when we don't have 'capitalism'. We have cartelism in which the market is stitched up by the very Cult billionaires and corporations bankrolling Woke. Billionaires love Marxism which keeps the people in servitude while they control from the top. Terminally naïve Wokers think they are 'changing the world' when it's the Cult that is doing the changing and when they have played their vital part and become surplus to requirements they, too, will be targeted. The Illuminati-Jacobins were behind the period known as 'The Terror' in the French Revolution in 1793 and 1794 when Jacobin Maximillian de Robespierre and his Orwellian 'Committee of Public Safety' killed 17,000 'enemies of the Revolution' who had once been 'friends of the Revolution'. Karl Marx (1818-1883), whose Sabbatian creed of Marxism has cost the lives of at least 100 million people, is a hero once again to Wokers who have been systematically kept ignorant of real history by their 'education' programming. As a result they now promote a Sabbatian 'Marxist' abomination destined at some point to consume them. Rabbi Antelman, who spent decades researching the Sabbatian plot, said of the League of the Just and Karl Marx:

> Contrary to popular opinion Karl Marx did not originate the Communist Manifesto. He was paid for his services by the League of the Just, which was known in its country of origin, Germany, as the Bund der Geaechteten.

Antelman said the text attributed to Marx was the work of other people and Marx 'was only repeating what others already said'. Marx was 'a hired hack – lackey of the wealthy Illuminists'. Marx famously said that religion was the 'opium of the people' (part of the Sabbatian plan to demonise religion) and Antelman called his books, *To Eliminate the Opiate*. Marx was born Jewish, but his family converted to Christianity (Sabbatian modus operandi) and he attacked Jews, not least in his book, *A World Without Jews*. In doing so he supported the Sabbatian plan to destroy traditional Jewishness and Judaism which we are clearly seeing today with the vindictive targeting of orthodox Jews by the Sabbatian government of Israel over 'Covid' laws. I don't follow any religion and it has done much damage to the world over centuries and acted as a perceptual straightjacket. Renegade Minds, however, are always asking *why* something is being done. It doesn't matter if they agree or disagree with what is happening – *why* is it happening is the question. The 'why?' can be answered with regard to religion in that religions create interacting communities of believers when the Cult wants to dismantle all discourse, unity and interaction (see 'Covid' lockdowns) and the ultimate goal is to delete all religions for a one-world religion of Cult Satanism worshipping their 'god' of which more later. We see the same 'why?' with gun control in America. I don't have guns and don't want them, but why is the Cult seeking to disarm the population at the same time that law enforcement agencies are armed to their molars and why has every tyrant in history sought to disarm people before launching the final takeover? They include Hitler, Stalin, Pol Pot and Mao who followed confiscation with violent seizing of power. You know it's a Cult agenda by the people who immediately race to the microphones to exploit dead people in multiple shootings. Ultra-Zionist Cult lackey Senator Chuck Schumer was straight on the case after ten people were killed in Boulder, Colorado in March, 2121. Simple rule … if Schumer wants it the Cult wants it and the same with his ultra-Zionist mate the wild-eyed Senator Adam Schiff. At the same time they were calling for the disarmament of Americans, many of whom live a long way from a police response, Schumer, Schiff and the rest of these pampered clowns were sitting on Capitol Hill behind a razor-wired security fence protected by thousands of armed troops in addition to their own armed bodyguards. Mom and pop in an isolated home? They're just potential mass shooters.

Zion Mainframe

Sabbatian-Frankists and most importantly the Rothschilds were behind the creation of ‘Zionism’, a political movement that demanded a Jewish homeland in Israel as promised by Sabbatai Zevi. The very symbol of Israel comes from the German meaning of the name Rothschild. Dynasty founder Mayer Amschel Rothschild changed the family name from Bauer to Rothschild, or ‘Red-Shield’ in German, in deference to the six-pointed ‘Star of David’ hexagram displayed on the family’s home in Frankfurt. The symbol later appeared on the flag of Israel after the Rothschilds were centrally involved in its creation. Hexagrams are not a uniquely Jewish symbol and are widely used in occult (‘hidden’) networks often as a symbol for Saturn (see my other books for why). Neither are Zionism and Jewishness interchangeable. Zionism is a political movement and philosophy and not a ‘race’ or a people. Many Jews oppose Zionism and many non-Jews, including US President Joe Biden, call themselves Zionists as does Israel-centric Donald Trump. America’s support for the Israel government is pretty much a gimme with ultra-Zionist billionaires and corporations providing fantastic and dominant funding for both political parties. Former Congresswoman Cynthia McKinney has told how she was approached immediately she ran for office to ‘sign the pledge’ to Israel and confirm that she would always vote in that country’s best interests. All American politicians are approached in this way. Anyone who refuses will get no support or funding from the enormous and all-powerful Zionist lobby that includes organisations like mega-lobby group AIPAC, the American Israel Public Affairs Committee. Trump’s biggest funder was ultra-Zionist casino and media billionaire Sheldon Adelson while major funders of the Democratic Party include ultra-Zionist George Soros and ultra-Zionist financial and media mogul, Haim Saban. Some may reel back at the suggestion that Soros is an Israel-firster (Sabbatian-controlled Israel-firster), but Renegade Minds watch the actions not the words and everywhere Soros donates his billions the Sabbatian agenda benefits. In the spirit of Sabbatian inversion Soros pledged $1 billion for a new university network to promote ‘liberal values and tackle intolerance’. He made the announcement during his annual speech at the Cult-owned World Economic Forum in Davos, Switzerland, in January, 2020, after his ‘harsh criticism’ of ‘authoritarian rulers’ around the world. You can only laugh at such brazen mendacity. How *he* doesn’t laugh is the mystery. Translated from the Orwellian ‘liberal values and tackle intolerance’ means teaching non-white people to hate white people and for white people to loathe themselves for being born white. The reason for that will become clear.

The ‘Anti-Semitism’ fraud

Zionists support the Jewish homeland in the land of Palestine which has been the Sabbatian-Rothschild goal for so long, but not for the benefit of Jews. Sabbatians and their global Anti-Semitism Industry have skewed public and political opinion to equate opposing the violent extremes of Zionism to be a blanket attack and condemnation of all Jewish people. Sabbatians and their global Anti-Semitism Industry have skewed public and political opinion to equate opposing the violent extremes of Zionism to be a blanket attack and condemnation of all Jewish people. This is nothing more than a Sabbatian protection racket to stop legitimate investigation and exposure of their agendas and activities. The official definition of ‘anti-Semitism’ has more recently been expanded to include criticism of Zionism – a *political movement* – and this was done to further stop exposure of Sabbatian infiltrators who created Zionism as we know it today in the 19th century. Renegade Minds will talk about these subjects when they know the shit that will come their way. People must decide if they want to know the truth or just cower in the corner in fear of what others will say. Sabbatians have been trying to label me as ‘anti-Semitic’ since the 1990s as I have uncovered more and more about their background and agendas. Useless, gutless, fraudulent ‘journalists’ then just repeat the smears without question and on the day I was writing this section a pair of unquestioning repeaters called Ben Quinn and Archie Bland (how appropriate) outright called me an ‘anti-Semite’ in the establishment propaganda sheet, the London *Guardian*, with no supporting evidence. The

Sabbatian Anti-Semitism Industry said so and who are they to question that? They wouldn’t dare. Ironically ‘Semitic’ refers to a group of languages in the Middle East that are almost entirely Arabic. ‘Anti-Semitism’ becomes ‘anti-Arab’ which if the consequences of this misunderstanding were not so grave would be hilarious. Don’t bother telling Quinn and Bland. I don’t want to confuse them, bless ‘em. One reason I am dubbed ‘anti-Semitic’ is that I wrote in the 1990s that Jewish operatives (Sabbatians) were heavily involved in the Russian Revolution when Sabbatians overthrew the Romanov dynasty. This apparently made me ‘anti-Semitic’. Oh, really? Here is a section from *The Trigger*:

> British journalist Robert Wilton confirmed these themes in his 1920 book *The Last Days of the Romanovs* when he studied official documents from the Russian government to identify the members of the Bolshevik ruling elite between 1917 and 1919. The Central Committee included 41 Jews among 62 members; the Council of the People’s Commissars had 17 Jews out of 22 members; and 458 of the 556 most important Bolshevik positions between 1918 and 1919 were occupied by Jewish people. Only 17 were Russian. Then there were the 23 Jews among the 36 members of the vicious Cheka Soviet secret police established in 1917 who would soon appear all across the country.
>
> Professor Robert Service of Oxford University, an expert on 20th century Russian history, found evidence that [‘Jewish’] Leon Trotsky had sought to make sure that Jews were enrolled in the Red Army and were disproportionately represented in the Soviet civil bureaucracy that included the Cheka which performed mass arrests, imprisonment and executions of ‘enemies of the people’. A US State Department Decimal File (861.00/5339) dated November 13th, 1918, names [Rothschild banking agent in America] Jacob Schiff and a list of ultra-Zionists as funders of the Russian Revolution leading to claims of a ‘Jewish plot’, but the key point missed by all is they were not ‘Jews’ – they were Sabbatian-Frankists.

Britain’s Winston Churchill made the same error by mistake or otherwise. He wrote in a 1920 edition of the *Illustrated Sunday Herald* that those behind the Russian revolution were part of a ‘worldwide conspiracy for the overthrow of civilisation and for the reconstitution of society on the basis of arrested development, of envious malevolence, and impossible equality’ (see ‘Woke’ today because that has been created by the same network). Churchill said there was no need to exaggerate the part played in the creation of Bolshevism and in the actual bringing about of the Russian

Revolution ‘by these international and for the most part atheistical Jews’ [‘atheistical Jews’ = Sabbatians]. Churchill said it is certainly a very great one and probably outweighs all others: ‘With the notable exception of Lenin, the majority of the leading figures are Jews.’ He went on to describe, knowingly or not, the Sabbatian modus operandi of placing puppet leaders nominally in power while they control from the background:

> Moreover, the principal inspiration and driving power comes from the Jewish leaders. Thus Tchitcherin, a pure Russian, is eclipsed by his nominal subordinate, Litvinoff, and the influence of Russians like Bukharin or Lunacharski cannot be compared with the power of Trotsky, or of Zinovieff, the Dictator of the Red Citadel (Petrograd), or of Krassin or Radek – all Jews. In the Soviet institutions the predominance of Jews is even more astonishing. And the prominent, if not indeed the principal, part in the system of terrorism applied by the Extraordinary Commissions for Combatting Counter-Revolution has been taken by Jews, and in some notable cases by Jewesses.

What I said about seriously disproportionate involvement in the Russian Revolution by Jewish ‘revolutionaries’ (Sabbatians) is provable fact, but truth is no defence against the Sabbatian Anti-Semitism Industry, its repeater parrots like Quinn and Bland, and the now breathtaking network of so-called ‘Woke’ ‘anti-hate’ groups with interlocking leaderships and funding which have the role of discrediting and silencing anyone who gets too close to exposing the Sabbatians. We have seen ‘truth is no defence’ confirmed in legal judgements with the Saskatchewan Human Rights Commission in Canada decreeing this: ‘Truthful statements can be presented in a manner that would meet the definition of hate speech, and not all truthful statements must be free from restriction.’ Most ‘anti-hate’ activists, who are themselves consumed by hatred, are too stupid and ignorant of the world to know how they are being used. They are far too far up their own virtue-signalling arses and it’s far too dark for them to see anything.

The ‘revolution’ game

The background and methods of the ‘Russian’ Revolution are straight from the Sabbatian playbook seen in the French Revolution and endless others around the world that appear to start as a revolution of the people against tyrannical rule and end up with a regime change to more tyrannical rule overtly or covertly. Wars, terror attacks and regime overthrows follow the Sabbatian cult through history with its agents creating them as Problem-Reaction-Solutions to remove opposition on the road to world domination. Sabbatian dots connect the Rothschilds with the Illuminati, Jacobins of the French Revolution, the 'Bund' or League of the Just, the International Communist Party, Communist League and the Communist Manifesto of Karl Marx and Friedrich Engels that would lead to the Rothschild-funded Russian Revolution. The sequence comes under the heading of 'creative destruction' when you advance to your global goal by continually destroying the status quo to install a new status quo which you then also destroy. The two world wars come to mind. With each new status quo you move closer to your planned outcome. Wars and mass murder are to Sabbatians a collective blood sacrifice ritual. They are obsessed with death for many reasons and one is that death is an inversion of life. Satanists and Sabbatians are obsessed with death and often target churches and churchyards for their rituals. Inversion-obsessed Sabbatians explain the use of inverted symbolism including the *inverted* pentagram and *inverted* cross. The inversion of the cross has been related to targeting Christianity, but the cross was a religious symbol long before Christianity and its inversion is a statement about the Sabbatian mentality and goals more than any single religion.

Sabbatians operating in Germany were behind the rise of the occult-obsessed Nazis and the subsequent Jewish exodus from Germany and Europe to Palestine and the United States after World War Two. The Rothschild dynasty was at the forefront of this both as political manipulators and by funding the operation. Why would Sabbatians help to orchestrate the horrors inflicted on Jews by the Nazis and by Stalin after they organised the Russian Revolution? Sabbatians hate Jews and their religion, that's why. They pose as Jews and secure positions of control within Jewish society and play the 'anti-Semitism' card to protect themselves from exposure through a global network of organisations answering to the Sabbatian-created-and-controlled globe-spanning intelligence network that involves a stunning web of military-intelligence operatives and operations for a tiny country of just nine million. Among them are Jewish assets who are not Sabbatians but have been convinced by them that what they are doing is for the good of Israel and the Jewish community to protect them from what they have been programmed since childhood to believe is a Jew-hating hostile world. The Jewish community is just a highly convenient cover to hide the true nature of Sabbatians. Anyone getting close to exposing their game is accused by Sabbatian place-people and gofers of 'anti-Semitism' and claiming that all Jews are part of a plot to take over the world. I am not saying that. I am saying that Sabbatians – the *real* Jew-haters – have infiltrated the Jewish community to use them both as a cover and an 'anti-Semitic' defence against exposure. Thus we have the Anti-Semitism Industry targeted researchers in this way and most Jewish people think this is justified and genuine. They don't know that their 'Jewish' leaders and institutions of state, intelligence and military are not controlled by Jews at all, but cultists and stooges of Sabbatian-Frankism. I once added my name to a pro-Jewish freedom petition online and the next time I looked my name was gone and text had been added to the petition blurb to attack me as an 'anti-Semite' such is the scale of perceptual programming.

Moving on America

I tell the story in *The Trigger* and a chapter called 'Atlantic Crossing' how particularly after Israel was established the Sabbatians moved in on the United States and eventually grasped control of government administration, the political system via both Democrats and Republicans, the intelligence community like the CIA and National Security Agency (NSA), the Pentagon and mass media. Through this seriously compartmentalised network Sabbatians and their operatives in Mossad, Israeli Defense Forces (IDF) and US agencies pulled off 9/11 and blamed it on 19 'Al-Qaeda hijackers' dominated by men from, or connected to, Sabbatian-ruled Saudi

Arabia. The ‘19’ were not even on the planes let alone flew those big passenger jets into buildings while being largely incompetent at piloting one-engine light aircraft. ‘Hijacker’ Hani Hanjour who is said to have flown American Airlines Flight 77 into the Pentagon with a turn and manoeuvre most professional pilots said they would have struggled to do was banned from renting a small plane by instructors at the Freeway Airport in Bowie, Maryland, just *six weeks* earlier on the grounds that he was an incompetent pilot. The Jewish population of the world is just 0.2 percent with even that almost entirely concentrated in Israel (75 percent Jewish) and the United States (around two percent). This two percent and globally 0.2 percent refers to *Jewish* people and not Sabbatian interlopers who are a fraction of that fraction. What a sobering thought when you think of the fantastic influence on world affairs of tiny Israel and that the Project for the New America Century (PNAC) which laid out the blueprint in September, 2000, for America’s war on terror and regime change wars in Iraq, Libya and Syria was founded and dominated by Sabbatians known as ‘Neocons’. The document conceded that this plan would not be supported politically or publicly without a major attack on American soil and a Problem-Reaction-Solution excuse to send troops to war across the Middle East. Sabbatian Neocons said:

> … [The] process of transformation … [war and regime change] … is likely to be a long one, absent some catastrophic and catalysing event – like a new Pearl Harbor.

Four months later many of those who produced that document came to power with their inane puppet George Bush from the long-time Sabbatian Bush family. They included Sabbatian Dick Cheney who was officially vice-president, but really de-facto president for the entirety of the ‘Bush’ government. Nine months after the ‘Bush’ inauguration came what Bush called at the time ‘the Pearl Harbor of the 21st century’ and with typical Sabbatian timing and symbolism 2001 was the 60th anniversary of the attack in 1941 by the Japanese Air Force on Pearl Harbor, Hawaii, which allowed President Franklin Delano Roosevelt to take the United States into a Sabbatian-instigated Second World War that he said in his election campaign that he never would. The evidence is overwhelming that Roosevelt and his military and intelligence networks knew the attack was coming and did nothing to stop it, but they did make sure that America's most essential naval ships were not in Hawaii at the time. Three thousand Americans died in the Pearl Harbor attacks as they did on September 11th. By the 9/11 year of 2001 Sabbatians had widely infiltrated the US government, military and intelligence operations and used their compartmentalised assets to pull off the 'Al-Qaeda' attacks. If you read *The Trigger* it will blow your mind to see the utterly staggering concentration of 'Jewish' operatives (Sabbatian infiltrators) in essential positions of political, security, legal, law enforcement, financial and business power before, during, and after the attacks to make them happen, carry them out, and then cover their tracks – and I do mean *staggering* when you think of that 0.2 percent of the world population and two percent of Americans which are Jewish while Sabbatian infiltrators are a fraction of that. A central foundation of the 9/11 conspiracy was the hijacking of government, military, Air Force and intelligence computer systems in real time through 'back-door' access made possible by Israeli (Sabbatian) 'cyber security' software. Sabbatian-controlled Israel is on the way to rivalling Silicon Valley for domination of cyberspace and is becoming the dominant force in cyber-security which gives them access to entire computer systems and their passcodes across the world. Then add to this that Zionists head (officially) Silicon Valley giants like Google (Larry Page and Sergey Brin), Google-owned YouTube (Susan Wojcicki), Facebook (Mark Zuckerberg and Sheryl Sandberg), and Apple (Chairman Arthur D. Levinson), and that ultra-Zionist hedge fund billionaire Paul Singer has a $1 billion stake in Twitter which is only nominally headed by 'CEO' pothead Jack Dorsey. As cable news host Tucker Carlson said of Dorsey: 'There used to be debate in the medical community whether dropping a ton of acid had permanent effects and I think that debate has now ended.' Carlson made the comment after Dorsey told a hearing on Capitol Hill (if you cut through his bullshit) that he believed in free speech so long as he got to decide what you can hear and see. These 'big names' of Silicon Valley are only front men and women for the Global Cult, not least the Sabbatians, who are the true controllers of these corporations. Does anyone still wonder why these same people and companies have been ferociously censoring and banning people (like me) for exposing any aspect of the Cult agenda and especially the truth about the 'Covid' hoax which Sabbatians have orchestrated?

The Jeffrey Epstein paedophile ring was a Sabbatian operation. He was officially 'Jewish' but he was a Sabbatian and women abused by the ring have told me about the high number of 'Jewish' people involved. The Epstein horror has Sabbatian written all over it and matches perfectly their modus operandi and obsession with sex and ritual. Epstein was running a Sabbatian blackmail ring in which famous people with political and other influence were provided with young girls for sex while everything was being filmed and recorded on hidden cameras and microphones at his New York house, Caribbean island and other properties. Epstein survivors have described this surveillance system to me and some have gone public. Once the famous politician or other figure knew he or she was on video they tended to do whatever they were told. Here we go again …when you've got them by the balls their hearts and minds will follow. Sabbatians use this blackmail technique on a wide scale across the world to entrap politicians and others they need to act as demanded. Epstein's private plane, the infamous 'Lolita Express', had many well-known passengers including Bill Clinton while Bill Gates has flown on an Epstein plane and met with him four years after Epstein had been jailed for paedophilia. They subsequently met many times at Epstein's home in New York according to a witness who was there. Epstein's infamous side-kick was Ghislaine Maxwell, daughter of Mossad agent and ultra-Zionist mega-crooked British businessman, Bob Maxwell, who at one time owned the *Daily Mirror* newspaper. Maxwell was murdered at sea on his boat in 1991 by Sabbatian-controlled Mossad when he became a liability with his business empire collapsing as a former Mossad operative has confirmed (see *The Trigger*).

Money, money, money, funny money ...

Before I come to the Sabbatian connection with the last three US presidents I will lay out the crucial importance to Sabbatians of controlling banking and finance. Sabbatian Mayer Amschel Rothschild set out to dominate this arena in his family's quest for total global control. What is freedom? It is, in effect, choice. The more choices you have the freer you are and the fewer your choices the more you are enslaved. In the global structure created over centuries by Sabbatians the biggest decider and restrictor of choice is … money. Across the world if you ask people what they would like to do with their lives and why they are not doing that they will reply 'I don't have the money'. This is the idea. A global elite of multi-billionaires are described as 'greedy' and that is true on one level; but control of money – who has it and who doesn't – is not primarily about greed. It's about control. Sabbatians have seized ever more control of finance and sucked the wealth of the world out of the hands of the population. We talk now, after all, about the 'One-percent' and even then the wealthiest are a lot fewer even than that. This has been made possible by a money scam so outrageous and so vast it could rightly be called the scam of scams founded on creating 'money' out of nothing and 'loaning' that with interest to the population. Money out of nothing is called 'credit'. Sabbatians have asserted control over governments and banking ever more completely through the centuries and secured financial laws that allow banks to lend hugely more than they have on deposit in a confidence trick known as fractional reserve lending. Imagine if you could lend money that doesn't exist and charge the recipient interest for doing so. You would end up in jail. Bankers by contrast end up in mansions, private jets, Malibu and Monaco.

Banks are only required to keep a fraction of their deposits and wealth in their vaults and they are allowed to lend 'money' they don't have called 'credit. Go into a bank for a loan and if you succeed the banker will not move any real wealth into your account. They will type into your account the amount of the agreed ‘loan’ – say £100,000. This is not wealth that really exists; it is non-existent, fresh-air, created-out-of-nothing ‘credit’ which has never, does not, and will never exist except in theory. Credit is backed by nothing except wind and only has buying power because people think that it has buying power and accept it in return for property, goods and services. I have described this situation as like those cartoon characters you see chasing each other and when they run over the edge of a cliff they keep running forward on fresh air until one of them looks down, realises what’s happened, and they all crash into the ravine. The whole foundation of the Sabbatian financial system is to stop people looking down except for periodic moments when they want to crash the system (as in 2008 and 2020 ongoing) and reap the rewards from all the property, businesses and wealth their borrowers had signed over as ‘collateral’ in return for a ‘loan’ of fresh air. Most people think that money is somehow created by governments when it comes into existence from the start as a debt through banks ‘lending’ illusory money called credit. Yes, the very currency of exchange is a *debt* from day one issued as an interest-bearing loan. Why don’t governments create money interest-free and lend it to their people interest-free? Governments are controlled by Sabbatians and the financial system is controlled by Sabbatians for whom interest-free money would be a nightmare come true. Sabbatians underpin their financial domination through their global network of central banks, including the privately-owned US Federal Reserve and Britain’s Bank of England, and this is orchestrated by a privately-owned central bank coordination body called the Bank for International Settlements in Basle, Switzerland, created by the usual suspects including the Rockefellers and Rothschilds. Central bank chiefs don’t answer to governments or the people. They answer to the Bank for International Settlements or, in other words, the Global Cult which is dominated today by Sabbatians.

Built-in disaster

There are so many constituent scams within the overall banking scam. When you take out a loan of thin-air credit only the amount of that loan is theoretically brought into circulation to add to the amount in circulation; but you are paying back the principle plus interest. The additional interest is not created and this means that with every 'loan' there is a shortfall in the money in circulation between what is borrowed and what has to be paid back. There is never even close to enough money in circulation to repay all outstanding public and private debt including interest. Coldly weaved in the very fabric of the system is the certainty that some will lose their homes, businesses and possessions to the banking 'lender'. This is less obvious in times of 'boom' when the amount of money in circulation (and the debt) is expanding through more people wanting and getting loans. When a downturn comes and the money supply contracts it becomes painfully obvious that there is not enough money to service all debt and interest. This is less obvious in times of 'boom' when the amount of money in circulation (and the debt) is expanding through more people wanting and getting loans. When a downturn comes and the money supply contracts and it becomes painfully obvious – as in 2008 and currently – that there is not enough money to service all debt and interest. Sabbatian banksters have been leading the human population through a calculated series of booms (more debt incurred) and busts (when the debt can't be repaid and the banks get the debtor's tangible wealth in exchange for non-existent 'credit'). With each 'bust' Sabbatian bankers have absorbed more of the world's tangible wealth and we end up with the One-percent. Governments are in bankruptcy levels of debt to the same system and are therefore owned by a system they do not control. The Federal Reserve, 'America's central bank', is privately-owned and American presidents only nominally appoint its chairman or woman to maintain the illusion that it's an arm of government. It's not. The 'Fed' is a cartel of private banks which handed billions to its associates and friends after the crash of 2008 and has been Sabbatian-controlled since it was manipulated into being in 1913 through the covert trickery of Rothschild banking agents Jacob Schiff and Paul

Warburg, and the Sabbatian Rockefeller family. Somehow from a Jewish population of two-percent and globally 0.2 percent (Sabbatian interlopers remember are far smaller) ultra-Zionists headed the Federal Reserve for 31 years between 1987 and 2018 in the form of Alan Greenspan, Bernard Bernanke and Janet Yellen (now Biden's Treasury Secretary) with Yellen's deputy chairman a Israeli-American duel citizen and ultra-Zionist Stanley Fischer, a former governor of the Bank of Israel. Ultra-Zionist Fed chiefs spanned the presidencies of Ronald Reagan ('Republican'), Father George Bush ('Republican'), Bill Clinton ('Democrat'), Boy George Bush ('Republican') and Barack Obama ('Democrat'). We should really add the pre-Greenspan chairman, Paul Adolph Volcker, 'appointed' by Jimmy Carter ('Democrat') who ran the Fed between 1979 and 1987 during the Carter and Reagan administrations before Greenspan took over. Volcker was a long-time associate and business partner of the Rothschilds. No matter what the 'party' officially in power the United States economy was directed by the same force. Here are members of the Obama, Trump and Biden administrations and see if you can make out a common theme.

Barack Obama ('Democrat')

Ultra-Zionists Robert Rubin, Larry Summers, and Timothy Geithner ran the US Treasury in the Clinton administration and two of them reappeared with Obama. Ultra-Zionist Fed chairman Alan Greenspan had manipulated the crash of 2008 through deregulation and jumped ship just before the disaster to make way for ultra-Zionist Bernard Bernanke to hand out trillions to Sabbatian 'too big to fail' banks and businesses, including the ubiquitous ultra-Zionist Goldman Sachs which has an ongoing staff revolving door operation between itself and major financial positions in government worldwide. Obama inherited the fallout of the crash when he took office in January, 2009, and fortunately he had the support of his ultra-Zionist White House Chief of Staff Rahm Emmanuel, son of a terrorist who helped to bomb Israel into being in 1948, and his ultra-Zionist senior adviser David Axelrod, chief strategist in Obama's two successful presidential campaigns. Emmanuel, later mayor of Chicago and former senior fundraiser and strategist for Bill Clinton, is an example of the Sabbatian policy after Israel was established of migrating insider families to America so their children would be born American citizens. 'Obama' chose this financial team throughout his administration to respond to the Sabbatian-instigated crisis:

Timothy Geithner (ultra-Zionist) Treasury Secretary; Jacob J. Lew, Treasury Secretary; Larry Summers (ultra-Zionist), director of the White House National Economic Council; Paul Adolph Volcker (Rothschild business partner), chairman of the Economic Recovery Advisory Board; Peter Orszag (ultra-Zionist), director of the Office of Management and Budget overseeing all government spending; Penny Pritzker (ultra-Zionist), Commerce Secretary; Jared Bernstein (ultra-Zionist), chief economist and economic policy adviser to Vice President Joe Biden; Mary Schapiro (ultra-Zionist), chair of the Securities and Exchange Commission (SEC); Gary Gensler (ultra-Zionist), chairman of the Commodity Futures Trading Commission (CFTC); Sheila Bair (ultra-Zionist), chair of the Federal Deposit Insurance Corporation (FDIC); Karen Mills (ultra-Zionist), head of the Small Business Administration (SBA); Kenneth Feinberg (ultra-Zionist), Special Master for Executive [bail-out] Compensation. Feinberg would be appointed to oversee compensation (with strings) to 9/11 victims and families in a campaign to stop them having their day in court to question the official story. At the same time ultra-Zionist Bernard Bernanke was chairman of the Federal Reserve and these are only some of the ultra-Zionists with allegiance to Sabbatian-controlled Israel in the Obama government. Obama's biggest corporate donor was ultra-Zionist Goldman Sachs which had employed many in his administration.

Donald Trump ('Republican')

Trump claimed to be an outsider (he wasn't) who had come to 'drain the swamp'. He embarked on this goal by immediately appointing ultra-Zionist Steve Mnuchin, a Goldman Sachs employee for 17 years, as his Treasury Secretary. Others included Gary Cohn (ultra-Zionist), chief operating officer of Goldman Sachs, his first Director of the National Economic Council and chief economic adviser, who was later replaced by Larry Kudlow (ultra-Zionist). Trump's senior adviser throughout his four years in the White House was his sinister son-in-law Jared Kushner, a life-long friend of Israel Prime Minister Benjamin Netanyahu. Kushner is the son of a convicted crook who was pardoned by Trump in his last days in office. Other ultra-Zionists in the Trump administration included: Stephen Miller, Senior Policy Adviser; Avrahm Berkowitz, Deputy Adviser to Trump and his Senior Adviser Jared Kushner; Ivanka Trump, Adviser to the President, who converted to Judaism when she married Jared Kushner; David Friedman, Trump lawyer and Ambassador to Israel; Jason Greenblatt, Trump Organization executive vice president and chief legal officer, who was made Special Representative for International Negotiations and the Israeli-Palestinian Conflict; Rod Rosenstein, Deputy Attorney General; Elliot Abrams, Special Representative for Venezuela, then Iran; John Eisenberg, National Security Council Legal Adviser and Deputy Council to the President for National Security Affairs; Anne Neuberger, Deputy National Manager, National Security Agency; Ezra Cohen-Watnick, Acting Under Secretary of Defense for Intelligence; Elan Carr, Special Envoy to monitor and combat anti-Semitism; Len Khodorkovsky, Deputy Special Envoy to monitor and combat anti-Semitism; Reed Cordish, Assistant to the President, Intragovernmental and Technology Initiatives. Trump Vice President Mike Pence and Secretary of State Mike Pompeo, both Christian Zionists, were also vehement supporters of Israel and its goals and ambitions.

Donald 'free-speech believer' Trump pardoned a number of financial and violent criminals while ignoring calls to pardon Julian Assange and Edward Snowden whose crimes are revealing highly relevant information about government manipulation and corruption and the widespread illegal surveillance of the American people by US 'security' agencies. It's so good to know that Trump is on the side of freedom and justice and not mega-criminals with allegiance to Sabbatian-controlled Israel. These included a pardon for Israeli spy Jonathan Pollard who was jailed for life in 1987 under the Espionage Act. Aviem Sella, the Mossad agent who recruited Pollard, was also pardoned by Trump while Assange sat in jail and Snowden remained in exile in Russia. Sella had 'fled' (was helped to escape) to Israel in 1987 and was never extradited despite being charged under the Espionage Act. A Trump White House statement said that Sella's clemency had been 'supported by Benjamin Netanyahu, Ron Dermer, Israel's US Ambassador, David Friedman, US Ambassador to Israel and Miriam Adelson, wife of leading Trump donor Sheldon Adelson who died shortly before. Other friends of Jared Kushner were pardoned along with Sholom Weiss who was believed to be serving the longest-ever white-collar prison sentence of more than 800 years in 2000. The sentence was commuted of Ponzi-schemer Eliyahu Weinstein who defrauded Jews and others out of $200 million. I did mention that Assange and Snowden were ignored, right? Trump gave Sabbatians almost everything they asked for in military and political support, moving the US Embassy from Tel Aviv to Jerusalem with its critical symbolic and literal implications for Palestinian statehood, and the 'deal of the Century' designed by Jared Kushner and David Friedman which gave the Sabbatian Israeli government the green light to substantially expand its already widespread program of building illegal Jewish-only settlements in the occupied land of the West Bank. This made a two-state 'solution' impossible by seizing all the land of a potential Palestinian homeland and that had been the plan since 1948 and then 1967 when the Arab-controlled Gaza Strip, West Bank, Sinai Peninsula and Syrian Golan Heights were occupied by Israel. All the talks about talks and road maps and delays have been buying time until the West Bank was physically occupied by Israeli real estate. Trump would have to be a monumentally ill-informed idiot not to see that this was the plan he was helping to complete. The Trump administration was in so many ways the Kushner administration which means the Netanyahu administration which means the Sabbatian administration. I understand why many opposing Cult fascism in all its forms gravitated to Trump, but he was a crucial part of the Sabbatian plan and I will deal with this in the next chapter.

Joe Biden ('Democrat')

A barely cognitive Joe Biden took over the presidency in January, 2021, along with his fellow empty shell, Vice-President Kamala Harris, as the latest Sabbatian gofers to enter the White House. Names on the door may have changed and the 'party' – the force behind them remained the same as Zionists were appointed to a stream of pivotal areas relating to Sabbatian plans and policy. They included: Janet Yellen, Treasury Secretary, former head of the Federal Reserve, and still another ultra-Zionist running the US Treasury after Mnuchin (Trump), Lew and Geithner (Obama), and Summers and Rubin (Clinton); Anthony Blinken, Secretary of State; Wendy Sherman, Deputy Secretary of State (so that's 'Biden's' Sabbatian foreign policy sorted); Jeff Zients, White House coronavirus coordinator; Rochelle Walensky, head of the Centers for Disease Control; Rachel Levine, transgender deputy health secretary (that's 'Covid' hoax policy under control); Merrick Garland, Attorney General; Alejandro Mayorkas, Secretary of Homeland Security; Cass Sunstein, Homeland Security with responsibility for new immigration laws; Avril Haines, Director of National Intelligence; Anne Neuberger, National Security Agency cybersecurity director (note, cybersecurity); David Cohen, CIA Deputy Director; Ronald Klain, Biden's Chief of Staff (see Rahm Emanuel); Eric Lander, a 'leading geneticist', Office of Science and Technology Policy director (see Smart Grid, synthetic biology agenda); Jessica Rosenworcel, acting head of the Federal Communications Commission (FCC) which controls Smart Grid technology policy and electromagnetic communication systems including 5G. How can it be that so many pivotal positions are held by two-percent of the American population and 0.2 percent of the world population administration after administration no matter who is the president and what is the party? It's a coincidence? Of course it's not and this is why Sabbatians have built their colossal global web of interlocking 'anti-hate' hate groups to condemn anyone who asks these glaring questions as an 'anti-Semite'. The way that Jewish people horrifically abused in Sabbatian-backed Nazi Germany are exploited to this end is stomach-turning and disgusting beyond words.

Political fusion

Sabbatian manipulation has reversed the roles of Republicans and Democrats and the same has happened in Britain with the Conservative and Labour Parties. Republicans and Conservatives were always labelled the 'right' and Democrats and Labour the 'left', but look at the policy positions now and the Democrat-Labour 'left' has moved further to the 'right' than Republicans and Conservatives under the banner of 'Woke', the Cult-created far-right tyranny. Where once the Democrat-Labour 'left' defended free speech and human rights they now seek to delete them and as I said earlier despite the 'Covid' fascism of the Jackboot Johnson Conservative government in the UK the Labour Party of leader Keir Starmer demanded even more extreme measures. The Labour Party has been very publicly absorbed by Sabbatians after a political and media onslaught against the previous leader, the weak and inept Jeremy Corbyn, over made-up allegations of 'anti-Semitism' both by him and his party. The plan was clear with this 'anti-Semite' propaganda and what was required in response was a swift and decisive 'fuck off' from Corbyn and a statement to expose the Anti-Semitism Industry (Sabbatian) attempt to silence Labour criticism of the Israeli government (Sabbatians) and purge the party of all dissent against the extremes of ultra-Zionism (Sabbatians). Instead Corbyn and his party fell to their knees and appeased the abusers which, by definition, is impossible. Appeasing one demand leads only to a new demand to be appeased until takeover is complete. Like I say – 'fuck off' would have been a much more effective policy and I have used it myself with great effect over the years when Sabbatians are on my case which is most of the time. I consider that fact a great compliment, by the way. The outcome of the Labour Party capitulation is that we now have a Sabbatian-controlled

Conservative Party ‘opposed’ by a Sabbatian-controlled Labour Party in a one-party Sabbatian state that hurtles towards the extremes of tyranny (the Sabbatian cult agenda). In America the situation is the same. Labour’s Keir Starmer spends his days on his knees with his tongue out pointing to Tel Aviv, or I guess now Jerusalem, while Boris Johnson has an ‘anti-Semitism czar’ in the form of former Labour MP John Mann who keeps Starmer company on his prayer mat.

Sabbatian influence can be seen in Jewish members of the Labour Party who have been ejected for criticism of Israel including those from families that suffered in Nazi Germany. Sabbatians despise real Jewish people and target them even more harshly because it is so much more difficult to dub them ‘anti-Semitic’ although in their desperation they do try.

CHAPTER THREE

The Pushbacker sting

Until you realize how easy it is for your mind to be manipulated, you remain the puppet of someone else's game
Evita Ochel

I will use the presidencies of Trump and Biden to show how the manipulation of the one-party state plays out behind the illusion of political choice across the world. No two presidencies could – on the face of it – be more different and apparently at odds in terms of direction and policy.

A Renegade Mind sees beyond the obvious and focuses on outcomes and consequences and not image, words and waffle. The Cult embarked on a campaign to divide America between those who blindly support its agenda (the mentality known as 'Woke') and those who are pushing back on where the Cult and its Sabbatians want to go. This presents infinite possibilities for dividing and ruling the population by setting them at war with each other and allows a perceptual ring fence of demonisation to encircle the Pushbackers in a modern version of the Little Big Horn in 1876 when American cavalry led by Lieutenant Colonel George Custer were drawn into a trap, surrounded and killed by Native American tribes defending their land of thousands of years from being seized by the government. In this modern version the roles are reversed and it's those defending themselves from the Sabbatian government who are surrounded and the government that's seeking to destroy them. This trap was set years ago and to explain how we must return to 2016 and the emergence of Donald Trump as a candidate to be President of the United States. He set out to overcome the best part of 20 other candidates in the Republican Party before and during the primaries and was not considered by many in those early stages to have a prayer of living in the White House. The Republican Party was said to have great reservations about Trump and yet somehow he won the nomination. When you know how American politics works – politics in general – there is no way that Trump could have become the party's candidate unless the Sabbatian-controlled 'Neocons' that run the Republican Party wanted that to happen. We saw the proof in emails and documents made public by WikiLeaks that the Democratic Party hierarchy, or Democons, systematically undermined the campaign of Bernie Sanders to make sure that Sabbatian gofer Hillary Clinton won the nomination to be their presidential candidate. If the Democons could do that then the Neocons in the Republican Party could have derailed Trump in the same way. But they didn't and at that stage I began to conclude that Trump could well be the one chosen to be president. If that was the case the 'why' was pretty clear to see – the goal of dividing America between Cult agenda-supporting Wokers and Pushbackers who gravitated to Trump because he was telling them what they wanted to hear. His constituency of support had been increasingly ignored and voiceless for decades and profoundly through the eight years of Sabbatian puppet Barack Obama. Now here was someone speaking their language of pulling back from the incessant globalisation of political and economic power, the exporting of American jobs to China and elsewhere by 'American' (Sabbatian) corporations, the deletion of free speech, and the mass immigration policies that had further devastated job opportunities for the urban working class of all races and the once American heartlands of the Midwest.

Beware the forked tongue

Those people collectively sighed with relief that at last a political leader was apparently on their side, but another trait of the Renegade Mind is that you look even harder at people telling you what you want to hear than those who are telling you otherwise. Obviously as I said earlier people wish what they want to hear to be true and genuine and they are much more likely to believe that than someone saying what they don't want to here and don't want to be true. Sales people are taught to be skilled in eliciting by calculated questioning what their customers want to hear and repeating that back to them as their own opinion to get their targets to like and trust them. Assets of the Cult are also sales people in the sense of selling perception. To read Cult manipulation you have to play the long and expanded game and not fall for the Vaudeville show of party politics. Both American parties are vehicles for the Cult and they exploit them in different ways depending on what the agenda requires at that moment. Trump and the Republicans were used to be the focus of dividing America and isolating Pushbackers to open the way for a Biden presidency to become the most extreme in American history by advancing the full-blown Woke (Cult) agenda with the aim of destroying and silencing Pushbackers now labelled Nazi Trump supporters and white supremacists.

Sabbatians wanted Trump in office for the reasons described by ultra-Zionist Saul Alinsky (1909-1972) who was promoting the Woke philosophy through 'community organising' long before anyone had heard of it. In those days it still went by its traditional name of Marxism. The reason for the manipulated Trump phenomenon was laid out in Alinsky's 1971 book, *Rules for Radicals*, which was his blueprint for overthrowing democratic and other regimes and replacing them with Sabbatian Marxism. Not surprisingly his to-do list was evident in the Sabbatian French and Russian 'Revolutions' and that in China which will become very relevant in the next chapter about the 'Covid' hoax. Among Alinsky's followers have been the deeply corrupt Barack Obama, House Speaker Nancy Pelosi and Hillary Clinton who described him as a 'hero'. All three are Sabbatian stooges with Pelosi personifying the arrogant corrupt idiocy that so widely fronts up for the Cult inner core. Predictably as a Sabbatian advocate of the 'light-bringer' Alinsky features Lucifer on the dedication page of his book as the original radical who gained his own kingdom ('Earth' as we shall see). One of Alinsky's golden radical rules was to pick an individual and focus all attention, hatred and blame on them and not to target faceless bureaucracies and corporations. *Rules for Radicals* is really a Sabbatian handbook with its contents repeatedly employed all over the world for centuries and why wouldn't Sabbatians bring to power their designer-villain to be used as the individual on which all attention, hatred and blame was bestowed? This is what they did and the only question for me is how much Trump knew that and how much he was manipulated. A bit of both, I suspect. This was Alinsky's Trump technique from a man who died in 1972. The technique has spanned history:

> Pick the target, freeze it, personalize it, polarize it. Don't try to attack abstract corporations or bureaucracies. Identify a responsible individual. Ignore attempts to shift or spread the blame.

From the moment Trump came to illusory power everything was about him. It wasn't about Republican policy or opinion, but all about Trump. Everything he did was presented in negative, derogatory and abusive terms by the Sabbatian-dominated media led by Cult operations such as CNN, MSNBC, *The New York Times* and the Jeff Bezos-owned *Washington Post* – 'Pick the target, freeze it, personalize it, polarize it.' Trump was turned into a demon to be vilified by those who hated him and a demi-god loved by those who worshipped him. This, in turn, had his supporters, too, presented as equally demonic in preparation for the punchline later down the line when Biden was about to take office. It was here's a Trump, there's a Trump, everywhere a Trump, Trump. Virtually every news story or happening was filtered through the lens of 'The Donald'. You loved him or hated him and which one you chose was said to define you as Satan's spawn or a paragon of virtue. Even supporting some Trump policies or statements and not others was enough for an assault on your character. No shades of grey were or are allowed. Everything is black and white (literally and figuratively). A Californian I knew had her head utterly scrambled by her hatred for Trump while telling people they should love each other. She was so totally consumed by

Trump Derangement Syndrome as it became to be known that this glaring contradiction would never have occurred to her. By definition anyone who criticised Trump or praised his opponents was a hero and this lady described Joe Biden as 'a kind, honest gentleman' when he's a provable liar, mega-crook and vicious piece of work to boot. Sabbatians had indeed divided America using Trump as the fall-guy and all along the clock was ticking on the consequences for his supporters.

In hock to his masters

Trump gave Sabbatians via Israel almost everything they wanted in his four years. Ask and you shall receive was the dynamic between himself and Benjamin Netanyahu orchestrated by Trump's ultra-Zionist son-in-law Jared Kushner, his ultra-Zionist Ambassador to Israel, David Friedman, and ultra-Zionist 'Israel adviser', Jason Greenblatt. The last two were central to the running and protecting from collapse of his business empire, the Trump Organisation, and colossal business failures made him forever beholding to Sabbatian networks that bailed him out. By the start of the 1990s Trump owed $4 billion to banks that he couldn't pay and almost $1billion of that was down to him personally and not his companies. This mega-disaster was the result of building two new casinos in Atlantic City and buying the enormous Taj Mahal operation which led to crippling debt payments. He had borrowed fantastic sums from 72 banks with major Sabbatian connections and although the scale of debt should have had him living in a tent alongside the highway they never foreclosed. A plan was devised to lift Trump from the mire by BT Securities Corporation and Rothschild Inc. and the case was handled by Wilber Ross who had worked for the Rothschilds for 27 years. Ross would be named US Commerce Secretary after Trump's election. Another crucial figure in saving Trump was ultra-Zionist 'investor' Carl Icahn who bought the Taj Mahal casino. Icahn was made special economic adviser on financial regulation in the Trump administration. He didn't stay long but still managed to find time to make a tidy sum of a reported $31.3 million when he sold his holdings affected by the price of steel three days before Trump imposed a 235 percent tariff on steel imports. What amazing bits of luck these people have. Trump and Sabbatian operatives have long had a close association and his mentor and legal adviser from the early 1970s until 1986 was the dark and genetically corrupt ultra-Zionist Roy Cohn who was chief counsel to Senator Joseph McCarthy's 'communist' witch-hunt in the 1950s. *Esquire* magazine published an article about Cohn with the headline 'Don't mess with Roy Cohn'. He was described as the most feared lawyer in New York and 'a ruthless master of dirty tricks ... [with] ... more than one Mafia Don on speed dial'. Cohn's influence, contacts, support and protection made Trump a front man for Sabbatians in New York with their connections to one of Cohn's many criminal employers, the 'Russian' Sabbatian Mafia. Israel-centric media mogul Rupert Murdoch was introduced to Trump by Cohn and they started a long friendship. Cohn died in 1986 weeks after being disbarred for unethical conduct by the Appellate Division of the New York State Supreme Court. The wheels of justice do indeed run slow given the length of Cohn's crooked career.

QAnon-sense

We are asked to believe that Donald Trump with his fundamental connections to Sabbatian networks and operatives has been leading the fight to stop the Sabbatian agenda for the fascistic control of America and the world. Sure he has. A man entrapped during his years in the White House by Sabbatian operatives and whose biggest financial donor was casino billionaire Sheldon Adelson who was Sabbatian to his DNA?? Oh, do come on. Trump has been used to divide America and isolate Pushbackers on the Cult agenda under the heading of 'Trump supporters', 'insurrectionists' and 'white supremacists'. The US Intelligence/Mossad Psyop or psychological operation known as QAnon emerged during the Trump years as a central pillar in the Sabbatian campaign to lead Pushbackers into the trap set by those that wished to destroy them. I knew from the start that QAnon was a scam because I had seen the same scenario many times before over 30 years under different names and I had written about one in particular in the books. ‘Not again’ was my reaction when QAnon came to the fore. The same script is pulled out every few years and a new name added to the letterhead. The story always takes the same form: ‘Insiders’ or ‘the good guys’ in the government-intelligence-military ‘Deep State’ apparatus were going to instigate mass arrests of the ‘bad guys’ which would include the Rockefellers, Rothschilds, Barack Obama, Hillary Clinton, George Soros, etc., etc. Dates are given for when the ‘good guys’ are going to move in, but the dates pass without incident and new dates are given which pass without incident. The central message to Pushbackers in each case is that they don’t have to do anything because there is ‘a plan’ and it is all going to be sorted by the ‘good guys’ on the inside. ‘Trust the plan’ was a QAnon mantra when the only plan was to misdirect Pushbackers into putting their trust in a Psyop they believed to be real. Beware, beware, those who tell you what you want to hear and always check it out. Right up to Biden’s inauguration QAnon was still claiming that ‘the Storm’ was coming and Trump would stay on as president when Biden and his cronies were arrested and jailed. It was never going to happen and of course it didn’t, but what did happen as a result provided that punchline to the Sabbatian Trump/QAnon Psyop.

On January 6th, 2021, a very big crowd of Trump supporters gathered in the National Mall in Washington DC down from the Capitol Building to protest at what they believed to be widespread corruption and vote fraud that stopped Trump being re-elected for a second term as president in November, 2020. I say as someone that does not support Trump or Biden that the evidence is clear that major vote-fixing went on to favour Biden, a man with cognitive problems so advanced he can often hardly string a sentence together without reading the words written for him on the Teleprompter. Glaring ballot discrepancies included serious questions about electronic voting machines that make vote rigging a comparative cinch and hundreds of thousands of paper votes that suddenly appeared during already advanced vote counts and virtually all of them for Biden. Early Trump leads in crucial swing states suddenly began to close and disappear. The pandemic hoax was used as the excuse to issue almost limitless numbers of mail-in ballots with no checks to establish that the recipients were still alive or lived at that address. They were sent to streams of people who had not even asked for them. Private organisations were employed to gather these ballots and who knows what they did with them before they turned up at the counts. The American election system has been manipulated over decades to become a sick joke with more holes than a Swiss cheese for the express purpose of dictating the results. Then there was the criminal manipulation of information by Sabbatian tech giants like Facebook, Twitter and Google-owned YouTube which deleted pro-Trump, anti-Biden accounts and posts while everything in support of Biden was left alone. Sabbatians wanted Biden to win because after the dividing of America it was time for full-on Woke and every aspect of the Cult agenda to be unleashed.

Hunter gatherer

Extreme Silicon Valley bias included blocking information by the *New York Post* exposing a Biden scandal that should have ended his bid for president in the final weeks of the campaign. Hunter Biden, his monumentally corrupt son, is reported to have sent a laptop to be repaired at a local store and failed to return for it. Time passed until the laptop became the property of the store for non-payment of the bill. When the owner saw what was on the hard drive he gave a copy to the FBI who did nothing even though it confirmed widespread corruption in which the Joe Biden family were using his political position, especially when he was vice president to Obama, to make multiple millions in countries around the world and most notably Ukraine and China. Hunter Biden's one-time business partner Tony Bobulinski went public when the story broke in the *New York Post* to confirm the corruption he saw and that Joe Biden not only knew what was going on he also profited from the spoils. Millions were handed over by a Chinese company with close connections – like all major businesses in China – to the Chinese communist party of President Xi Jinping. Joe Biden even boasted at a meeting of the Cult's World Economic Forum that as vice president he had ordered the government of Ukraine to fire a prosecutor. What he didn't mention was that the same man just happened to be investigating an energy company which was part of Hunter Biden's corrupt portfolio. The company was paying him big bucks for no other reason than the influence his father had. Overnight Biden's presidential campaign should have been over given that he had lied publicly about not knowing what his son was doing. Instead almost the entire Sabbatian-owned mainstream media and Sabbatian-owned Silicon Valley suppressed circulation of the story. This alone went a mighty way to rigging the election of 2020. Cult assets like Mark Zuckerberg at Facebook also spent hundreds of millions to be used in support of Biden and vote 'administration'.

The Cult had used Trump as the focus to divide America and was now desperate to bring in moronic, pliable, corrupt Biden to complete the double-whammy. No way were they going to let little things like the will of the people thwart their plan. Silicon Valley widely censored claims that the election was rigged because it *was* rigged. For the same reason anyone claiming it was rigged was denounced as a 'white supremacist' including the pathetically few Republican politicians willing to say so. Right across the media where the claim was mentioned it was described as a 'false claim' even though these excuses for 'journalists' would have done no research into the subject whatsoever. Trump won seven million more votes than any sitting president had ever achieved while somehow a cognitively-challenged soon to be 78-year-old who was hidden away from the public for most of the campaign managed to win more votes than any presidential candidate in history. It makes no sense. You only had to see election rallies for both candidates to witness the enthusiasm for Trump and the apathy for Biden. Tens of thousands would attend Trump events while Biden was speaking in empty car parks with often only television crews attending and framing their shots to hide the fact that no one was there. It was pathetic to see footage come to light of Biden standing at a podium making speeches only to TV crews and party fixers while reading the words written for him on massive Teleprompter screens. So, yes, those protestors on January 6th had a point about election rigging, but some were about to walk into a trap laid for them in Washington by the Cult Deep State and its QAnon Psyop. This was the Capitol Hill riot ludicrously dubbed an 'insurrection'.

The spider and the fly

Renegade Minds know there are not two 'sides' in politics, only one side, the Cult, working through all 'sides'. It's a stage show, a puppet show, to direct the perceptions of the population into focusing on diversions like parties and candidates while missing the puppeteers with their hands holding all the strings. The Capitol Hill 'insurrection' brings us back to the Little Big Horn. Having created two distinct opposing groupings – Woke and Pushbackers – the trap was about to be sprung. Pushbackers were to be encircled and isolated by associating them all in the public mind with Trump and then labelling Trump as some sort of Confederate leader. I knew immediately that the Capitol riot was a set-up because of two things. One was how easy the rioters got into the building with virtually no credible resistance and secondly I could see – as with the 'Covid' hoax in the West at the start of 2020 – how the Cult could exploit the situation to move its agenda forward with great speed. My experience of Cult techniques and activities over more than 30 years has showed me that while they do exploit situations they haven't themselves created this never happens with events of fundamental agenda significance. Every time major events giving cultists the excuse to rapidly advance their plan you find they are manipulated into being for the specific reason of providing that excuse – Problem-Reaction-Solution. Only a tiny minority of the huge crowd of Washington protestors sought to gain entry to the Capitol by smashing windows and breaching doors. That didn't matter. The whole crowd and all Pushbackers, even if they did not support Trump, were going to be lumped together as dangerous insurrectionists and conspiracy theorists. The latter term came into widespread use through a CIA memo in the 1960s aimed at discrediting those questioning the nonsensical official story of the Kennedy assassination and it subsequently became widely employed by the media. It's still being used by inept 'journalists' with no idea of its origin to discredit anyone questioning anything that authority claims to be true. When you are perpetrating a conspiracy you need to discredit the very word itself even though the dictionary definition of conspiracy is merely 'the activity of secretly planning with other people to do something bad or illegal' and 'a general agreement to keep silent about a subject for the purpose of keeping it secret'. On that basis there are conspiracies almost wherever you look. For obvious reasons the Cult and its lapdog media have to claim there are no conspiracies even though the word appears in state laws as with conspiracy to defraud, to murder, and to corrupt public morals.

Agent provocateurs are widely used by the Cult Deep State to manipulate genuine people into acting in ways that suit the desired outcome. By genuine in this case I mean protestors genuinely supporting Trump and claims that the election was stolen. In among them, however, were agents of the state wearing the garb of Trump supporters and QAnon to pump-prime the Capital riot which some genuine Trump supporters naively fell for. I described the situation as 'Come into my parlour said the spider to the fly'. Leaflets appeared through the Woke paramilitary arm Antifa, the anti-fascist fascists, calling on supporters to turn up in Washington looking like Trump supporters even though they hated him. Some of those arrested for breaching the Capitol Building were sourced to Antifa and its stable mate Black Lives Matter. Both organisations are funded by Cult billionaires and corporations. One man charged for the riot was according to his lawyer a former FBI agent who had held top secret security clearance for 40 years. Attorney Thomas Plofchan said of his client, 66-year-old Thomas Edward Caldwell:

> He has held a Top Secret Security Clearance since 1979 and has undergone multiple Special Background Investigations in support of his clearances. After retiring from the Navy, he

> worked as a section chief for the Federal Bureau of Investigation from 2009-2010 as a GS-12 [mid-level employee].
>
> He also formed and operated a consulting firm performing work, often classified, for U.S government customers including the US. Drug Enforcement Agency, Department of Housing and Urban Development, the US Coast Guard, and the US Army Personnel Command.

A judge later released Caldwell pending trial in the absence of evidence about a conspiracy or that he tried to force his way into the building. *The New York Post* reported a 'law enforcement source' as saying that 'at least two known Antifa members were spotted' on camera among Trump supporters during the riot while one of the rioters arrested was John Earle Sullivan, a seriously extreme Black Lives Matter Trump-hater from Utah who was previously arrested and charged in July, 2020, over a BLM-Antifa riot in which drivers were threatened and one was shot. Sullivan is the founder of Utah-based Insurgence USA which is an affiliate of the Cult-created-and-funded Black Lives Matter movement. Footage appeared and was then deleted by Twitter of Trump supporters calling out Antifa infiltrators and a group was filmed changing into pro-Trump clothing before the riot. Security at the building was *pathetic* – as planned. Colonel Leroy Fletcher Prouty, a man with long experience in covert operations working with the US security apparatus, once described the tell-tale sign to identify who is involved in an assassination. He said:

> No one has to direct an assassination – it happens. The active role is played secretly by permitting it to happen. This is the greatest single clue. Who has the power to call off or reduce the usual security precautions?

This principle applies to many other situations and certainly to the Capitol riot of January 6th, 2021.

The sting

With such a big and potentially angry crowd known to be gathering near the Capitol the security apparatus would have had a major police detail to defend the building with National Guard troops on standby given the strength of feeling among people arriving from all over America encouraged by the QAnon Psyop and statements by Donald Trump. Instead Capitol Police 'security' was flimsy, weak, and easily breached. The same number of officers was deployed as on a regular day and that is a blatant red flag. They were not staffed or equipped for a possible riot that had been an obvious possibility in the circumstances. No protective and effective fencing worth the name was put in place and there were no contingency plans. The whole thing was basically a case of standing aside and waving people in. Once inside police mostly backed off apart from one Capitol police officer who ridiculously shot dead unarmed Air Force veteran protestor Ashli Babbitt without a warning as she climbed through a broken window. The 'investigation' refused to name or charge the officer after what must surely be considered a murder in the circumstances. They just lifted a carpet and swept. The story was endlessly repeated about five people dying in the 'armed insurrection' when there was no report of rioters using weapons. Apart from Babbitt the other four died from a heart attack, strokes and apparently a drug overdose. Capitol police officer Brian Sicknick was reported to have died after being bludgeoned with a fire extinguisher when he was alive after the riot was over and died later of what the Washington Medical Examiner's Office said was a stroke. Sicknick had no external injuries. The lies were delivered like rapid fire. There was a narrative to build with incessant repetition of the lie until the lie became the accepted 'everybody knows that' truth. The 'Big Lie' technique of Nazi Propaganda Minister Joseph Goebbels is constantly used by the Cult which was behind the Nazis and is today behind the 'Covid' and 'climate change' hoaxes. Goebbels said:

> If you tell a lie big enough and keep repeating it, people will eventually come to believe it. The lie can be maintained only for such time as the State can shield the people from the political, economic and/or military consequences of the lie. It thus becomes vitally important for the State to use all of its powers to repress dissent, for the truth is the mortal enemy of the lie, and thus by extension, the truth is the greatest enemy of the State.

Most protestors had a free run of the Capitol Building. This allowed pictures to be taken of rioters in iconic parts of the building including the Senate chamber which could be used as propaganda images against all Pushbackers. One Congresswoman described the scene as 'the worst kind of non-security anybody could ever imagine'. Well, the first part was true, but someone obviously did imagine it and made sure it happened. Some photographs most widely circulated featured people wearing QAnon symbols and now the Psyop would be used to dub all QAnon followers with the ubiquitous fit-all label of 'white supremacist' and 'insurrectionists'. When a Muslim extremist called Noah Green drove his car at two police officers at the Capitol Building killing one in April, 2021, there was no such political and media hysteria. They were just disappointed he wasn't white.

The witch-hunt

Government prosecutor Michael Sherwin, an aggressive, dark-eyed, professional Rottweiler led the 'investigation' and to call it over the top would be to understate reality a thousand fold. Hundreds were tracked down and arrested for the crime of having the wrong political views and people were jailed who had done nothing more than walk in the building, committed no violence or damage to property, took a few pictures and left. They were labelled a 'threat to the Republic' while Biden sat in the White House signing executive orders written for him that were dismantling 'the Republic'. Even when judges ruled that a mother and son should not be in jail the government kept them there. Some of those arrested have been badly beaten by prison guards in Washington and lawyers for one man said he suffered a fractured skull and was made blind in one eye. Meanwhile a woman is shot dead for no reason by a Capitol Police officer and we are not allowed to know who he is never mind what has happened to him although that will be *nothing*. The Cult's QAnon/Trump sting to identify and isolate Pushbackers and then target them on the road to crushing and deleting them was a resounding success. You would have thought the Russians had invaded the building at gunpoint and lined up senators for a firing squad to see the political and media reaction. Congresswoman Alexandria Ocasio-Cortez is a child in a woman's body, a terrible-twos, me, me, me, Woker narcissist of such proportions that words have no meaning. She said she thought she was going to die when 'insurrectionists' banged on her office door. It turned out she wasn't even in the Capitol Building when the riot was happening and the 'banging' was a Capitol Police officer. She referred to herself as a 'survivor' which is an insult to all those true survivors of violent and sexual abuse while she lives her pampered and privileged life talking drivel for a living. Her Woke colleague and fellow mega-narcissist Rashida Tlaib broke down describing the devastating effect on her, too, of *not being* in the building when the rioters were there. Ocasio-Cortez and Tlaib are members of a fully-Woke group of Congresswomen known as 'The Squad' along with Ilhan Omar and Ayanna Pressley. The Squad from what I can see can be identified by its vehement anti-white racism, anti-white men agenda, and, as always in these cases, the absence of brain cells on active duty.

The usual suspects were on the riot case immediately in the form of Democrat ultra-Zionist senators and operatives Chuck Schumer and Adam Schiff demanding that Trump be impeached for 'his part in the insurrection'. The same pair of prats had led the failed impeachment of Trump over the invented 'Russia collusion' nonsense which claimed Russia had helped Trump win the 2016 election. I didn't realise that Tel Aviv had been relocated just outside Moscow. I must find an up-to-date map. The Russia hoax was a Sabbatian operation to keep Trump occupied and impotent and to stop any rapport with Russia which the Cult wants to retain as a perceptual enemy to be pulled out at will. Puppet Biden began attacking Russia when he came to office as the Cult seeks more upheaval, division and war across the world. A two-year stage show 'Russia collusion inquiry' headed by the not-very-bright former 9/11 FBI chief Robert Mueller, with support from 19 lawyers, 40 FBI agents plus intelligence analysts, forensic accountants and other staff, devoured tens of millions of dollars and found no evidence of Russia collusion which a ten-year-old could have told them on day one. Now the same moronic Schumer and Schiff wanted a second impeachment of Trump over the Capitol 'insurrection' (riot) which the arrested development of Schumer called another 'Pearl Harbor' while others compared it with 9/11 in which 3,000 died and, in the case of CNN, with the Rwandan genocide in the 1990s in which an estimated 500,000 to 600,000 were murdered, between 250, 000 and 500,000 women were raped, and populations of whole towns were hacked to death with machetes. To make those comparisons purely for Cult political reasons is beyond insulting to those that suffered and lost their lives and confirms yet again the callous inhumanity that we are dealing with. Schumer is a monumental idiot and so is Schiff, but they serve the Cult agenda and do whatever they're told so they get looked after. Talking of idiots – another inane man who spanned the Russia and Capitol impeachment attempts was Senator Eric Swalwell who had the nerve to accuse Trump of collusion with the Russians while sleeping with a Chinese spy called Christine Fang or 'Fang Fang' which is straight out of a Bond film no doubt starring Klaus Schwab as the bloke living on a secret island and controlling laser weapons positioned in space and pointing at world capitals. Fang Fang plays the part of Bond's infiltrator girlfriend which I'm sure she would enjoy rather more than sharing a bed with the brainless Swalwell, lying back and thinking of China. The FBI eventually warned Swalwell about Fang Fang which gave her time to escape back to the Chinese dictatorship. How very thoughtful of them. The second Trump impeachment also failed and hardly surprising when an impeachment is supposed to remove a sitting president and by the time it happened Trump was no longer president. These people are running your country America, well, officially anyway. Terrifying isn't it?

Outcomes tell the story - always

The outcome of all this – and it's the *outcome* on which Renegade Minds focus, not the words – was that a vicious, hysterical and obviously pre-planned assault was launched on Pushbackers to censor, silence and discredit them and even targeted their right to earn a living. They have since been condemned as 'domestic terrorists' that need to be treated like Al-Qaeda and Islamic State. 'Domestic terrorists' is a label the Cult has been trying to make stick since the period of the Oklahoma bombing in 1995 which was blamed on 'far-right domestic terrorists'. If you read *The Trigger* you will see that the bombing was clearly a Problem-Reaction-Solution carried out by the Deep State during a Bill Clinton administration so corrupt that no dictionary definition of the term would even nearly suffice. Nearly 30, 000 troops were deployed from all over America to the empty streets of Washington for Biden's inauguration. Ten thousand of them stayed on with the pretext of protecting the capital from insurrectionists when it was more psychological programming to normalise the use of the military in domestic law enforcement in support of the Cult plan for a police-military state. Biden's fascist administration began a purge of 'wrong-thinkers' in the military which means anyone that is not on board with Woke. The Capitol Building was surrounded by a fence with razor wire and the Land of the Free was further symbolically and literally dismantled. The circle was completed with the installation of Biden and the exploitation of the QAnon Psyop.

America had never been so divided since the civil war of the 19th century, Pushbackers were isolated and dubbed terrorists and now, as was always going to happen, the Cult immediately set about deleting what little was left of freedom and transforming American society through a swish of the hand of the most controlled 'president' in American history leading (officially at least) the most extreme regime since the country was declared an independent state on July 4th, 1776. Biden issued undebated, dictatorial executive orders almost by the hour in his opening days in office across the whole spectrum of the Cult wish-list including diluting controls on the border with Mexico allowing thousands of migrants to illegally enter the United States to transform the demographics of America and import an election-changing number of perceived Democrat voters. Then there were Biden deportation amnesties for the already illegally resident (estimated to be as high as 20 or even 30 million). A bill before Congress awarded American citizenship to anyone who could prove they had worked in agriculture for just 180 days in the previous two years as 'Big Ag' secured its slave labour long-term. There were the plans to add new states to the union such as Puerto Rico and making Washington DC a state. They are all parts of a plan to ensure that the Cult-owned Woke Democrats would be permanently in power.

Border – what border?

I have exposed in detail in other books how mass immigration into the United States and Europe is the work of Cult networks fuelled by the tens of billions spent to this and other ends by George Soros and his global Open Society (open borders) Foundations. The impact can be seen in America alone where the population has increased by *100 million* in little more than 30 years mostly through immigration. I wrote in *The Answer* that the plan was to have so many people crossing the southern border that the numbers become unstoppable and we are now there under Cult-owned Biden. El Salvador in Central America puts the scale of what is happening into context. A third of the population now lives in the United States, much of it illegally, and many more are on the way. The methodology is to crush Central and South American countries economically and spread violence through machete-wielding psychopathic gangs like MS-13 based in El Salvador and now operating in many American cities. Biden-imposed lax security at the southern border means that it is all but open. He said before his 'election' that he wanted to see a surge towards the border if he became president and that was the green light for people to do just that after election day to create the human disaster that followed for both America and the migrants. When that surge came the imbecilic Alexandria Ocasio-Cortez said it wasn't a 'surge' because they are 'children, not insurgents' and the term 'surge' (used by Biden) was a claim of 'white supremacists'.

This disingenuous lady may one day enter the realm of the most basic intelligence, but it won't be any time soon.

Sabbatians and the Cult are in the process of destroying America by importing violent people and gangs in among the genuine to terrorise American cities and by overwhelming services that cannot cope with the sheer volume of new arrivals. Something similar is happening in Europe as Western society in general is targeted for demographic and cultural transformation and upheaval. The plan demands violence and crime to create an environment of intimidation, fear and division and Soros has been funding the election of district attorneys across America who then stop prosecuting many crimes, reduce sentences for violent crimes and free as many violent criminals as they can. Sabbatians are creating the chaos from which order – their order – can respond in a classic Problem-Reaction-Solution. A Freemasonic moto says 'Ordo Ab Chao' (Order out of Chaos) and this is why the Cult is constantly creating chaos to impose a new 'order'. Here you have the reason the Cult is constantly creating chaos. The 'Covid' hoax can be seen with those entering the United States by plane being forced to take a 'Covid' test while migrants flooding through southern border processing facilities do not. Nothing is put in the way of mass migration and if that means ignoring the government's own 'Covid' rules then so be it. They know it's all bullshit anyway. Any pushback on this is denounced as 'racist' by Wokers and Sabbatian fronts like the ultra-Zionist Anti-Defamation League headed by the appalling Jonathan Greenblatt which at the same time argues that Israel should not give citizenship and voting rights to more Palestinian Arabs or the 'Jewish population' (in truth the Sabbatian network) will lose control of the country.

Society-changing numbers

Biden's masters have declared that countries like El Salvador are so dangerous that their people must be allowed into the United States for humanitarian reasons when there are fewer murders in large parts of many Central American countries than in US cities like

Baltimore. That is not to say Central America cannot be a dangerous place and Cult-controlled American governments have been making it so since way back, along with the dismantling of economies, in a long-term plan to drive people north into the United States. Parts of Central America are very dangerous, but in other areas the story is being greatly exaggerated to justify relaxing immigration criteria. Migrants are being offered free healthcare and education in the United States as another incentive to head for the border and there is no requirement to be financially independent before you can enter to prevent the resources of America being drained. You can't blame migrants for seeking what they believe will be a better life, but they are being played by the Cult for dark and nefarious ends. The numbers since Biden took office are huge. In February, 2021, more than 100,000 people were known to have tried to enter the US illegally through the southern border (it was 34,000 in the same month in 2020) and in March it was 170,000 – a 418 percent increase on March, 2020. These numbers are only known people, not the ones who get in unseen. The true figure for migrants illegally crossing the border in a single month was estimated by one congressman at 250,000 and that number will only rise under Biden's current policy. Gangs of murdering drug-running thugs that control the Mexican side of the border demand money – thousands of dollars – to let migrants cross the Rio Grande into America. At the same time gun battles are breaking out on the border several times a week between rival Mexican drug gangs (which now operate globally) who are equipped with sophisticated military-grade weapons, grenades and armoured vehicles. While the Capitol Building was being 'protected' from a non-existent 'threat' by thousands of troops, and others were still deployed at the time in the Cult Neocon war in Afghanistan, the southern border of America was left to its fate. This is not incompetence, it is cold calculation.

By March, 2021, there were 17,000 unaccompanied children held at border facilities and many of them are ensnared by people traffickers for paedophile rings and raped on their journey north to America. This is not conjecture – this is fact. Many of those designated children are in reality teenage boys or older. Meanwhile Wokers posture their self-purity for encouraging poor and tragic people to come to America and face this nightmare both on the journey and at the border with the disgusting figure of House Speaker Nancy Pelosi giving disingenuous speeches about caring for migrants. The woman's evil. Wokers condemned Trump for having children in cages at the border (so did Obama, *Shhhh*), but now they are sleeping on the floor without access to a shower with one border facility 729 percent over capacity. The Biden insanity even proposed flying migrants from the southern border to the northern border with Canada for 'processing'. The whole shambles is being overseen by ultra-Zionist Secretary of Homeland Security, the moronic liar Alejandro Mayorkas, who banned news cameras at border facilities to stop Americans seeing what was happening. Mayorkas said there was not a ban on news crews; it was just that they were not allowed to film. Alongside him at Homeland Security is another ultra-Zionist Cass Sunstein appointed by Biden to oversee new immigration laws. Sunstein despises conspiracy researchers to the point where he suggests they should be banned or *taxed* for having such views. The man is not bonkers or anything. He's perfectly well-adjusted, but adjusted to what is the question. Criticise what is happening and you are a 'white supremacist' when earlier non-white immigrants also oppose the numbers which effect their lives and opportunities. Black people in poor areas are particularly damaged by uncontrolled immigration and the increased competition for work opportunities with those who will work for less. They are also losing voting power as Hispanics become more dominant in former black areas. It's a downward spiral for them while the billionaires behind the policy drone on about how much they care about black people and 'racism'. None of this is about compassion for migrants or black people – that's just wind and air. Migrants are instead being mercilessly exploited to transform America while the countries they leave are losing their future and the same is true in Europe. Mass immigration may now be the work of Woke Democrats, but it can be traced back to the 1986 Immigration Reform and Control Act (it wasn't) signed into law by Republican hero President Ronald Reagan which gave amnesty to millions living in the United States illegally and other incentives for people to head for the southern border. Here we have the one-party state at work again.

Save me syndrome

Almost every aspect of what I have been exposing as the Cult agenda was on display in even the first days of 'Biden' with silencing of Pushbackers at the forefront of everything. A Renegade Mind will view the Trump years and QAnon in a very different light to their supporters and advocates as the dots are connected. The QAnon/Trump Psyop has given the Cult all it was looking for. We may not know how much, or little, that Trump realised he was being used, but that's a side issue. This pincer movement produced the desired outcome of dividing America and having Pushbackers isolated. To turn this around we have to look at new routes to empowerment which do not include handing our power to other people and groups through what I will call the 'Save Me Syndrome' – 'I want someone else to do it so that I don't have to'. We have seen this at work throughout human history and the QAnon/Trump Psyop is only the latest incarnation alongside all the others. Religion is an obvious expression of this when people look to a 'god' or priest to save them or tell them how to be saved and then there are 'save me' politicians like Trump. Politics is a diversion and not a 'saviour'. It is a means to block positive change, not make it possible.

Save Me Syndrome always comes with the same repeating theme of handing your power to whom or what you believe will save you while your real 'saviour' stares back from the mirror every morning. Renegade Minds are constantly vigilant in this regard and always asking the question 'What can I do?' rather than 'What can someone else do for me?' Gandhi was right when he said: 'You must be the change you want to see in the world.' We are indeed the people we have been waiting for. We are presented with a constant raft of reasons to concede that power to others and forget where the real power is. Humanity has the numbers and the Cult does not. It has to use diversion and division to target the unstoppable power that comes from unity. Religions, governments, politicians, corporations, media, QAnon, are all different manifestations of this power-diversion and dilution. Refusing to give your power to governments and instead handing it to Trump and QAnon is not to take a new direction, but merely to recycle the old one with new names on the posters. I will explore this phenomenon as we proceed and how to break the cycles and recycles that got us here through the mists of repeating perception and so repeating history.

For now we shall turn to the most potent example in the entire human story of the consequences that follow when you give your power away. I am talking, of course, of the ‘Covid’ hoax.

CHAPTER FOUR

'Covid': Calculated catastrophe

Facts are threatening to those invested in fraud
DaShanne Stokes

We can easily unravel the real reason for the 'Covid pandemic' hoax by employing the Renegade Mind methodology that I have outlined this far. We'll start by comparing the long-planned Cult outcome with the 'Covid pandemic' outcome. Know the outcome and you'll see the journey.

I have highlighted the plan for the Hunger Games Society which has been in my books for so many years with the very few controlling the very many through ongoing dependency. To create this dependency it is essential to destroy independent livelihoods, businesses and employment to make the population reliant on the state (the Cult) for even the basics of life through a guaranteed pittance income. While independence of income remained these Cult ambitions would be thwarted. With this knowledge it was easy to see where the 'pandemic' hoax was going once talk of 'lockdowns' began and the closing of all but perceived 'essential' businesses to 'save' us from an alleged 'deadly virus'. Cult corporations like Amazon and Walmart were naturally considered 'essential' while mom and pop shops and stores had their doors closed by fascist decree. As a result with every new lockdown and new regulation more small and medium, even large businesses not owned by the Cult, went to the wall while Cult giants and their frontmen and women grew financially fatter by the second. Mom and pop were denied an income and the right to earn a living and the wealth of people like Jeff Bezos (Amazon), Mark Zuckerberg (Facebook) and Sergei Brin and Larry Page (Google/Alphabet) have reached record levels. The Cult was increasing its own power through further dramatic concentrations of wealth while the competition was being destroyed and brought into a state of dependency. Lockdowns have been instigated to secure that very end and were never anything to do with health. My brother Paul spent 45 years building up a bus repair business, but lockdowns meant buses were running at a fraction of normal levels for months on end. Similar stories can told in their hundreds of millions worldwide. Efforts of a lifetime coldly destroyed by Cult multi-billionaires and their lackeys in government and law enforcement who continued to earn their living from the taxation of the people while denying the right of the same people to earn theirs. How different it would have been if those making and enforcing these decisions had to face the same financial hardships of those they affected, but they never do.

Gates of Hell

Behind it all in the full knowledge of what he is doing and why is the psychopathic figure of Cult operative Bill Gates. His puppet Tedros at the World Health Organization declared ‘Covid’ a pandemic in March, 2020. The WHO had changed the definition of a ‘pandemic’ in 2009 just a month before declaring the ‘swine flu pandemic’ which would not have been so under the previous definition. The same applies to ‘Covid’. The definition had included… ‘an infection by an infectious agent, occurring simultaneously in different countries, with a significant mortality rate relative to the proportion of the population infected’. The new definition removed the need for ‘significant mortality’. The ‘pandemic’ has been fraudulent even down to the definition, but Gates demanded economy-destroying lockdowns, school closures, social distancing, mandatory masks, a ‘vaccination’ for every man, woman and child on the planet and severe consequences and restrictions for those that refused. Who gave him this power? The

Cult did which he serves like a little boy in short trousers doing what his daddy tells him. He and his psychopathic missus even smiled when they said that much worse was to come (what they knew was planned to come). Gates responded in the matter-of-fact way of all psychopaths to a question about the effect on the world economy of what he was doing:

> Well, it won't go to zero but it will shrink. Global GDP is probably going to take the biggest hit ever [Gates was smiling as he said this] … in my lifetime this will be the greatest economic hit. But you don't have a choice. People act as if you have a choice. People don't feel like going to the stadium when they might get infected … People are deeply affected by seeing these stats, by knowing they could be part of the transmission chain, old people, their parents and grandparents, could be affected by this, and so you don't get to say ignore what is going on here.
>
> There will be the ability to open up, particularly in rich countries, if things are done well over the next few months, but for the world at large normalcy only returns when we have largely vaccinated the entire population.

The man has no compassion or empathy. How could he when he's a psychopath like all Cult players? My own view is that even beyond that he is very seriously mentally ill. Look in his eyes and you can see this along with his crazy flailing arms. You don't do what he has done to the world population since the start of 2020 unless you are mentally ill and at the most extreme end of psychopathic. You especially don't do it when to you know, as we shall see, that cases and deaths from 'Covid' are fakery and a product of monumental figure massaging. 'These stats' that Gates referred to are based on a 'test' that's not testing for the 'virus' as he has known all along. He made his fortune with big Cult support as an infamously ruthless software salesman and now buys global control of 'health' (death) policy without the population he affects having any say. It's a breathtaking outrage. Gates talked about people being deeply affected by fear of 'Covid' when that was because of *him* and his global network lying to them minute-by-minute supported by a lying media that he seriously influences and funds to the tune of hundreds of millions. He's handed big sums to media operations including the BBC, NBC, Al Jazeera, Univision, *PBS NewsHour,* *ProPublica, National Journal, The Guardian, The Financial Times, The Atlantic, Texas Tribune, USA Today* publisher Gannett, *Washington Monthly, Le Monde,* Center for Investigative Reporting, Pulitzer Center on Crisis Reporting, National Press Foundation, International Center for Journalists, Solutions Journalism Network, the Poynter Institute for Media Studies, and many more. Gates is everywhere in the 'Covid' hoax and the man must go to prison – or a mental facility – for the rest of his life and his money distributed to those he has taken such enormous psychopathic pleasure in crushing.

The Muscle

The Hunger Games global structure demands a police-military state – a fusion of the two into one force – which viciously imposes the will of the Cult on the population and protects the Cult from public rebellion. In that regard, too, the 'Covid' hoax just keeps on giving. Often unlawful, ridiculous and contradictory 'Covid' rules and regulations have been policed across the world by moronic automatons and psychopaths made faceless by face-nappy masks and acting like the Nazi SS and fascist blackshirts and brownshirts of Hitler and Mussolini. The smallest departure from the rules decreed by the psychos in government and their clueless gofers were jumped upon by the face-nappy fascists. Brutality against public protestors soon became commonplace even on girls, women and old people as the brave men with the batons – the Face-Nappies as I call them – broke up peaceful protests and handed out fines like confetti to people who couldn't earn a living let alone pay hundreds of pounds for what was once an accepted human right. Robot Face-Nappies of Nottingham police in the English East Midlands fined one group £11,000 for attending a child's birthday party. For decades I charted the transformation of law enforcement as genuine, decent officers were replaced with psychopaths and the brain dead who would happily and brutally do whatever their masters told them. Now they were let loose on the public and I would emphasise the point that none of this just happened. The step-by-step change in the dynamic between police and public was orchestrated from the shadows by those who knew where this was all going and the same with the perceptual reframing of those in all levels of authority and official administration through ‘training courses’ by organisations such as Common Purpose which was created in the late 1980s and given a massive boost in Blair era Britain until it became a global phenomenon. Supposed public ‘servants’ began to view the population as the enemy and the same was true of the police. This was the start of the explosion of behaviour manipulation organisations and networks preparing for the all-war on the human psyche unleashed with the dawn of 2020. I will go into more detail about this later in the book because it is a core part of what is happening.

Police desecrated beauty spots to deter people gathering and arrested women for walking in the countryside alone ‘too far’ from their homes. We had arrogant, clueless sergeants in the Isle of Wight police where I live posting on Facebook what they insisted the population must do or else. A schoolmaster sergeant called Radford looked young enough for me to ask if his mother knew he was out, but he was posting what he *expected* people to do while a Sergeant Wilkinson boasted about fining lads for meeting in a McDonald’s car park where they went to get a lockdown takeaway. Wilkinson added that he had even cancelled their order. What a pair of prats these people are and yet they have increasingly become the norm among Jackboot Johnson’s Yellowshirts once known as the British police. This was the theme all over the world with police savagery common during lockdown protests in the United States, the Netherlands, and the fascist state of Victoria in Australia under its tyrannical and again moronic premier Daniel Andrews. Amazing how tyrannical and moronic tend to work as a team and the same combination could be seen across America as arrogant, narcissistic Woke governors and mayors such as Gavin Newsom (California), Andrew Cuomo (New York), Gretchen Whitmer (Michigan), Lori Lightfoot (Chicago) and Eric Garcetti (Los Angeles) did their Nazi and Stalin impressions with the full support of the compliant brutality of their enforcers in uniform as they arrested small business owners defying fascist shutdown orders and took them to jail in ankle shackles and handcuffs. This happened to bistro owner Marlena Pavlos-Hackney in Gretchen Whitmer's fascist state of Michigan when police arrived to enforce an order by a state-owned judge for 'putting the community at risk' at a time when other states like Texas were dropping restrictions and migrants were pouring across the southern border without any 'Covid' questions at all. I'm sure there are many officers appalled by what they are ordered to do, but not nearly enough of them. If they were truly appalled they would not do it. As the months passed every opportunity was taken to have the military involved to make their presence on the streets ever more familiar and 'normal' for the longer-term goal of police-military fusion.

Another crucial element to the Hunger Games enforcement network has been encouraging the public to report neighbours and others for 'breaking the lockdown rules'. The group faced with £11,000 in fines at the child's birthday party would have been dobbed-in by a neighbour with a brain the size of a pea. The technique was most famously employed by the Stasi secret police in communist East Germany who had public informants placed throughout the population. A police chief in the UK says his force doesn't need to carry out 'Covid' patrols when they are flooded with so many calls from the public reporting other people for visiting the beach. Dorset police chief James Vaughan said people were so enthusiastic about snitching on their fellow humans they were now operating as an auxiliary arm of the police: 'We are still getting around 400 reports a week from the public, so we will respond to reports …We won't need to be doing hotspot patrols because people are very quick to pick the phone up and tell us.' Vaughan didn't say that this is a pillar of all tyrannies of whatever complexion and the means to hugely extend the reach of enforcement while spreading distrust among the people and making them wary of doing anything that might get them reported. Those narcissistic Isle of Wight sergeants Radford and Wilkinson never fail to add a link to their Facebook posts where the public can inform on their fellow slaves.

Neither would be self-aware enough to realise they were imitating the Stasi which they might well never have heard of. Government psychologists that I will expose later laid out a policy to turn communities against each other in the same way.

A coincidence? Yep, and I can knit fog

I knew from the start of the alleged pandemic that this was a Cult operation. It presented limitless potential to rapidly advance the Cult agenda and exploit manipulated fear to demand that every man, woman and child on the planet was 'vaccinated' in a process never used on humans before which infuses self-replicating *synthetic* material into human cells. Remember the plan to transform the human body from a biological to a synthetic biological state. I'll deal with the 'vaccine' (that's not actually a vaccine) when I focus on the genetic agenda. Enough to say here that mass global 'vaccination' justified by this 'new virus' set alarms ringing after 30 years of tracking these people and their methods. The 'Covid' hoax officially beginning in China was also a big red flag for reasons I will be explaining. The agenda potential was so enormous that I could dismiss any idea that the 'virus' appeared naturally. Major happenings with major agenda implications never occur without Cult involvement in making them happen. My questions were twofold in early 2020 as the media began its campaign to induce global fear and hysteria: Was this alleged infectious agent released on purpose by the Cult or did it even exist at all? I then did what I always do in these situations. I sat, observed and waited to see where the evidence and information would take me. By March and early April synchronicity was strongly – and ever more so since then – pointing me in the direction of *there is no 'virus'*. I went public on that with derision even from swathes of the alternative media that voiced a scenario that the Chinese government released the 'virus' in league with Deep State elements in the United States from a top-level bio-lab in Wuhan where the 'virus' is said to have first appeared. I looked at that possibility, but I didn't buy it for several reasons. Deaths from the 'virus' did not in any way match what they would have been with a ‘deadly bioweapon’ and it is much more effective if you sell the *illusion* of an infectious agent rather than having a real one unless you can control through injection who has it and who doesn’t. Otherwise you lose control of events. A made-up ‘virus’ gives you a blank sheet of paper on which you can make it do whatever you like and have any symptoms or mutant ‘variants’ you choose to add while a real infectious agent would limit you to what it actually does. A phantom disease allows you to have endless ludicrous ‘studies’ on the ‘Covid’ dollar to widen the perceived impact by inventing ever more ‘at risk’ groups including one study which said those who walk slowly may be almost four times more likely to die from the ‘virus’. People are in psychiatric wards for less.

A real ‘deadly bioweapon’ can take out people in the hierarchy that are not part of the Cult, but essential to its operation. Obviously they don’t want that. Releasing a real disease means you immediately lose control of it. Releasing an illusory one means you don’t. Again it’s vital that people are extra careful when dealing with what they want to hear. A bioweapon unleashed from a Chinese laboratory in collusion with the American Deep State may fit a conspiracy narrative, but is it true? Would it not be far more effective to use the excuse of a ‘virus’ to justify the real bioweapon – the ‘vaccine’? That way your disease agent does not have to be transmitted and arrives directly through a syringe. I saw a French virologist Luc Montagnier quoted in the alternative media as saying he had discovered that the alleged ‘new’ severe acute respiratory syndrome coronavirus , or SARS-CoV-2, was made artificially and included elements of the human immunodeficiency ‘virus’ (HIV) and a parasite that causes malaria. SARS-CoV-2 is alleged to trigger an alleged illness called Covid-19. I remembered Montagnier’s name from my research years before into claims that an HIV ‘retrovirus’ causes AIDs – claims that were demolished by Berkeley virologist Peter Duesberg who showed that no one had ever proved that HIV causes acquired immunodeficiency syndrome or AIDS. Claims that become accepted as fact, publicly and medically, with no proof whatsoever are an ever-recurring story that profoundly applies to

'Covid'. Nevertheless, despite the lack of proof, Montagnier's team at the Pasteur Institute in Paris had a long dispute with American researcher Robert Gallo over which of them discovered and isolated the HIV 'virus' and with *no evidence* found it to cause AIDS. You will see later that there is also no evidence that any 'virus' causes any disease or that there is even such a thing as a 'virus' in the way it is said to exist. The claim to have 'isolated' the HIV 'virus' will be presented in its real context as we come to the shocking story – and it is a story – of SARS-CoV-2 and so will Montagnier's assertion that he identified the full SARS-CoV-2 genome.

Hoax in the making

We can pick up the 'Covid' story in 2010 and the publication by the Rockefeller Foundation of a document called 'Scenarios for the Future of Technology and International Development'. The inner circle of the Rockefeller family has been serving the Cult since John D. Rockefeller (1839-1937) made his fortune with Standard Oil. It is less well known that the same Rockefeller – the Bill Gates of his day – was responsible for establishing what is now referred to as 'Big Pharma', the global network of pharmaceutical companies that make outrageous profits dispensing scalpel and drug 'medicine' and are obsessed with pumping vaccines in ever-increasing number into as many human arms and backsides as possible. John D. Rockefeller was the driving force behind the creation of the 'education' system in the United States and elsewhere specifically designed to program the perceptions of generations thereafter. The Rockefeller family donated exceptionally valuable land in New York for the United Nations building and were central in establishing the World Health Organization in 1948 as an agency of the UN which was created from the start as a Trojan horse and stalking horse for world government. Now enter Bill Gates. His family and the Rockefellers have long been extremely close and I have seen genealogy which claims that if you go back far enough the two families fuse into the same bloodline. Gates has said that the Bill and Melinda Gates Foundation was inspired by the Rockefeller Foundation and why not when both are serving the same Cult? Major tax-exempt foundations are overwhelmingly criminal enterprises in which Cult assets fund the Cult agenda in the guise of 'philanthropy' while avoiding tax in the process. Cult operatives can become mega-rich in their role of front men and women for the psychopaths at the inner core and they, too, have to be psychopaths to knowingly serve such evil. Part of the deal is that a big percentage of the wealth gleaned from representing the Cult has to be spent advancing the ambitions of the Cult and hence you have the Rockefeller Foundation, Bill and Melinda Gates Foundation (and *so* many more) and people like George Soros with his global Open Society Foundations spending their billions in pursuit of global Cult control. Gates is a global public face of the Cult with his interventions in world affairs including Big Tech influence; a central role in the 'Covid' and 'vaccine' scam; promotion of the climate change shakedown; manipulation of education; geoengineering of the skies; and his food-control agenda as the biggest owner of farmland in America, his GMO promotion and through other means. As one writer said: 'Gates monopolizes or wields disproportionate influence over the tech industry, global health and vaccines, agriculture and food policy (including biopiracy and fake food), weather modification and other climate technologies, surveillance, education and media.' The almost limitless wealth secured through Microsoft and other not-allowed-to-fail ventures (including vaccines) has been ploughed into a long, long list of Cult projects designed to enslave the entire human race. Gates and the Rockefellers have been working as one unit with the Rockefeller-established World Health Organization leading global 'Covid' policy controlled by Gates through his mouth-piece Tedros. Gates became the WHO's biggest funder when Trump announced that the American government would cease its donations, but Biden immediately said he would restore the money when he took office in January, 2021. The Gates Foundation (the Cult) owns through limitless funding the world health system and the major players across the globe in the 'Covid' hoax.

Okay, with that background we return to that Rockefeller Foundation document of 2010 headed 'Scenarios for the Future of Technology and International Development' and its 'imaginary' epidemic of a virulent and deadly influenza strain which infected 20 percent of the global population and killed eight million in seven months. The Rockefeller scenario was that the epidemic destroyed economies, closed shops, offices and other businesses and led to governments imposing fierce rules and restrictions that included mandatory wearing of face masks and body-temperature checks to enter communal spaces like railway stations and supermarkets. The document predicted that even after the height of the Rockefeller-envisaged epidemic the authoritarian rule would continue to deal with further pandemics, transnational terrorism, environmental crises and rising poverty. Now you may think that the Rockefellers are our modern-day seers or alternatively, and rather more likely, that they well knew what was planned a few years further on. Fascism had to be imposed, you see, to 'protect citizens from risk and exposure'. The Rockefeller scenario document said:

> During the pandemic, national leaders around the world flexed their authority and imposed airtight rules and restrictions, from the mandatory wearing of face masks to body-temperature checks at the entries to communal spaces like train stations and supermarkets. Even after the pandemic faded, this more authoritarian control and oversight of citizens and their activities stuck and even intensified. In order to protect themselves from the spread of increasingly global problems – from pandemics and transnational terrorism to environmental crises and rising poverty – leaders around the world took a firmer grip on power.
>
> At first, the notion of a more controlled world gained wide acceptance and approval. Citizens willingly gave up some of their sovereignty – and their privacy – to more paternalistic states in exchange for greater safety and stability. Citizens were more tolerant, and even eager, for top-down direction and oversight, and national leaders had more latitude to impose order in the ways they saw fit.
>
> In developed countries, this heightened oversight took many forms: biometric IDs for all citizens, for example, and tighter regulation of key industries whose stability was deemed vital to national interests. In many developed countries, enforced cooperation with a suite of new regulations and agreements slowly but steadily restored both order and, importantly, economic growth.

There we have the prophetic Rockefellers in 2010 and three years later came their paper for the Global Health Summit in Beijing, China, when government representatives, the private sector, international organisations and groups met to discuss the next 100 years of 'global health'. The Rockefeller Foundation-funded paper was called 'Dreaming the Future of Health for the Next 100 Years and more prophecy ensued as it described a dystopian future: 'The abundance of data, digitally tracking and linking people may mean the 'death of privacy' and may replace physical interaction with transient, virtual connection, generating isolation and raising questions of how values are shaped in virtual networks.' Next in the 'Covid' hoax preparation sequence came a 'table top' simulation in 2018 for another 'imaginary' pandemic of a disease called Clade X which was said to kill 900 million people. The exercise was organised by the Gates-funded Johns Hopkins University's Center for Health Security in the United States and this is the very same university that has been compiling the disgustingly and systematically erroneous global figures for 'Covid' cases and deaths. Similar Johns Hopkins health crisis scenarios have included the Dark Winter exercise in 2001 and Atlantic Storm in 2005.

Nostradamus 201

For sheer predictive genius look no further prophecy-watchers than the Bill Gates-funded Event 201 held only six weeks before the 'coronavirus pandemic' is supposed to have broken out in China and Event 201 was based on a scenario of a global 'coronavirus pandemic'. Melinda Gates, the great man's missus, told the BBC that he had 'prepared for years' for a coronavirus pandemic which told us what we already knew. Nostradamugates had predicted in a TED talk in 2015 that a pandemic was coming that would kill a lot of people and demolish the world economy. My god, the man is a machine – possibly even literally. Now here he was only weeks before the real thing funding just such a simulated scenario and involving his friends and associates at Johns Hopkins, the World Economic Forum Cult-front of Klaus Schwab, the United Nations, Johnson & Johnson, major banks, and officials from China and the Centers for Disease Control in the United States. What synchronicity – Johns Hopkins would go on to compile the fraudulent 'Covid' figures, the World Economic Forum and Schwab would push the 'Great Reset' in response to 'Covid', the Centers for Disease Control would be at the forefront of 'Covid' policy in the United States, Johnson & Johnson would produce a 'Covid vaccine', and everything would officially start just weeks later in China. Spooky, eh? They were even accurate in creating a simulation of a 'virus' pandemic because the 'real thing' would also be a simulation. Event 201 was not an exercise preparing for something that might happen; it was a rehearsal for what those in control knew was *going* to happen and very shortly. Hours of this simulation were posted on the Internet and the various themes and responses mirrored what would soon be imposed to transform human society. News stories were inserted and what they said would be commonplace a few weeks later with still more prophecy perfection. Much discussion focused on the need to deal with misinformation and the 'anti-vax movement' which is exactly what happened when the 'virus' arrived – was said to have arrived – in the West.

Cult-owned social media banned criticism and exposure of the official 'virus' narrative and when I said there *was* no 'virus' in early April, 2020, I was banned by one platform after another including YouTube, Facebook and later Twitter. The mainstream broadcast media in Britain was in effect banned from interviewing me by the Tony-Blair-created government broadcasting censor Ofcom headed by career government bureaucrat Melanie Dawes who was appointed just as the 'virus' hoax was about to play out in January, 2020. At the same time the Ickonic media platform was using Vimeo, another ultra-Zionist-owned operation, while our own player was being created and they deleted in an instant hundreds of videos, documentaries, series and shows to confirm their unbelievable vindictiveness. We had copies, of course, and they had to be restored one by one when our player was ready. These people have no class. Sabbatian Facebook promised free advertisements for the Gates-controlled World Health Organization narrative while deleting 'false claims and conspiracy theories' to stop 'misinformation' about the alleged coronavirus. All these responses could be seen just a short while earlier in the scenarios of Event 201. Extreme censorship was absolutely crucial for the Cult because the official story was so ridiculous and unsupportable by the evidence that it could never survive open debate and the free-flow of information and opinion. If you can't win a debate then don't have one is the Cult's approach throughout history. Facebook's little boy front man – front boy – Mark Zuckerberg equated 'credible and accurate information' with official sources and exposing their lies with 'misinformation'.

Silencing those that can see

The censorship dynamic of Event 201 is now the norm with an army of narrative-supporting 'fact-checker' organisations whose entire reason for being is to tell the public that official narratives are true and those exposing them are lying. One of the most appalling of these 'fact-checkers' is called NewsGuard founded by ultra-Zionist Americans Gordon Crovitz and Steven Brill. Crovitz is a former publisher of *The Wall Street Journal*, former Executive Vice President of Dow Jones, a member of the Council on Foreign Relations (CFR), and on the board of the American Association of Rhodes Scholars. The CFR and Rhodes Scholarships, named after Rothschild agent Cecil Rhodes who plundered the gold and diamonds of South Africa for his masters and the Cult, have featured widely in my books. NewsGuard don't seem to like me for some reason – I really can't think why – and they have done all they can to have me censored and discredited which is, to quote an old British politician, like being savaged by a dead sheep. They are, however, like all in the censorship network, very well connected and funded by organisations themselves funded by, or connected to, Bill Gates. As you would expect with anything associated with Gates NewsGuard has an offshoot called HealthGuard which 'fights online health care hoaxes'. How very kind. Somehow the NewsGuard European Managing Director Anna-Sophie Harling, a remarkably young-looking woman with no broadcasting experience and little hands-on work in journalism, has somehow secured a position on the ‘Content Board’ of UK government broadcast censor Ofcom. An executive of an organisation seeking to discredit dissidents of the official narratives is making decisions for the government broadcast ‘regulator’ about content?? Another appalling ‘fact-checker’ is Full Fact funded by George Soros and global censors Google and Facebook.

It’s amazing how many activists in the ‘fact-checking’, ‘anti-hate’, arena turn up in government-related positions – people like UK Labour Party activist Imran Ahmed who heads the Center for Countering Digital Hate founded by people like Morgan McSweeney, now chief of staff to the Labour Party’s hapless and useless ‘leader’ Keir Starmer. Digital Hate – which is what it really is – uses the American spelling of Center to betray its connection to a transatlantic network of similar organisations which in 2020 shapeshifted from attacking people for ‘hate’ to attacking them for questioning the ‘Covid’ hoax and the dangers of the ‘Covid vaccine’. It’s just a coincidence, you understand. This is one of Imran Ahmed’s hysterical statements: ‘I would go beyond calling anti-vaxxers conspiracy theorists to say they are an extremist group that pose a national security risk.’ No one could ever accuse this prat of understatement and he’s including in that those parents who are now against vaccines after their children were damaged for life or killed by them. He’s such a nice man. Ahmed does the rounds of the Woke media getting soft-ball questions from spineless ‘journalists’ who never ask what right he has to campaign to destroy the freedom of speech of others while he demands it for himself. There also seems to be an overrepresentation in Ofcom of people connected to the narrative-worshipping BBC. This incredible global network of narrative-support was super-vital when the ‘Covid’ hoax was played in the light of the mega-whopper lies that have to be defended from the spotlight cast by the most basic intelligence.

Setting the scene

The Cult plays the long game and proceeds step-by-step ensuring that everything is in place before major cards are played and they don't come any bigger than the 'Covid' hoax. The psychopaths can't handle events where the outcome isn't certain and as little as possible – preferably nothing – is left to chance. Politicians, government and medical officials who would follow direction were brought to illusory power in advance by the Cult web whether on the national stage or others like state governors and mayors of America. For decades the dynamic between officialdom, law enforcement and the public was changed from one of service to one of control and dictatorship. Behaviour manipulation networks established within government were waiting to impose the coming 'Covid' rules and regulations specifically designed to subdue and rewire the psyche of the people in the guise of protecting health. These included in the UK the Behavioural Insights Team part-owned by the British government Cabinet Office; the Scientific Pandemic Insights Group on Behaviours (SPI-B); and a whole web of intelligence and military groups seeking to direct the conversation on social media and control the narrative. Among them are the cyberwarfare (on the people) 77th Brigade of the British military which is also coordinated through the Cabinet Office as civilian and military leadership continues to combine in what they call the Fusion Doctrine. The 77th Brigade is a British equivalent of the infamous Israeli (Sabbatian) military cyberwarfare and Internet manipulation operation Unit 8200 which I expose at length in *The Trigger*. Also carefully in place were the medical and science advisers to government – many on the payroll past or present of Bill Gates – and a whole alternative structure of unelected government stood by to take control when elected parliaments were effectively closed down once the 'Covid' card was slammed on the table. The structure I have described here and so much more was installed in every major country through the Cult networks. The top-down control hierarchy looks like this: The Cult – Cult-owned Gates – the World Health Organization and Tedros – Gates-funded or controlled chief medical officers and science 'advisers' (dictators) in each country –

political 'leaders'– law enforcement – The People. Through this simple global communication and enforcement structure the policy of the Cult could be imposed on virtually the entire human population so long as they acquiesced to the fascism. With everything in place it was time for the button to be pressed in late 2019/early 2020.

These were the prime goals the Cult had to secure for its will to prevail:

1) Locking down economies, closing all but designated 'essential' businesses (Cult-owned corporations were 'essential'), and putting the population under house arrest was an imperative to destroy independent income and employment and ensure dependency on the Cult-controlled state in the Hunger Games Society. Lockdowns had to be established as the global blueprint from the start to respond to the 'virus' and followed by pretty much the entire world.

2) The global population had to be terrified into believing in a deadly 'virus' that didn't actually exist so they would unquestioningly obey authority in the belief that authority must know how best to protect them and their families. Software salesman Gates would suddenly morph into the world's health expert and be promoted as such by the Cult-owned media.

3) A method of testing that wasn't testing for the 'virus', but was only claimed to be, had to be in place to provide the illusion of 'cases' and subsequent 'deaths' that had a very different cause to the 'Covid-19' that would be scribbled on the death certificate.

4) Because there was no 'virus' and the great majority testing positive with a test not testing for the 'virus' would have no symptoms of anything the lie had to be sold that people without symptoms (without the 'virus') could still pass it on to others. This was crucial to justify for the first time quarantining – house arresting – healthy people. Without this the economy-destroying lockdown of *everybody* could not have been credibly sold.

5) The 'saviour' had to be seen as a vaccine which beyond evil drug companies were working like angels of mercy to develop as quickly as possible, with all corners cut, to save the day. The public must absolutely not know that the 'vaccine' had nothing to do with a 'virus' or that the contents were ready and waiting with a very different motive long before the 'Covid' card was even lifted from the pack.

I said in March, 2020, that the 'vaccine' would have been created way ahead of the 'Covid' hoax which justified its use and the following December an article in the New York *Intelligencer* magazine said the Moderna 'vaccine' had been 'designed' by

January, 2020. This was 'before China had even acknowledged that the disease could be transmitted from human to human, more than a week before the first confirmed coronavirus case in the United States'. The article said that by the time the first American death was announced a month later 'the vaccine had already been manufactured and shipped to the National Institutes of Health for the beginning of its Phase I clinical trial'. The 'vaccine' was actually 'designed' long before that although even with this timescale you would expect the article to ask how on earth it could have been done that quickly. Instead it asked why the 'vaccine' had not been rolled out then and not months later. Journalism in the mainstream is truly dead. I am going to detail in the next chapter why the 'virus' has never existed and how a hoax on that scale was possible, but first the foundation on which the Big Lie of 'Covid' was built.

The test that doesn't test

Fraudulent 'testing' is the bottom line of the whole 'Covid' hoax and was the means by which a 'virus' that did not exist *appeared* to exist. They could only achieve this magic trick by using a test not testing for the 'virus'. To use a test that *was* testing for the 'virus' would mean that every test would come back negative given there was no 'virus'. They chose to exploit something called the RT-PCR test invented by American biochemist Kary Mullis in the 1980s who said publicly that his PCR test … *cannot detect infectious disease*. Yes, the 'test' used worldwide to detect infectious 'Covid' to produce all the illusory 'cases' and 'deaths' compiled by Johns Hopkins and others *cannot detect infectious disease*. This fact came from the mouth of the man who invented PCR and was awarded the Nobel Prize in Chemistry in 1993 for doing so. Sadly, and incredibly conveniently for the Cult, Mullis died in August, 2019, at the age of 74 just before his test would be fraudulently used to unleash fascism on the world. He was said to have died from pneumonia which was an irony in itself. A few months later he would have had 'Covid-19' on his death certificate. I say the timing of his death was convenient because had he lived Mullis, a brilliant, honest and decent man, would have been vociferously speaking out against the use of his test to detect 'Covid' when it was never designed, or able, to do that. I know that to be true given that Mullis made the same point when his test was used to 'detect' – not detect – HIV. He had been seriously critical of the Gallo/Montagnier claim to have isolated the HIV 'virus' and shown it to cause AIDS for which Mullis said there was no evidence. AIDS is actually not a disease but a series of diseases from which people die all the time. When they die from those *same diseases* after a positive 'test' for HIV then AIDS goes on their death certificate. I think I've heard that before somewhere. Countries instigated a policy with 'Covid' that anyone who tested positive with a test not testing for the 'virus' and died of any other cause within 28 days and even longer 'Covid-19' had to go on the death certificate. Cases have come from the test that can't test for infectious disease and the deaths are those who have died of *anything* after testing positive with a test not testing for the 'virus'. I'll have much more later about the death certificate scandal.

Mullis was deeply dismissive of the now US 'Covid' star Anthony Fauci who he said was a liar who didn't know anything about anything – 'and I would say that to his face – nothing.' He said of Fauci: 'The man thinks he can take a blood sample, put it in an electron microscope and if it's got a virus in there you'll know it – he doesn't understand electron microscopy and he doesn't understand medicine and shouldn't be in a position like he's in.' That position, terrifyingly, has made him the decider of 'Covid' fascism policy on behalf of the Cult in his role as director since 1984 of the National Institute of Allergy and Infectious Diseases (NIAID) while his record of being wrong is laughable; but being wrong, so long as it's the *right kind* of wrong, is why the Cult loves him. He'll say anything the Cult tells him to say. Fauci was made Chief Medical Adviser to the President immediately Biden took office. Biden was installed in the White House by Cult manipulation and one of his first decisions was to elevate Fauci to a position of even more control. This is a coincidence? Yes, and I identify as a flamenco dancer called Lola. How does such an incompetent criminal like Fauci remain in that pivotal position in American health since *the 1980s*? When you serve the Cult it looks after you until you are surplus to requirements. Kary Mullis said prophetically of Fauci and his like: 'Those guys have an agenda and it's not an agenda we would like them to have … they make their own rules, they change them when they want to, and Tony Fauci does not mind going on television in front of the people who pay his salary and lie directly into the camera.' Fauci has done that almost daily since the 'Covid' hoax began. Lying is in Fauci's DNA. To make the situation crystal clear about the PCR test this is a direct quote from its inventor Kary Mullis:

> It [the PCR test] doesn't tell you that you're sick and doesn't tell you that the thing you ended up with was really going to hurt you ...'

Ask yourself why governments and medical systems the world over have been using this very test to decide who is 'infected' with the SARS-CoV-2 'virus' and the alleged disease it allegedly causes, 'Covid-19'. The answer to that question will tell you what has been going on. By the way, here's a little show-stopper – the 'new' SARS-CoV-2 'virus' was 'identified' as such right from the start using … *the PCR test not testing for the 'virus'*. If you are new to this and find that shocking then stick around. I have hardly started yet. Even worse, other 'tests', like the 'Lateral Flow Device' (LFD), are considered so useless that they have to be *confirmed* by the PCR test! Leaked emails written by Ben Dyson, adviser to UK 'Health' Secretary Matt Hancock, said they were 'dangerously unreliable'. Dyson, executive director of strategy at the Department of Health, wrote: 'As of today, someone who gets a positive LFD result in (say) London has at best a 25 per cent chance of it being a true positive, but if it is a self-reported test potentially as low as 10 per cent (on an optimistic assumption about specificity) or as low as 2 per cent (on a more pessimistic assumption).' These are the 'tests' that schoolchildren and the public are being urged to have twice a week or more and have to isolate if they get a positive. Each fake positive goes in the statistics as a 'case' no matter how ludicrously inaccurate and the ‘cases’ drive lockdown, masks and the pressure to ‘vaccinate’. The government said in response to the email leak that the ‘tests’ were accurate which confirmed yet again what shocking bloody liars they are. The real false positive rate is *100 percent* as we’ll see. In another ‘you couldn’t make it up’ the UK government agreed to pay £2.8 billion to California’s Innova Medical Group to supply the irrelevant lateral flow tests. The company’s primary test-making centre is in China. Innova Medical Group, established in March, 2020, is owned by Pasaca Capital Inc, chaired by Chinese-American millionaire Charles Huang who was born in Wuhan.

How it works – and how it doesn’t

The RT-PCR test, known by its full title of Polymerase chain reaction, is used across the world to make millions, even billions, of copies of a DNA/RNA genetic information sample. The process is called ‘amplification’ and means that a tiny sample of genetic material is amplified to bring out the detailed content. I stress that it is not testing for an infectious disease. It is simply amplifying a sample of genetic material. In the words of Kary Mullis: ‘PCR is … just a process that’s used to make a whole lot of something out of something.’ To emphasise the point companies that make the PCR tests circulated around the world to ‘test’ for ‘Covid’ warn on the box that it can’t be used to detect ‘Covid’ or infectious disease and is for research purposes only. It’s okay, rest for a minute and you’ll be fine. This is the test that produces the ‘cases’ and ‘deaths’ that have been used to destroy human society. All those global and national medical and scientific ‘experts’ demanding this destruction to ‘save us’ *KNOW* that the test is not testing for the ‘virus’ and the cases and deaths they claim to be real are an almost unimaginable fraud. Every one of them and so many others including politicians and psychopaths like Gates and Tedros must be brought before Nuremburg-type trials and jailed for the rest of their lives. The more the genetic sample is amplified by PCR the more elements of that material become sensitive to the test and by that I don’t mean sensitive for a ‘virus’ but for elements of the genetic material which is *naturally* in the body or relates to remnants of old conditions of various kinds lying dormant and causing no disease. Once the amplification of the PCR reaches a certain level *everyone* will test positive. So much of the material has been made sensitive to the test that everyone will have some part of it in their body. Even lying criminals like Fauci have said that once PCR amplifications pass 35 cycles everything will be a false positive that cannot be trusted for the reasons I have described. I say, like many proper doctors and scientists, that 100 percent of the 'positives' are false, but let's just go with Fauci for a moment.

He says that any amplification over 35 cycles will produce false positives and yet the US Centers for Disease Control (CDC) and Food and Drug Administration (FDA) have recommended up to *40 cycles* and the National Health Service (NHS) in Britain admitted in an internal document for staff that it was using *45 cycles* of amplification. A long list of other countries has been doing the same and at least one 'testing' laboratory has been using *50 cycles*. Have you ever heard a doctor, medical 'expert' or the media ask what level of amplification has been used to claim a 'positive'. The 'test' comes back 'positive' and so you have the 'virus', end of story. Now we can see how the government in Tanzania could send off samples from a goat and a pawpaw fruit under human names and both came back positive for 'Covid-19'. Tanzania president John Magufuli mocked the 'Covid' hysteria, the PCR test and masks and refused to import the DNA-manipulating 'vaccine'. The Cult hated him and an article sponsored by the Bill Gates Foundation appeared in the London *Guardian* in February, 2021, headed 'It's time for Africa to rein in Tanzania's anti-vaxxer president'. Well, 'reined in' he shortly was. Magufuli appeared in good health, but then, in March, 2021, he was dead at 61 from 'heart failure'. He was replaced by Samia Hassan Suhulu who is connected to Klaus Schwab's World Economic Forum and she immediately reversed Magufuli's 'Covid' policy. A sample of cola tested positive for 'Covid' with the PCR test in Germany while American actress and singer-songwriter Erykah Badu tested positive in one nostril and negative in the other. Footballer Ronaldo called the PCR test ‘bullshit’ after testing positive three times and being forced to quarantine and miss matches when there was nothing wrong with him. The mantra from Tedros at the World Health Organization and national governments (same thing) has been test, test, test. They know that the more tests they can generate the more fake ‘cases’ they have which go on to become ‘deaths’ in ways I am coming to. The UK government has its Operation Moonshot planned to test multiple millions every day in workplaces and schools with free tests for everyone to use twice a week at home in line with the Cult plan from the start to make testing part of life. A government advertisement for an ‘Interim Head of Asymptomatic Testing Communication’ said the job included responsibility for delivering a ‘communications strategy’ (propaganda) ‘to support the expansion of asymptomatic testing that *‘normalises testing as part of everyday life*’. More tests means more fake ‘cases’, ‘deaths’ and fascism. I have heard of, and from, many people who booked a test, couldn’t turn up, and yet got a positive result through the post for a test they’d never even had. The whole thing is crazy, but for the Cult there’s method in the madness. Controlling and manipulating the level of amplification of the test means the authorities can control whenever they want the number of apparent ‘cases’ and ‘deaths’. If they want to justify more fascist lockdown and destruction of livelihoods they keep the amplification high. If they want to give the illusion that lockdowns and the ‘vaccine’ are working then they lower the amplification and ‘cases’ and ‘deaths’ will appear to fall. In January, 2021, the Cult-owned World Health Organization suddenly warned laboratories about over-amplification of the test and to lower the threshold. Suddenly headlines began appearing such as: ‘Why ARE “Covid” cases plummeting?’ This was just when the vaccine rollout was underway and I had predicted months before they would make cases appear to fall through amplification tampering when the ‘vaccine’ came. These people are so predictable.

Cow vaccines?

The question must be asked of what is on the test swabs being poked far up the nose of the population to the base of the brain? A nasal swab punctured one woman's brain and caused it to leak fluid. Most of these procedures are being done by people with little training or medical knowledge. Dr Lorraine Day, former orthopaedic trauma surgeon and Chief of Orthopaedic Surgery at San Francisco General Hospital, says the tests are really a *'vaccine'*. Cows have long been vaccinated this way. She points out that masks have to cover the nose and the mouth where it is claimed the 'virus' exists in saliva. Why then don't they take saliva from the mouth as they do with a DNA test instead of pushing a long swab up the nose towards the brain? The ethmoid bone separates the nasal cavity from the brain and within that bone is the cribriform plate. Dr Day says that when the swab is pushed up against this plate and twisted the procedure is 'depositing things back there'. She claims that among these 'things' are nanoparticles that can enter the brain. Researchers have noted that a team at the Gates-funded Johns Hopkins have designed tiny, star-shaped micro-devices that can latch onto intestinal mucosa and release drugs into the body. Mucosa is the thin skin that covers the inside surface of parts of the body such as *the nose* and mouth and produces mucus to protect them. The Johns Hopkins micro-devices are called 'theragrippers' and were 'inspired' by a parasitic worm that digs its sharp teeth into a host's intestines. Nasal swabs are also coated in the sterilisation agent ethylene oxide. The US National Cancer Institute posts this explanation on its website:

> At room temperature, ethylene oxide is a flammable colorless gas with a sweet odor. It is used primarily to produce other chemicals, including antifreeze. In smaller amounts, ethylene oxide is used as a pesticide and a sterilizing agent. The ability of ethylene oxide to damage DNA makes it an effective sterilizing agent but also accounts for its cancer-causing activity.

The Institute mentions lymphoma and leukaemia as cancers most frequently reported to be associated with occupational exposure to ethylene oxide along with stomach and breast cancers. How does anyone think this is going to work out with the constant testing regime being inflicted on adults and children at home and at school that will accumulate in the body anything that's on the swab?

Doctors know best

It is vital for people to realise that 'hero' doctors 'know' only what the Big Pharma-dominated medical authorities tell them to 'know' and if they refuse to 'know' what they are told to 'know' they are out the door. They are mostly not physicians or healers, but repeaters of the official narrative – or else. I have seen alleged professional doctors on British television make shocking statements that we are supposed to take seriously. One called 'Dr' Amir Khan, who is actually telling patients how to respond to illness, said that men could take the birth pill to 'help slow down the effects of Covid-19'. In March, 2021, another ridiculous 'Covid study' by an American doctor proposed injecting men with the female sex hormone progesterone as a 'Covid' treatment. British doctor Nighat Arif told the BBC that face coverings were now going to be part of ongoing normal. Yes, the vaccine protects you, she said (evidence?) … but the way to deal with viruses in the community was always going to come down to hand washing, face covering and keeping a physical distance. That's not what we were told before the 'vaccine' was circulating. Arif said she couldn't imagine ever again going on the underground or in a lift without a mask. I was just thanking my good luck that she was not my doctor when she said – in March, 2021 – that if 'we are *behaving* and we are doing all the right things' she thought we could 'have our nearest and dearest around us at home … around *Christmas* and N*ew Year!* Her patronising delivery was the usual school teacher talking to six-year-olds as she repeated every government talking point and probably believed them all. If we have learned anything from the 'Covid' experience surely it must be that humanity's perception of doctors needs a fundamental rethink. NHS 'doctor' Sara Kayat told her television audience that the 'Covid vaccine' would '100 percent prevent hospitalisation and death'. Not even Big Pharma claimed that. We have to stop taking 'experts' at their word without question when so many of them are clueless and only repeating the party line on which their careers depend. That is not to say there are not brilliants doctors – there are and I have spoken to many of them since all this began – but you won't see them in the mainstream media or quoted by the psychopaths and yes-people in government.

Remember the name – Christian Drosten

German virologist Christian Drosten, Director of Charité Institute of Virology in Berlin, became a national star after the pandemic hoax began. He was feted on television and advised the German government on 'Covid' policy. Most importantly to the wider world Drosten led a group that produced the 'Covid' testing protocol for the PCR test. What a remarkable feat given the PCR cannot test for infectious disease and even more so when you think that Drosten said that his method of testing for SARS-CoV-2 was developed 'without having virus material available'. *He developed a test for a 'virus' that he didn't have and had never seen*. Let that sink in as you survey the global devastation that came from what he did. The whole catastrophe of Drosten's 'test' was based on the alleged genetic sequence published by Chinese scientists on the Internet. We will see in the next chapter that this alleged 'genetic sequence' has never been produced by China or anyone and cannot be when there *is no* SARS-CoV-2. Drosten, however, doesn't seem to let little details like that get in the way. He was the lead author with Victor Corman from the same Charité Hospital of the paper 'Detection of 2019 novel coronavirus (2019-nCoV) by real-time PCR' published in a magazine called *Eurosurveillance*. This became known as the Corman-Drosten paper. In November, 2020, with human society devastated by the effects of the Corman-Drosten test baloney, the protocol was publicly challenged by 22 international scientists and independent researchers from Europe, the United States, and Japan. Among them were senior molecular geneticists, biochemists, immunologists, and microbiologists. They produced a document headed 'External peer review of the RTPCR test to detect SARS-Cov-2 Reveals 10 Major Flaws At The Molecular and Methodological Level: Consequences

For False-Positive Results'. The flaws in the Corman-Drosten test included the following:

- The test is non-specific because of erroneous design
- Results are enormously variable
- The test is unable to discriminate between the whole 'virus' and viral fragments
- It doesn't have positive or negative controls
- The test lacks a standard operating procedure
- It is unsupported by proper peer view

The scientists said the PCR 'Covid' testing protocol was not founded on science and they demanded the Corman-Drosten paper be retracted by *Eurosurveillance*. They said all present and previous Covid deaths, cases, and 'infection rates' should be subject to a massive retroactive inquiry. Lockdowns and travel restrictions should be reviewed and relaxed and those diagnosed through PCR to have 'Covid-19' should not be forced to isolate. Dr Kevin Corbett, a health researcher and nurse educator with a long academic career producing a stream of peer-reviewed publications at many UK universities, made the same point about the PCR test debacle. He said of the scientists' conclusions: 'Every scientific rationale for the development of that test has been totally destroyed by this paper. It's like Hiroshima/Nagasaki to the Covid test.' He said that China hadn't given them an isolated 'virus' when Drosten developed the test. Instead they had developed the test from *a sequence in a gene bank*.' Put another way … *they made it up*! The scientists were supported in this contention by a Portuguese appeals court which ruled in November, 2020, that PCR tests are unreliable and it is unlawful to quarantine people based solely on a PCR test. The point about China not providing an isolated virus must be true when the 'virus' has never been isolated to this day and the consequences of that will become clear. Drosten and company produced this useless 'protocol' right on cue in January, 2020, just as the 'virus' was said to be moving westward and it somehow managed to successfully pass a peer-review in 24 hours. In other words there was no peer-review for a test that would be used to decide who had ‘Covid’ and who didn’t across the world. The Cult-created, Gates-controlled World Health Organization immediately recommended all its nearly 200 member countries to use the Drosten PCR protocol to detect ‘cases’ and ‘deaths’. The sting was underway and it continues to this day.

So who is this Christian Drosten that produced the means through which death, destruction and economic catastrophe would be justified? His education background, including his doctoral thesis, would appear to be somewhat shrouded in mystery and his track record is dire as with another essential player in the ‘Covid’ hoax, the Gates-funded Professor Neil Ferguson at the Gates-funded Imperial College in London of whom more shortly. Drosten predicted in 2003 that the alleged original SARS ‘virus’ (SARS-1’) was an epidemic that could have serious effects on economies and an effective vaccine would take at least two years to produce. Drosten’s answer to every alleged ‘outbreak’ is a vaccine which you won’t be shocked to know. What followed were just 774 official deaths worldwide and none in Germany where there were only nine cases. That is even if you believe there ever was a SARS ‘virus’ when the evidence is zilch and I will expand on this in the next chapter. Drosten claims to be co-discoverer of ‘SARS-1’ and developed a test for it in 2003. He was screaming warnings about ‘swine flu’ in 2009 and how it was a widespread infection far more severe than any dangers from a vaccine could be and people should get vaccinated. It would be helpful for Drosten’s vocal chords if he simply recorded the words ‘the virus is deadly and you need to get vaccinated’ and copies could be handed out whenever the latest made-up threat comes along. Drosten’s swine flu epidemic never happened, but Big Pharma didn’t mind with governments spending hundreds of millions on vaccines that hardly anyone bothered to use and many who did wished they hadn’t. A study in 2010 revealed that the risk of dying from swine flu, or H1N1, was no higher than that of the annual seasonal flu which is what at least most of ‘it’ really was as in the case of 'Covid-19'. A media investigation into Drosten asked how with such a record of inaccuracy he could be *the* government adviser on these issues. The answer to that question is the same with Drosten, Ferguson and Fauci – they keep on giving the authorities the 'conclusions' and 'advice' they want to hear. Drosten certainly produced the goods for them in January, 2020, with his PCR protocol garbage and provided the foundation of what German internal medicine specialist Dr Claus Köhnlein, co-author of *Virus Mania*, called the 'test pandemic'. The 22 scientists in the *Eurosurveillance* challenge called out conflicts of interest within the Drosten 'protocol' group and with good reason. Olfert Landt, a regular co-author of Drosten 'studies', owns the biotech company TIB Molbiol Syntheselabor GmbH in Berlin which manufactures and sells the tests that Drosten and his mates come up with. They have done this with SARS, Enterotoxigenic E. coli (ETEC), MERS, Zika 'virus', yellow fever, and now 'Covid'. Landt told the *Berliner Zeitung* newspaper:

> The testing, design and development came from the Charité [Drosten and Corman]. We simply implemented it immediately in the form of a kit. And if we don't have the virus, which originally only existed in Wuhan, we can make a synthetic gene to simulate the genome of the virus. That's what we did very quickly.

This is more confirmation that the Drosten test was designed without access to the 'virus' and only a synthetic simulation which is what SARS-CoV-2 really is – a computer-generated synthetic fiction. It's quite an enterprise they have going here. A Drosten team decides what the test for something should be and Landt's biotech company flogs it to governments and medical systems across the world. His company must have made an absolute fortune since the 'Covid' hoax began. Dr Reiner Fuellmich, a prominent German consumer protection trial lawyer in Germany and California, is on Drosten's case and that of Tedros at the World Health Organization for crimes against humanity with a class-action lawsuit being prepared in the United States and other legal action in Germany.

Why China?

Scamming the world with a 'virus' that doesn't exist would seem impossible on the face of it, but not if you have control of the relatively few people that make policy decisions and the great majority of the global media. Remember it's not about changing 'real' reality it's about controlling *perception* of reality. You don't have to make something happen you only have make people *believe* that it's happening. Renegade Minds understand this and are therefore much harder to swindle. 'Covid-19' is not a 'real' 'virus'. It's a mind virus, like a computer virus, which has infected the minds, not the bodies, of billions. It all started, publically at least, in China and that alone is of central significance. The Cult was behind the revolution led by its asset Mao Zedong, or Chairman Mao, which established the People's Republic of China on October 1st, 1949. It should have been called The Cult's Republic of China, but the name had to reflect the recurring illusion that vicious dictatorships are run by and for the people (see all the 'Democratic Republics' controlled by tyrants). In the same way we have the 'Biden' Democratic Republic of America officially ruled by a puppet tyrant (at least temporarily) on behalf of Cult tyrants. The creation of Mao's merciless communist/fascist dictatorship was part of a frenzy of activity by the Cult at the conclusion of World War Two which, like the First World War, it had instigated through its assets in Germany, Britain, France, the United States and elsewhere. Israel was formed in 1948; the Soviet Union expanded its 'Iron Curtain' control, influence and military power with the Warsaw Pact communist alliance in 1955; the United Nations was formed in 1945 as a Cult precursor to world government; and a long list of world bodies would be established including the World Health Organization (1948), World Trade Organization (1948 under another name until 1995), International Monetary Fund (1945) and World Bank (1944). Human society was redrawn and hugely centralised in the global Problem-Reaction-Solution that was World War Two. All these changes were significant. Israel would become the headquarters of the Sabbatians and the revolution in China would prepare the ground and control system for the events of 2019/2020.

Renegade Minds know there are no borders except for public consumption. The Cult is a seamless, borderless global entity and to understand the game we need to put aside labels like borders, nations, countries, communism, fascism and democracy. These delude the population into believing that countries are ruled within their borders by a government of whatever shade when these are mere agencies of a global power. America's illusion of democracy and China's communism/fascism are subsidiaries – vehicles – for the same agenda. We may hear about conflict and competition between America and China and on the lower levels that will be true; but at the Cult level they are branches of the same company in the way of the McDonald's example I gave earlier. I have tracked in the books over the years support by US governments of both parties for Chinese Communist Party infiltration of American society through allowing the sale of land, even military facilities, and the acquisition of American business and university influence. All this is underpinned by the infamous stealing of intellectual property and technological know-how. Cult-owned Silicon Valley corporations waive their fraudulent 'morality' to do business with human-rights-free China; Cult-controlled Disney has become China's PR department; and China in effect owns 'American' sports such as basketball which depends for much of its income on Chinese audiences. As a result any sports player, coach or official speaking out against China's horrific human rights record is immediately condemned or fired by the China-worshipping National Basketball Association. One of the first acts of China-controlled Biden was to issue an executive order telling federal agencies to stop making references to the 'virus' by the 'geographic location of its origin'. Long-time Congressman Jerry Nadler warned that criticising China, America's biggest rival, leads to hate crimes against Asian people in the United States. So shut up you bigot. China is fast closing in on Israel as a country that must not be criticised which is apt, really, given that Sabbatians control them both. The two countries have developed close economic, military, technological and strategic ties which include involvement in China's 'Silk Road' transport and economic initiative to connect China with Europe. Israel was the first country in the Middle East to recognise the establishment of Mao's tyranny in 1950 months after it was established.

Project Wuhan – the 'Covid' Psyop

I emphasise again that the Cult plays the long game and what is happening to the world today is the result of centuries of calculated manipulation following a script to take control step-by-step of every aspect of human society. I will discuss later the common force behind all this that has spanned those centuries and thousands of years if the truth be told. Instigating the Mao revolution in China in 1949 with a 2020 'pandemic' in mind is not only how they work – the 71 years between them is really quite short by the Cult's standards of manipulation preparation. The reason for the Cult's Chinese revolution was to create a fiercely-controlled environment within which an extreme structure for human control could be incubated to eventually be unleashed across the world. We have seen this happen since the 'pandemic' emerged from China with the Chinese control-structure founded on AI technology and tyrannical enforcement sweep across the West. Until the moment when the Cult went for broke in the West and put its fascism on public display Western governments had to pay some lip-service to freedom and democracy to not alert too many people to the tyranny-in-the-making. Freedoms were more subtly eroded and power centralised with covert government structures put in place waiting for the arrival of 2020 when that smokescreen of 'freedom' could be dispensed with. The West was not able to move towards tyranny before 2020 anything like as fast as China which was created as a tyranny and had no limits on how fast it could construct the Cult's blueprint for global control. When the time came to impose that structure on the world it was the same Cult-owned Chinese communist/fascist government that provided the excuse – the 'Covid pandemic'. It was absolutely crucial to the Cult plan for the Chinese response to the 'pandemic' –

draconian lockdowns of the entire population – to become the blueprint that Western countries would follow to destroy the livelihoods and freedom of their people. This is why the Cult-owned, Gates-owned, WHO Director-General Tedros said early on:

> The Chinese government is to be congratulated for the extraordinary measures it has taken to contain the outbreak. China is actually setting a new standard for outbreak response and it is not an exaggeration.

Forbes magazine said of China: '… those measures protected untold millions from getting the disease'. The Rockefeller Foundation 'epidemic scenario' document in 2010 said 'prophetically':

> However, a few countries did fare better – China in particular. The Chinese government's quick imposition and enforcement of mandatory quarantine for all citizens, as well as its instant and near-hermetic sealing off of all borders, saved millions of lives, stopping the spread of the virus far earlier than in other countries and enabling a swifter post-pandemic recovery.

Once again – *spooky*.

The first official story was the 'bat theory' or rather the bat diversion. The source of the 'virus outbreak' we were told was a ''wet market' in Wuhan where bats and other animals are bought and eaten in horrifically unhygienic conditions. Then another story emerged through the alternative media that the 'virus' had been released on purpose or by accident from a BSL-4 (biosafety level 4) laboratory in Wuhan not far from the wet market. The lab was reported to create and work with lethal concoctions and bioweapons. Biosafety level 4 is the highest in the World Health Organization system of safety and containment. Renegade Minds are aware of what I call designer manipulation. The ideal for the Cult is for people to buy its prime narrative which in the opening salvoes of the 'pandemic' was the wet market story. It knows, however, that there is now a considerable worldwide alternative media of researchers sceptical of anything governments say and they are often given a version of events in a form they can perceive as credible while misdirecting them from the real truth. In this case let them think that the conspiracy involved is a 'bioweapon virus' released from the Wuhan lab to keep them from the real conspiracy – *there is no 'virus'*. The WHO's current position on the source of the outbreak at the time of writing appears to be: 'We haven't got a clue, mate.' This is a good position to maintain mystery and bewilderment. The inner circle will know where the 'virus' came from – *nowhere*. The bottom line was to ensure the public believed there *was* a 'virus' and it didn't much matter if they thought it was natural or had been released from a lab. The belief that there was a 'deadly virus' was all that was needed to trigger global panic and fear. The population was terrified into handing their power to authority and doing what they were told. They had to or they were 'all gonna die'.

In March, 2020, information began to come my way from real doctors and scientists and my own additional research which had my intuition screaming: 'Yes, that's it! *There is no virus*.' The 'bioweapon' was not the 'virus'; it was the *'vaccine'* already being talked about that would be the bioweapon. My conclusion was further enhanced by happenings in Wuhan. The 'virus' was said to be sweeping the city and news footage circulated of people collapsing in the street (which they've never done in the West with the same 'virus'). The Chinese government was building 'new hospitals' in a matter of ten days to 'cope with demand' such was the virulent nature of the 'virus'. Yet in what seemed like no time the 'new hospitals' closed – even if they even opened – and China declared itself 'virus-free'. It was back to business as usual. This was more propaganda to promote the Chinese draconian lockdowns in the West as the way to 'beat the virus'. Trouble was that we subsequently had lockdown after lockdown, but never business as usual. As the people of the West and most of the rest of the world were caught in an ever-worsening spiral of lockdown, social distancing, masks, isolated old people, families forced apart, and livelihood destruction, it was party-time in Wuhan. Pictures emerged of thousands of people enjoying pool parties and concerts. It made no sense until you realised there never was a 'virus' and the whole thing was a Cult set-up to transform human society out of one its major global strongholds – China.

How is it possible to deceive virtually the entire world population into believing there is a deadly virus when there is not even a 'virus' let alone a deadly one? It's nothing like as difficult as you would think and that's clearly true because it happened.

Postscript: See end of book Postscript for more on the 'Wuhan lab virus release' story which the authorities and media were pushing heavily in the summer of 2021 to divert attention from the truth that the 'Covid virus' is pure invention.

CHAPTER FIVE

There *is no* 'virus'

You can fool some of the people all of the time, and all of the people some of the time, but you cannot fool all of the people all of the time
Abraham Lincoln

The greatest form of mind control is repetition. The more you repeat the same mantra of alleged 'facts' the more will accept them to be true. It becomes an 'everyone knows that, mate'. If you can also censor any other version or alternative to your alleged 'facts' you are pretty much home and cooking.

By the start of 2020 the Cult owned the global mainstream media almost in its entirety to spew out its 'Covid' propaganda and ignore or discredit any other information and view. Cult-owned social media platforms in Cult-owned Silicon Valley were poised and ready to unleash a campaign of ferocious censorship to obliterate all but the official narrative. To complete the circle many demands for censorship by Silicon Valley were led by the mainstream media as 'journalists' became full-out enforcers for the Cult both as propagandists and censors. Part of this has been the influx of young people straight out of university who have become 'journalists' in significant positions. They have no experience and a headful of programmed perceptions from their years at school and university at a time when today's young are the most perceptually-targeted generations in known human history given the insidious impact of technology. They enter the media perceptually prepared and ready to repeat the narratives of the system that programmed them to repeat its narratives. The BBC has a truly pathetic ‘specialist disinformation reporter’ called Marianna Spring who fits this bill perfectly. She is clueless about the world, how it works and what is really going on. Her role is to discredit anyone doing the job that a proper journalist would do and system-serving hacks like Spring wouldn’t dare to do or even see the need to do. They are too busy licking the arse of authority which can never be wrong and, in the case of the BBC propaganda programme, *Panorama*, contacting payments systems such as PayPal to have a donations page taken down for a film company making documentaries questioning vaccines. Even the BBC soap opera *EastEnders* included a disgracefully biased scene in which an inarticulate white working class woman was made to look foolish for questioning the ‘vaccine’ while a well-spoken black man and Asian woman promoted the government narrative. It ticked every BBC box and the fact that the black and minority community was resisting the ‘vaccine’ had nothing to do with the way the scene was written. The BBC has become a disgusting tyrannical propaganda and censorship operation that should be defunded and disbanded and a free media take its place with a brief to stop censorship instead of demanding it. A BBC ‘interview’ with Gates goes something like: ‘Mr Gates, sir, if I can call you sir, would you like to tell our audience why you are such a great man, a wonderful humanitarian philanthropist, and why you should absolutely be allowed as a software salesman to decide health policy for approaching eight billion people? Thank you, sir, please sir.’ Propaganda programming has been incessant and merciless and when all you hear is the same story from the media, repeated by those around you who have only heard the same story, is it any wonder that people on a grand scale believe absolute mendacious garbage to be true? You are about to see, too, why this level of information control is necessary when the official ‘Covid’ narrative is so nonsensical and unsupportable by the evidence.

Structure of Deceit

The pyramid structure through which the ‘Covid’ hoax has been manifested is very simple and has to be to work. As few people as possible have to be involved with full knowledge of what they are doing – and why – or the real story would get out. At the top of the pyramid are the inner core of the Cult which controls Bill Gates who, in turn, controls the World Health Organization through his pivotal funding and his puppet Director-General mouthpiece, Tedros. Before he was appointed Tedros was chair of the Gates-founded Global Fund to ‘fight against AIDS, tuberculosis and malaria’, a board member of the Gates-funded ‘vaccine alliance’ GAVI, and on the board of another Gates-funded organisation. Gates owns him and picked him for a specific reason – Tedros is a crook and worse. ‘Dr’ Tedros (he’s not a medical doctor, the first WHO chief not to be) was a member of the tyrannical Marxist government of Ethiopia for decades with all its human rights abuses. He has faced allegations of corruption and misappropriation of funds and was exposed three times for covering up cholera epidemics while Ethiopia’s health minister. Tedros appointed the mass-murdering genocidal Zimbabwe dictator Robert Mugabe as a WHO goodwill ambassador for public health which, as with Tedros, is like appointing a psychopath to run a peace and love campaign. The move was so ridiculous that he had to drop Mugabe in the face of widespread condemnation. American economist David Steinman, a Nobel peace prize nominee, lodged a complaint with the International Criminal Court in The Hague over alleged genocide by Tedros when he was Ethiopia’s foreign minister. Steinman says Tedros was a ‘crucial decision maker’ who directed the actions of Ethiopia’s security forces from 2013 to 2015 and one of three officials in charge when those security services embarked on the ‘killing’ and ‘torturing’ of Ethiopians. You can see where Tedros is coming from and it’s sobering to think that he has been the vehicle for Gates and the Cult to direct the global response to ‘Covid’. Think about that. A psychopathic Cult dictates to psychopath Gates who dictates to psychopath Tedros who dictates how countries of the world must respond to a ‘Covid virus’ never scientifically shown to exist. At the same time psychopathic Cult-owned Silicon Valley information giants like Google, YouTube, Facebook and Twitter announced very early on that they would give the Cult/Gates/Tedros/WHO version of the narrative free advertising and censor those who challenged their intelligence-insulting, mendacious story.

The next layer in the global 'medical' structure below the Cult, Gates and Tedros are the chief medical officers and science 'advisers' in each of the WHO member countries which means virtually all of them. Medical officers and arbiters of science (they're not) then take the WHO policy and recommended responses and impose them on their country's population while the political 'leaders' say they are deciding policy (they're clearly not) by 'following the science' on the advice of the 'experts' – the same medical officers and science 'advisers' (dictators). In this way with the rarest of exceptions the entire world followed the same policy of lockdown, people distancing, masks and 'vaccines' dictated by the psychopathic Cult, psychopathic Gates and psychopathic Tedros who we are supposed to believe give a damn about the health of the world population they are seeking to enslave. That, amazingly, is all there is to it in terms of crucial decision-making. Medical staff in each country then follow like sheep the dictates of the shepherds at the top of the national medical hierarchies – chief medical officers and science 'advisers' who themselves follow like sheep the shepherds of the World Health Organization and the Cult. Shepherds at the national level often have major funding and other connections to Gates and his Bill and Melinda Gates Foundation which carefully hands out money like confetti at a wedding to control the entire global medical system from the WHO down.

Follow the money

Christopher Whitty, Chief Medical Adviser to the UK Government at the centre of 'virus' policy, a senior adviser to the government's Scientific Advisory Group for Emergencies (SAGE), and Executive Board member of the World Health Organization, was gifted a grant of $40 million by the Bill and Melinda Gates Foundation for malaria research in Africa. The BBC described the unelected Whitty as 'the official who will probably have the greatest impact on our everyday lives of any individual policymaker in modern times' and so it turned out. What Gates and Tedros have said Whitty has done like his equivalents around the world. Patrick Vallance, co-chair of SAGE and the government's Chief Scientific Adviser, is a former executive of Big Pharma giant GlaxoSmithKline with its fundamental financial and business connections to Bill Gates. In September, 2020, it was revealed that Vallance owned a deferred bonus of shares in GlaxoSmithKline worth £600,000 while the company was 'developing' a 'Covid vaccine'. Move along now – nothing to see here – what could possibly be wrong with that? Imperial College in London, a major player in 'Covid' policy in Britain and elsewhere with its 'Covid-19' Response Team, is funded by Gates and has big connections to China while the now infamous Professor Neil Ferguson, the useless 'computer modeller' at Imperial College is also funded by Gates. Ferguson delivered the dramatically inaccurate excuse for the first lockdowns (much more in the next chapter). The Institute for Health Metrics and Evaluation (IHME) in the United States, another source of outrageously false 'Covid' computer models to justify lockdowns, is bankrolled by Gates who is a vehement promotor of lockdowns. America's version of Whitty and Vallance, the again now infamous Anthony Fauci, has connections to 'Covid vaccine' maker Moderna as does Bill Gates through funding from the Bill and Melinda Gates Foundation. Fauci is director of the National Institute of Allergy and Infectious Diseases (NIAID), a major recipient of Gates money, and they are very close. Deborah Birx who was appointed White House Coronavirus Response Coordinator in February, 2020, is yet another with ties to Gates. Everywhere you look at the different elements around the world behind the coordination and decision making of the 'Covid' hoax there is Bill Gates and his money. They include the World Health Organization; Centers for Disease Control (CDC) in the United States; National Institutes of Health (NIH) of Anthony Fauci; Imperial College and Neil Ferguson; the London School of Hygiene where Chris Whitty worked; Regulatory agencies like the UK Medicines & Healthcare products Regulatory Agency (MHRA) which gave emergency approval for ‘Covid vaccines’; Wellcome Trust; GAVI, the Vaccine Alliance; the Coalition for Epidemic Preparedness Innovations (CEPI); Johns Hopkins University which has compiled the false ‘Covid’ figures; and the World Economic Forum. A Nationalfile.com article said:

> Gates has a lot of pull in the medical world, he has a multi-million dollar relationship with Dr. Fauci, and Fauci originally took the Gates line supporting vaccines and casting doubt on [the drug hydroxychloroquine]. Coronavirus response team member Dr. Deborah Birx, appointed by former president Obama to serve as United States Global AIDS Coordinator, also sits on the board of a group that has received billions from Gates’ foundation, and Birx reportedly used a disputed Bill Gates-funded model for the White House’s Coronavirus effort. Gates is a big proponent for a population lockdown scenario for the Coronavirus outbreak.

Another funder of Moderna is the Defense Advanced Research Projects Agency (DARPA), the technology-development arm of the Pentagon and one of the most sinister organisations on earth. DARPA had a major role with the CIA covert technology-funding operation In-Q-Tel in the development of Google and social media which is now at the centre of global censorship. Fauci and Gates are extremely close and openly admit to talking regularly about ‘Covid’ policy, but then why wouldn’t Gates have a seat at every national ‘Covid’ table after his Foundation committed $1.75 billion to the ‘fight against Covid-19’. When passed through our Orwellian Translation Unit this means that he has bought and paid for the Cult-driven ‘Covid’ response worldwide. Research the major ‘Covid’ response personnel in your own country and you will find the same Gates funding and other connections again and again. Medical and science chiefs following World Health Organization ‘policy’ sit atop a medical hierarchy in their country of administrators, doctors and nursing staff. These ‘subordinates’ are told they must work and behave in accordance with the policy delivered from the ‘top’ of the national ‘health’ pyramid which is largely the policy delivered by the WHO which is the policy delivered by Gates and the Cult. The whole ‘Covid’ narrative has been imposed on medical staff by a climate of fear although great numbers don’t even need that to comply. They do so through breathtaking levels of ignorance and include doctors who go through life simply repeating what Big Pharma and their hierarchical masters tell them to say and believe. No wonder Big Pharma 'medicine' is one of the biggest killers on Planet Earth.

The same top-down system of intimidation operates with regard to the Cult Big Pharma cartel which also dictates policy through national and global medical systems in this way. The Cult and Big Pharma agendas are the same because the former controls and owns the latter. 'Health' administrators, doctors, and nursing staff are told to support and parrot the dictated policy or they will face consequences which can include being fired. How sad it's been to see medical staff meekly repeating and imposing Cult policy without question and most of those who can see through the deceit are only willing to speak anonymously off the record. They know what will happen if their identity is known. This has left the courageous few to expose the lies about the 'virus', face masks, overwhelmed hospitals that aren't, and the dangers of the 'vaccine' that isn't a vaccine. When these medical professionals and scientists, some renowned in their field, have taken to the Internet to expose the truth their articles, comments and videos have been deleted by Cult-owned Facebook, Twitter and YouTube. What a real head-shaker to see YouTube videos with leading world scientists and highly qualified medical specialists with an added link underneath to the notorious Cult propaganda website *Wikipedia* to find the 'facts' about the same subject.

HIV – the 'Covid' trial-run

I'll give you an example of the consequences for health and truth that come from censorship and unquestioning belief in official narratives. The story was told by PCR inventor Kary Mullis in his book *Dancing Naked in the Mind Field*. He said that in 1984 he accepted as just another scientific fact that Luc Montagnier of France's Pasteur Institute and Robert Gallo of America's National Institutes of Health had independently discovered that a 'retrovirus' dubbed HIV (human immunodeficiency virus) caused AIDS. They were, after all, Mullis writes, specialists in retroviruses. This is how the medical and science pyramids work. Something is announced or *assumed* and then becomes an everybody-knows-that purely through repetition of the assumption as if it is fact. Complete crap becomes accepted truth with no supporting evidence and only repetition of the crap. This is how a 'virus' that doesn't exist became the 'virus' that changed the world. The HIV-AIDS fairy story became a multi-billion pound industry and the media poured out propaganda terrifying the world about the deadly HIV 'virus' that caused the lethal AIDS. By then Mullis was working at a lab in Santa Monica, California, to detect retroviruses with his PCR test in blood donations received by the Red Cross. In doing so he asked a virologist where he could find a reference for HIV being the cause of AIDS. 'You don't need a reference,' the virologist said … *'Everybody knows it.'* Mullis said he wanted to quote a reference in the report he was doing and he said he felt a little funny about not knowing the source of such an important discovery when everyone else seemed to. The virologist suggested he cite a report by the Centers for Disease Control and Prevention (CDC) on morbidity and mortality. Mullis read the report, but it only said that an organism had been identified and did not say how. The report did not identify the original scientific work. Physicians, however, *assumed* (key recurring theme) that if the CDC was convinced that HIV caused AIDS then proof must exist. Mullis continues:

> I did computer searches. Neither Montagnier, Gallo, nor anyone else had published papers describing experiments which led to the conclusion that HIV probably caused AIDS. I read the papers in Science for which they had become well known as AIDS doctors, but all they had said there was that they had found evidence of a past infection by something which was probably HIV in some AIDS patients.
>
> They found antibodies. Antibodies to viruses had always been considered evidence of past disease, not present disease. Antibodies signaled that the virus had been defeated. The patient had saved himself. There was no indication in these papers that this virus caused a disease. They didn't show that everybody with the antibodies had the disease. In fact they found some healthy people with antibodies.

Mullis asked why their work had been published if Montagnier and Gallo hadn't really found this evidence, and why had they been fighting so hard to get credit for the discovery? He says he was hesitant to write 'HIV is the probable cause of AIDS' until he found published evidence to support that. 'Tens of thousands of scientists and researchers were spending billions of dollars a year doing research based on this idea,' Mullis writes. 'The reason had to be there somewhere; otherwise these people would not have allowed their research to settle into one narrow channel of investigation.' He said he lectured about PCR at numerous meetings where people were always talking about HIV and he asked them how they knew that HIV was the cause of AIDS:

> Everyone said something. Everyone had the answer at home, in the office, in some drawer. They all knew, and they would send me the papers as soon as they got back. But I never got any papers. Nobody ever sent me the news about how AIDS was caused by HIV.

Eventually Mullis was able to ask Montagnier himself about the reference proof when he lectured in San Diego at the grand opening of the University of California AIDS Research Center. Mullis says this was the last time he would ask his question without showing anger. Montagnier said he should reference the CDC report. 'I read it', Mullis said, and it didn't answer the question. 'If Montagnier didn't know the answer who the hell did?' Then one night Mullis was driving when an interview came on National Public Radio with Peter Duesberg, a prominent virologist at Berkeley and a California Scientist of the Year. Mullis says he finally understood why he could not find references that connected HIV to AIDS – *there weren't any!* No one had ever proved that HIV causes AIDS even though it had spawned a multi-billion pound global industry and the media was repeating this as fact every day in their articles and broadcasts terrifying the shit out of people about AIDS and giving the impression that a positive test for HIV (see 'Covid') was a death sentence. Duesberg was a threat to the AIDS gravy train and the agenda that underpinned it. He was therefore abused and castigated after he told the Proceedings of the National Academy of Sciences there was no good evidence implicating the new ‘virus’. Editors rejected his manuscripts and his research funds were deleted. Mullis points out that the CDC has defined AIDS as one of more than 30 diseases *if accompanied* by a positive result on a test that detects antibodies to HIV; but those same diseases are not defined as AIDS cases when antibodies are not detected:

> If an HIV-positive woman develops uterine cancer, for example, she is considered to have AIDS. If she is not HIV positive, she simply has uterine cancer. An HIV-positive man with tuberculosis has AIDS; if he tests negative he simply has tuberculosis. If he lives in Kenya or Colombia, where the test for HIV antibodies is too expensive, he is simply presumed to have the antibodies and therefore AIDS, and therefore he can be treated in the World Health Organization’s clinic. It’s the only medical help available in some places. And it’s free, because the countries that support WHO are worried about AIDS.

Mullis accuses the CDC of continually adding new diseases (see ever more ‘Covid symptoms’) to the grand AIDS definition and of virtually doctoring the books to make it appear as if the disease continued to spread. He cites how in 1993 the CDC enormously broadened its AIDS definition and county health authorities were delighted because they received $2,500 per year from the Federal government for every reported AIDS case. Ladies and gentlemen, I have just described, via Kary Mullis, the ‘Covid pandemic’ of 2020 and beyond. Every element is the same and it’s been pulled off in the same way by the same networks.

The ‘Covid virus’ exists? Okay – prove it. Er … still waiting

What Kary Mullis described with regard to ‘HIV’ has been repeated with ‘Covid’. A claim is made that a new, or ‘novel’, infection has been found and the entire medical system of the world repeats that as fact exactly as they did with HIV and AIDS. No one in the mainstream asks rather relevant questions such as ‘How do you know?’ and ‘Where is your proof?’ The SARS-Cov-2 ‘virus’ and the ‘Covid-19 disease’ became an overnight ‘everybody-knows-that’. The origin could be debated and mulled over, but what you could not suggest was that ‘SARS-Cov-2’ didn’t exist. That would be ridiculous. ‘Everybody knows’ the ‘virus’ exists. Well, I didn’t for one along with American proper doctors like Andrew Kaufman and Tom Cowan and long-time American proper journalist Jon Rappaport. We dared to pursue the obvious and simple question: ‘Where’s the evidence?’ The overwhelming majority in medicine, journalism and the general public did not think to ask that. After all, *everyone knew* there was a new ‘virus’. Everyone was saying so and I heard it on the BBC. Some would eventually argue that the ‘deadly virus’ was nothing like as deadly as claimed, but few would venture into the realms of its very existence. Had they done so they would have found that the evidence for that claim had gone AWOL as with HIV causes AIDS. In fact, not even that. For something to go AWOL it has to exist in the first place and scientific proof for a ‘SARS-Cov-2’ can be filed under nothing, nowhere and zilch.

Dr Andrew Kaufman is a board-certified forensic psychiatrist in New York State, a Doctor of Medicine and former Assistant Professor and Medical Director of Psychiatry at SUNY Upstate Medical University, and Medical Instructor of Hematology and Oncology at the Medical School of South Carolina. He also studied biology at the Massachusetts Institute of Technology (MIT) and trained in Psychiatry at Duke University. Kaufman is retired from allopathic medicine, but remains a consultant and educator on natural healing, I saw a video of his very early on in the ‘Covid’ hoax in which he questioned claims about the ‘virus’ in the absence of any supporting evidence and with plenty pointing the other way. I did everything I could to circulate his work which I felt was asking the pivotal questions that needed an answer. I can recommend an excellent pull-together interview he did with the website The Last Vagabond entitled *Dr Andrew Kaufman: Virus Isolation, Terrain Theory and Covid-19* and his website is andrewkaufmanmd.com. Kaufman is not only a forensic psychiatrist; he is forensic in all that he does. He always reads original scientific papers, experiments and studies instead of second-third-fourth-hand reports about the ‘virus’ in the media which are repeating the repeated repetition of the narrative. When he did so with the original Chinese ‘virus’ papers Kaufman realised that there was no evidence of a ‘SARS-Cov-2’. They had never – from the start – shown it to exist and every repeat of this claim worldwide was based on the accepted existence of proof that was nowhere to be found – see Kary Mullis and HIV. Here we go again.

Let’s postulate

Kaufman discovered that the Chinese authorities immediately concluded that the cause of an illness that broke out among about 200 initial patients in Wuhan was a ‘new virus’ when there were no grounds to make that conclusion. The alleged ‘virus’ was not isolated from other genetic material in their samples and then shown through a system known as Koch’s postulates to be the causative agent of the illness. The world was told that the SARS-Cov-2 ‘virus’ caused a disease they called ‘Covid-19’ which had ‘flu-like’ symptoms and could lead to respiratory problems and pneumonia. If it wasn’t so tragic it would almost be funny. *‘Flu-like’ symptoms’? Pneumonia? Respiratory disease?* What in *CHINA* and particularly in *Wuhan*, one of the most polluted cities in the world with a resulting epidemic of respiratory disease?? Three hundred thousand people get pneumonia in China every year and there are nearly a billion cases worldwide of ‘flu-like symptoms’. These have a whole range of causes – including pollution in Wuhan – but no other possibility was credibly considered in late 2019 when the world was told there was a new and deadly ‘virus’. The global prevalence of pneumonia and ‘flu-like systems’ gave the Cult networks unlimited potential to re-diagnose these other causes as the mythical ‘Covid-19’ and that is what they did from the very start. Kaufman revealed how Chinese medical and science authorities (all subordinates to the Cult-owned communist government) took genetic material from the lungs of only a few of the first patients. The material contained their own cells, bacteria, fungi and other microorganisms living in their bodies. The only way you could prove the existence of the ‘virus’ and its responsibility for the alleged ‘Covid-19’ was to isolate the virus from all the other material – a process also known as ‘purification’ – and then follow the postulates sequence developed in the late 19th century by German physician and bacteriologist Robert Koch which became the 'gold standard' for connecting an alleged causation agent to a disease:

1. The microorganism (bacteria, fungus, virus, etc.) must be present in every case of the disease and all patients must have the same symptoms. It must also *not be present in healthy individuals*.

2. The microorganism must be isolated from the host with the disease. If the microorganism is a bacteria or fungus it must be grown in a pure culture. If it is a virus, it must be purified (i.e. containing no other material except the virus particles) from a clinical sample.

3. The specific disease, with all of its characteristics, must be reproduced when the infectious agent (the purified virus or a pure culture of bacteria or fungi) is inoculated into a healthy, susceptible host.

4. The microorganism must be recoverable from the experimentally infected host as in step 2.

Not one of these criteria has been met in the case of 'SARS-Cov-2' and 'Covid-19'. Not ONE. *EVER*. Robert Koch refers to bacteria and not viruses. What are called 'viral particles' are so minute (hence masks are useless by any definition) that they could only be seen after the invention of the electron microscope in the 1930s and can still only be observed through that means. American bacteriologist and virologist Thomas Milton Rivers, the so-called 'Father of Modern Virology' who was very significantly director of the Rockefeller Institute for Medical Research in the 1930s, developed a less stringent version of Koch's postulates to identify 'virus' causation known as 'Rivers criteria'. 'Covid' did not pass that process either. Some even doubt whether any 'virus' can be isolated from other particles containing genetic material in the Koch method. Freedom of Information requests in many countries asking for scientific proof that the 'Covid virus' has been purified and isolated and shown to exist have all come back with a 'we don't have that' and when this happened with a request to the UK Department of Health they added this comment:

> However, outside of the scope of the [Freedom of Information Act] and on a discretionary basis, the following information has been advised to us, which may be of interest. Most infectious diseases are caused by viruses, bacteria or fungi. Some bacteria or fungi have the capacity to grow on their own in isolation, for example in colonies on a petri dish. Viruses are different in that they are what we call 'obligate pathogens' – that is, they cannot survive or reproduce without infecting a host ...
>
> … For some diseases, it is possible to establish causation between a microorganism and a disease by isolating the pathogen from a patient, growing it in pure culture and reintroducing it to a healthy organism. These are known as 'Koch's postulates' and were developed in 1882. However, as our understanding of disease and different disease-causing agents has advanced, these are no longer the method for determining causation [Andrew Kaufman asks why in that case are there two published articles falsely claiming to satisfy Koch's postulates].
>
> It has long been known that viral diseases cannot be identified in this way as viruses cannot be grown in 'pure culture'. When a patient is tested for a viral illness, this is normally done by looking for the presence of antigens, or viral genetic code in a host with molecular biology techniques [Kaufman asks how you could know the origin of these chemicals without having a pure culture for comparison].
>
> For the record 'antigens' are defined so:
>
> Invading microorganisms have antigens on their surface that the human body can recognise as being foreign – meaning not belonging to it. When the body recognises a foreign antigen, lymphocytes (white blood cells) produce antibodies, which are complementary in shape to the antigen.

Notwithstanding that this is open to question in relation to 'SARS-Cov-2' the presence of 'antibodies' can have many causes and they are found in people that are perfectly well. Kary Mullis said: 'Antibodies … had always been considered evidence of past disease, not present disease.'

'Covid' really is a *computer* 'virus'

Where the UK Department of Health statement says 'viruses' are now 'diagnosed' through a 'viral genetic code in a host with molecular biology techniques', they mean … *the PCR test* which its inventor said cannot test for infectious disease. They have no credible method of connecting a 'virus' to a disease and we will see that there is no scientific proof that any 'virus' causes any disease or there is any such thing as a 'virus' in the way that it is described. Tenacious Canadian researcher Christine Massey and her team made some 40 Freedom of Information requests to national public health agencies in different countries asking for proof that SARS-CoV-2 has been isolated and not one of them could supply that information. Massey said of her request in Canada: 'Freedom of Information reveals Public Health Agency of Canada has no record of 'SARS-COV-2' isolation performed by anyone, anywhere, ever.' If you accept the comment from the UK Department of Health it's because they can't isolate a 'virus'. Even so many 'science' papers claimed to have isolated the 'Covid virus' until they were questioned and had to admit they hadn't. A reply from the Robert Koch Institute in Germany was typical: 'I am not aware of a paper which purified isolated SARS-CoV-2.' So what the hell was Christian Drosten and his gang using to design the 'Covid' testing protocol that has produced all the illusory Covid' cases and 'Covid' deaths when the head of the Chinese version of the CDC admitted there was a problem right from the start in that the 'virus' had never been isolated/purified? Breathe deeply: What they are calling 'Covid' is actually created by a *computer program* i.e. *they made it up* – er, that's it. They took lung fluid, with many sources of genetic material, from one single person alleged to be infected with Covid-19 by a PCR test which they *claimed*, without clear evidence, contained a 'virus'. They used several computer programs to create a model of a theoretical virus genome sequence from more than fifty-six million small sequences of RNA, each of an unknown source, assembling them like a puzzle with no known solution. The computer filled in the gaps with sequences from bits in the gene bank to make it look like a bat SARS-like coronavirus! A wave of the magic wand and poof, an in silico (computer-generated) genome, a scientific fantasy, was created. UK health researcher Dr Kevin Corbett made the same point with this analogy:

> … It's like giving you a few bones and saying that's your fish. It could be any fish. Not even a skeleton. Here's a few fragments of bones. That's your fish … It's all from gene bank and the bits of the virus sequence that weren't there they made up.
>
> They synthetically created them to fill in the blanks. That's what genetics is; it's a code. So it's ABBBCCDDD and you're missing some what you think is EEE so you put it in. It's all

> synthetic. You just manufacture the bits that are missing. This is the end result of the geneticization of virology. This is basically a computer virus.

Further confirmation came in an email exchange between British citizen journalist Frances Leader and the government's Medicines & Healthcare Products Regulatory Agency (the Gates-funded MHRA) which gave emergency permission for untested 'Covid vaccines' to be used. The agency admitted that the 'vaccine' is not based on an isolated 'virus', but comes from a *computer-generated model*. Frances Leader was naturally banned from Cult-owned fascist Twitter for making this exchange public. The process of creating computer-generated alleged 'viruses' is called 'in silico' or 'in silicon' – computer chips – and the term 'in silico' is believed to originate with biological experiments using only a computer in 1989. 'Vaccines' involved with 'Covid' are also produced 'in silico' or by computer not a natural process. If the original 'virus' is nothing more than a made-up computer model how can there be 'new variants' of something that never existed in the first place? They are not new 'variants'; they are new *computer models* only minutely different to the original program and designed to further terrify the population into having the 'vaccine' and submitting to fascism. You want a 'new variant'? Click, click, enter – there you go. Tell the medical profession that you have discovered a 'South African variant', 'UK variants' or a 'Brazilian variant' and in the usual HIV-causes-AIDS manner they will unquestioningly repeat it with no evidence whatsoever to support these claims. They will go on television and warn about the dangers of 'new variants' while doing nothing more than repeating what they have been told to be true and knowing that any deviation from that would be career suicide. Big-time insiders will know it's a hoax, but much of the medical community is clueless about the way they are being played and themselves play the public without even being aware they are doing so. What an interesting 'coincidence' that AstraZeneca and Oxford University were conducting 'Covid vaccine trials' in the three countries – the UK, South Africa and Brazil – where the first three 'variants' were claimed to have 'broken out'.

Here's your 'virus' – it's a unicorn

Dr Andrew Kaufman presented a brilliant analysis describing how the 'virus' was imagined into fake existence when he dissected an article published by *Nature* and written by 19 authors detailing *alleged* 'sequencing of a complete viral genome' of the 'new SARS-CoV-2 virus'. This computer-modelled *in silico* genome was used as a template for all subsequent genome sequencing experiments that resulted in the so-called variants which he said now number more than *6,000*. The fake genome was constructed from more than 56 million individual short strands of RNA. Those little pieces were assembled into longer pieces by finding areas of overlapping sequences. The computer programs created over two million possible combinations from which the authors simply chose the longest one. They then compared this to a 'bat virus' and the computer 'alignment' rearranged the sequence and filled in the gaps! They called this computer-generated abomination the 'complete genome'. Dr Tom Cowan, a fellow medical author and collaborator with Kaufman, said such computer-generation constitutes scientific fraud and he makes this superb analogy:

> Here is an equivalency: A group of researchers claim to have found a unicorn because they found a piece of a hoof, a hair from a tail, and a snippet of a horn. They then add that information into a computer and program it to re-create the unicorn, and they then claim this computer re-creation is the real unicorn. Of course, they had never actually seen a unicorn so could not possibly have examined its genetic makeup to compare their samples with the actual unicorn's hair, hooves and horn.
>
> The researchers claim they decided which is the real genome of SARS-CoV-2 by 'consensus', sort of like a vote. Again, different computer programs will come up with different versions of the imaginary 'unicorn', so they come together as a group and decide which is the real imaginary unicorn.

This is how the 'virus' that has transformed the world was brought into fraudulent 'existence'. Extraordinary, yes, but as the Nazis said the bigger the lie the more will believe it. Cowan, however, wasn't finished and he went on to identify what he called the real blockbuster in the paper. He quotes this section from a paper written by virologists and published by the CDC and then explains what it means:

> Therefore, we examined the capacity of SARS-CoV-2 to infect and replicate in several common primate and human cell lines, including human adenocarcinoma cells (A549), human liver cells (HUH 7.0), and human embryonic kidney cells (HEK-293T). In addition to Vero E6 and Vero CCL81 cells. ... Each cell line was inoculated at high multiplicity of infection and examined 24h post-infection.
>
> No CPE was observed in any of the cell lines except in Vero cells, which grew to greater than 10 to the 7th power at 24 h post-infection. In contrast, HUH 7.0 and 293T showed only modest viral replication, and A549 cells were incompatible with SARS CoV-2 infection.

Cowan explains that when virologists attempt to prove infection they have three possible 'hosts' or models on which they can test. The first was humans. Exposure to humans was generally not done for ethical reasons and has never been done with SARS-CoV-2 or any coronavirus. The second possible host was animals. Cowan said that forgetting for a moment that they never actually use purified virus when exposing animals they do use solutions that they *claim* contain the virus. Exposure to animals has been done with SARS-CoV-2 in an experiment involving mice and this is what they found: *None of the wild (normal) mice got sick*. In a group of genetically-modified mice, a statistically insignificant number lost weight and had slightly bristled fur, but they experienced nothing like the illness called 'Covid-19'. Cowan said the third method – the one they mostly rely on – is to inoculate solutions they *say* contain the virus onto a variety of tissue cultures. This process had never been shown to kill tissue *unless* the sample material was starved of nutrients and poisoned as *part of the process*. Yes, incredibly, in tissue experiments designed to show the 'virus' is responsible for killing the tissue they starve the tissue of nutrients and add toxic drugs including antibiotics and they do not have control studies to see if it's the starvation and poisoning that is degrading the tissue rather than the 'virus' they allege to be in there somewhere. You want me to pinch you? Yep, I understand. Tom Cowan said this about the whole nonsensical farce as he explains what that quote from the CDC paper really means:

> The shocking thing about the above quote is that using their own methods, the virologists found that solutions containing SARS-CoV-2 – even in high amounts – were NOT, I repeat NOT, infective to any of the three human tissue cultures they tested. In plain English, this means they proved, on their terms, that this 'new coronavirus' is not infectious to human beings. It is ONLY infective to monkey kidney cells, and only then when you add two potent drugs (gentamicin and amphotericin), known to be toxic to kidneys, to the mix.
>
> My friends, read this again and again. These virologists, published by the CDC, performed a clear proof, on their terms, showing that the SARS-CoV-2 virus is harmless to human beings. That is the only possible conclusion, but, unfortunately, this result is not even mentioned in their conclusion. They simply say they can provide virus stocks cultured only on monkey Vero cells, thanks for coming.

Cowan concluded: 'If people really understood how this "science" was done, I would hope they would storm the gates and demand honesty, transparency and truth.' Dr Michael Yeadon, former Vice President and Chief Scientific Adviser at drug giant Pfizer has been a vocal critic of the 'Covid vaccine' and its potential for multiple harm. He said in an interview in April, 2021, that 'not one [vaccine] has the virus. He was asked why vaccines normally using a 'dead' version of a disease to activate the immune system were not used for 'Covid' and instead we had the synthetic methods of the 'mRNA Covid vaccine'. Yeadon said that to do the former 'you'd have to have some of [the virus] wouldn't you?' He added: 'No-one's got any – seriously.' Yeadon said that surely they couldn't have fooled the whole world for a year without having a virus, 'but oddly enough ask around – no one's got it'. He didn't know why with all the 'great labs' around the world that the virus had not been isolated – 'Maybe they've been too busy running bad PCR tests and vaccines that people don't need.' What is today called 'science' is not 'science' at all. Science is no longer what is, but whatever people can be manipulated to *believe* that it is. Real science has been hijacked by the Cult to dispense and produce the 'expert scientists' and contentions that suit the agenda of the Cult. How big-time this has happened with the 'Covid' hoax which is entirely based on fake science delivered by fake 'scientists' and fake 'doctors'. The human-caused climate change hoax is also entirely based on fake science delivered by fake 'scientists' and fake 'climate experts'. In both cases real scientists, climate experts and doctors have their views suppressed and deleted by the Cult-owned science establishment, media and Silicon Valley. This is the 'science' that politicians claim to be 'following' and a common denominator of 'Covid' and climate are Cult psychopaths Bill Gates and his mate Klaus Schwab at the Gates-funded World Economic Forum. But, don't worry, it's all just a coincidence and absolutely nothing to worry about. *Zzzzzzzz*.

What is a 'virus' REALLY?

Dr Tom Cowan is one of many contesting the very existence of viruses let alone that they cause disease. This is understandable when there is no scientific evidence for a disease-causing 'virus'. German virologist Dr Stefan Lanka won a landmark case in 2017 in the German Supreme Court over his contention that there is no such thing as a measles virus. He had offered a big prize for anyone who could prove there is and Lanka won his case when someone sought to claim the money. There is currently a prize of more than 225,000 euros on offer from an Isolate Truth Fund for anyone who can prove the isolation of SARS-CoV-2 and its genetic substance. Lanka wrote in an article headed 'The Misconception Called Virus' that scientists think a 'virus' is causing tissue to become diseased and degraded when in fact it is the *processes they are using* which do that – not a 'virus'. Lanka has done an important job in making this point clear as Cowan did in his analysis of the CDC paper. Lanka says that all claims about viruses as disease-causing pathogens are wrong and based on 'easily recognisable, understandable and verifiable misinterpretations.' Scientists believed they were working with 'viruses' in their laboratories when they were really working with 'typical particles of specific dying tissues or cells …' Lanka said that the tissue decaying process claimed to be caused by a 'virus' still happens when no alleged 'virus' is involved. It's the *process* that does the damage and not a 'virus'. The genetic sample is deprived of nutrients, removed from its energy supply through removal from the body and then doused in toxic antibiotics to remove any bacteria. He confirms again that establishment scientists do not (pinch me) conduct control experiments to see if this is the case and if they did they would see the claims that 'viruses' are doing the damage is nonsense. He adds that during the measles 'virus' court case he commissioned an independent laboratory to perform just such a control experiment and the result was that the tissues and cells died in the exact same way as with alleged 'infected' material. This is supported by a gathering number of scientists, doctors and researchers who reject what is called 'germ theory' or the belief in the body being infected by contagious sources emitted by other people. Researchers Dawn Lester and David Parker take the same stance in their highly-detailed and sourced book *What Really Makes You Ill – Why everything you thought you knew about disease is wrong* which was recommended to me by a number of medical professionals genuinely seeking the truth. Lester and Parker say there is no provable scientific evidence to show that a 'virus' can be transmitted between people or people and animals or animals and people:

> The definition also claims that viruses are the cause of many diseases, as if this has been definitively proven. But this is not the case; there is no original scientific evidence that definitively demonstrates that any virus is the cause of any disease. The burden of proof for any theory lies with those who proposed it; but none of the existing documents provides 'proof' that supports the claim that 'viruses' are pathogens.

Dr Tom Cowan employs one of his clever analogies to describe the process by which a 'virus' is named as the culprit for a disease when what is called a 'virus' is only material released by cells detoxing themselves from infiltration by chemical or radiation poisoning. The tidal wave of technologically-generated radiation in the 'smart' modern world plus all the toxic food and drink are causing this to happen more than ever. Deluded 'scientists' misread this as a gathering impact of what they wrongly label 'viruses'.

Paper can infect houses

Cowan said in an article for davidicke.com – with his tongue only mildly in his cheek – that he believed he had made a tremendous discovery that may revolutionise science. He had discovered that small bits of paper are alive, 'well alive-ish', can 'infect' houses, and then reproduce themselves inside the house. The result was that this explosion of growth in the paper inside the house causes the house to explode, blowing it to smithereens. His evidence for this new theory is that in the past months he had carefully examined many of the houses in his neighbourhood and found almost no scraps of paper on the lawns and surrounds of the house. There was an occasional stray label, but nothing more. Then he would return to these same houses a week or so later and with a few, not all of them, particularly the old and decrepit ones, he found to his shock and surprise they were littered with stray bits of paper. He knew then that the paper had infected these houses, made copies of itself, and blew up the house. A young boy on a bicycle at one of the sites told him he had seen a demolition crew using dynamite to explode the house the previous week, but Cowan dismissed this as the idle thoughts of silly boys because 'I was on to something big'. He was on to how 'scientists' mistake genetic material in the detoxifying process for something they call a 'virus'. Cowan said of his house and paper story:

> If this sounds crazy to you, it's because it should. This scenario is obviously nuts. But consider this admittedly embellished, for effect, current viral theory that all scientists, medical doctors and virologists currently believe.

He takes the example of the 'novel SARS-Cov2' virus to prove the point. First they take someone with an undefined illness called 'Covid-19' and don't even attempt to find any virus in their sputum. Never mind the scientists still describe how this 'virus', which they have not located attaches to a cell receptor, injects its genetic material, in 'Covid's' case, RNA, into the cell. The RNA once inserted exploits the cell to reproduce itself and makes 'thousands, nay millions, of copies of itself … Then it emerges victorious to claim its next victim':

If you were to look in the scientific literature for proof, actual scientific proof, that uniform SARS-CoV2 viruses have been properly isolated from the sputum of a sick person, that actual spike proteins could be seen protruding from the virus (which has not been found), you would find that such evidence doesn't exist.

If you go looking in the published scientific literature for actual pictures, proof, that these spike proteins or any viral proteins are ever attached to any receptor embedded in any cell membrane, you would also find that no such evidence exists. If you were to look for a video or documented evidence of the intact virus injecting its genetic material into the body of the cell, reproducing itself and then emerging victorious by budding off the cell membrane, you would find that no such evidence exists.

The closest thing you would find is electron micrograph pictures of cellular particles, possibly attached to cell debris, both of which to be seen were stained by heavy metals, a process that completely distorts their architecture within the living organism. This is like finding bits of paper stuck to the blown-up bricks, thereby proving the paper emerged by taking pieces of the bricks on its way out.

The Enders baloney

Cowan describes the 'Covid' story as being just as make-believe as his paper story and he charts back this fantasy to a Nobel Prize winner called John Enders (1897-1985), an American biomedical scientist who has been dubbed 'The Father of Modern Vaccines'. Enders is claimed to have 'discovered' the process of the viral culture which 'proved' that a 'virus' caused measles. Cowan explains how Enders did this 'by using the EXACT same procedure that has been followed by every virologist to find and characterize every new virus since 1954'. Enders took throat swabs from children with measles and immersed them in 2ml of milk. Penicillin (100u/ml) and the antibiotic streptomycin (50,g/ml) were added and the whole mix was centrifuged – rotated at high speed to separate large cellular debris from small particles and molecules as with milk and cream, for example. Cowan says that if the aim is to find little particles of genetic material ('viruses') in the snot from children with measles it would seem that the last thing you would do is mix the snot with other material – milk –that also has genetic material. 'How are you ever going to know whether whatever you found came from the snot or the milk?' He points out that streptomycin is a 'nephrotoxic' or poisonous-to-the-kidney drug. You will see the relevance of that shortly. Cowan says that it gets worse, much worse, when Enders describes the culture medium upon which the virus ‘grows’: ‘The culture medium consisted of bovine amniotic fluid (90%), beef embryo extract (5%), horse serum (5%), antibiotics and phenol red as an indicator of cell metabolism.’ Cowan asks incredulously: ‘Did he just say that the culture medium also contained fluids and tissues that are themselves rich sources of genetic material?’ The genetic cocktail, or ‘medium’, is inoculated onto tissue and cells from rhesus monkey *kidney* tissue. This is where the importance of streptomycin comes in and currently-used antimicrobials and other drugs that are *poisonous to kidneys* and used in ALL modern viral cultures (e.g. gentamicin, streptomycin, and amphotericin). Cowan asks: ‘How are you ever going to know from this witch’s brew where any genetic material comes from as we now have five different sources of rich genetic material in our mix?’ Remember, he says, that all genetic material, whether from monkey kidney tissues, bovine serum, milk, etc., is made from the exact same components. The same central question returns: ‘How are you possibly going to know that it was the virus that killed the kidney tissue and not the toxic antibiotic and starvation rations on which you are growing the tissue?’ John Enders answered the question himself – *you can’t*:

> A second agent was obtained from an uninoculated culture of monkey kidney cells. The cytopathic changes [death of the cells] it induced in the unstained preparations could not be distinguished with confidence from the viruses isolated from measles.

The death of the cells (‘cytopathic changes’) happened in exactly the same manner, whether they inoculated the kidney tissue with the measles snot or not, Cowan says. ‘This is evidence that the destruction of the tissue, the very proof of viral causation of illness, was not caused by anything in the snot because they saw the same destructive effect when the snot was not even used … the cytopathic, i.e., cell-killing, changes come from the process of the culture itself, not from any virus in any snot, period.’ Enders quotes in his 1957 paper a virologist called Ruckle as reporting similar findings ‘and in addition has isolated an agent from monkey kidney tissue that is so far indistinguishable from human measles virus'. In other words, Cowan says, these particles called 'measles viruses' are simply and clearly breakdown products of the starved and poisoned tissue. For measles 'virus' see all 'viruses' including the so-called 'Covid virus'. Enders, the 'Father of Modern Vaccines', also said:

> There is a potential risk in employing cultures of primate cells for the production of vaccines composed of attenuated virus, since the presence of other agents possibly latent in primate tissues cannot be definitely excluded by any known method.

Cowan further quotes from a paper published in the journal *Viruses* in May, 2020, while the 'Covid pandemic' was well underway in the media if not in reality. 'EVs' here refers to particles of genetic debris from our own tissues, such as exosomes of which more in a moment: 'The remarkable resemblance between EVs and viruses has caused quite a few problems in the studies focused on the analysis of EVs released during viral infections.' Later the paper adds that to date a reliable method that can actually guarantee a complete separation (of EVs from viruses) DOES NOT EXIST. This was published at a time when a fairy tale 'virus' was claimed in total certainty to be causing a fairy tale 'viral disease' called 'Covid-19' – a fairy tale that was already well on the way to transforming human society in the image that the Cult has worked to achieve for so long. Cowan concludes his article:

> To summarize, there is no scientific evidence that pathogenic viruses exist. What we think of as 'viruses' are simply the normal breakdown products of dead and dying tissues and cells. When we are well, we make fewer of these particles; when we are starved, poisoned, suffocated by wearing masks, or afraid, we make more.
>
> There is no engineered virus circulating and making people sick. People in laboratories all over the world are making genetically modified products to make people sick. These are called vaccines. There is no virome, no 'ecosystem' of viruses, viruses are not 8%, 50% or 100 % of our genetic material. These are all simply erroneous ideas based on the misconception called a virus.

What is 'Covid'? Load of bollocks

The background described here by Cowan and Lanka was emphasised in the first video presentation that I saw by Dr Andrew Kaufman when he asked whether the 'Covid virus' was in truth a natural defence mechanism of the body called 'exosomes'. These are released by cells when in states of toxicity – see the same themes returning over and over. They are released ever more profusely as chemical and radiation toxicity increases and think of the potential effect therefore of 5G alone as its destructive frequencies infest the human energetic information field with a gathering pace (5G went online in Wuhan in 2019 as the 'virus' emerged). I'll have more about this later. Exosomes transmit a warning to the rest of the body that 'Houston, we have a problem'. Kaufman presented images of exosomes and compared them with 'Covid' under an electron microscope and the similarity was remarkable. They both attach to the same cell receptors (*claimed* in the case of 'Covid'), contain the same genetic material in the form of RNA or ribonucleic acid, and both are found in 'viral cell cultures' with damaged or dying cells. James Hildreth MD, President and Chief Executive Officer of the Meharry Medical College at Johns Hopkins, said: 'The virus is fully an exosome in every sense of the word.' Kaufman's conclusion was that there is no 'virus': 'This entire pandemic is a completely manufactured crisis … there is no evidence of anyone dying from [this] illness.' Dr Tom Cowan and Sally Fallon Morell, authors of *The Contagion Myth*, published a statement with Dr Kaufman in February, 2021, explaining why the 'virus' does not exist and you can read it that in full in the Appendix.

'Virus' theory can be traced to the 'cell theory' in 1858 of German physician Rudolf Virchow (1821-1920) who contended that disease originates from a single cell infiltrated by a 'virus'. Dr Stefan Lanka said that findings and insights with respect to the structure, function and central importance of tissues in the creation of life, which were already known in 1858, comprehensively refute the cell theory. Virchow ignored them. We have seen the part later played by John Enders in the 1950s and Lanka notes that infection theories were only established as a global dogma through the policies and eugenics of the Third Reich in Nazi Germany (creation of the same Sabbatian cult behind the 'Covid' hoax). Lanka said: 'Before 1933, scientists dared to contradict this theory; after 1933, these critical scientists were silenced'. Dr Tom Cowan's view is that ill-heath is caused by too much of something, too little of something, or toxification from chemicals and radiation – not contagion. We must also highlight as a major source of the 'virus' theology a man still called the 'Father of Modern Virology' – Thomas Milton Rivers (1888-1962). There is no way given the Cult's long game policy that it was a coincidence for the 'Father of Modern Virology' to be director of the Rockefeller Institute for Medical Research from 1937 to 1956 when he is credited with making the Rockefeller Institute a leader in 'viral research'. Cult Rockefellers were the force behind the creation of Big Pharma 'medicine', established the World Health Organisation in 1948, and have long and close associations with the Gates family that now runs the WHO during the pandemic hoax through mega-rich Cult gofer and psychopath Bill Gates.

Only a Renegade Mind can see through all this bullshit by asking the questions that need to be answered, not taking 'no' or prevarication for an answer, and certainly not hiding from the truth in fear of speaking it. Renegade Minds have always changed the world for the better and they will change this one no matter how bleak it may currently appear to be.

CHAPTER SIX

Sequence of deceit

If you tell the truth, you don't have to remember anything
Mark Twain

Against the background that I have laid out this far the sequence that took us from an invented 'virus' in Cult-owned China in late 2019 to the fascist transformation of human society can be seen and understood in a whole new context.

We were told that a deadly disease had broken out in Wuhan and the world media began its campaign (coordinated by behavioural psychologists as we shall see) to terrify the population into unquestioning compliance. We were shown images of Chinese people collapsing in the street which never happened in the West with what was supposed to be the same condition. In the earliest days when alleged cases and deaths were few the fear register was hysterical in many areas of the media and this would expand into the common media narrative across the world. The real story was rather different, but we were never told that. The Chinese government, one of the Cult's biggest centres of global operation, said they had discovered a new illness with flu-like and pneumonia-type symptoms in a city with such toxic air that it is overwhelmed with flu-like symptoms, pneumonia and respiratory disease. Chinese scientists said it was a new – 'novel' – coronavirus which they called Sars-Cov-2 and that it caused a disease they labelled 'Covid-19'. There was no evidence for this and the 'virus' has never to this day been isolated, purified and its genetic code established from that. It was from the beginning a computer-generated fiction. Stories of Chinese whistleblowers saying the number of deaths was being supressed or that the 'new disease' was related to the Wuhan bio-lab misdirected mainstream and alternative media into cul-de-sacs to obscure the real truth – there was no 'virus'.

Chinese scientists took genetic material from the lung fluid of just a few people and said they had found a 'new' disease when this material had a wide range of content. There was no evidence for a 'virus' for the very reasons explained in the last two chapters. The 'virus' has never been shown to (a) exist and (b) cause any disease. People were diagnosed on symptoms that are so widespread in Wuhan and polluted China and with a PCR test that can't detect infectious disease. On this farce the whole global scam was sold to the rest of the world which would also diagnose respiratory disease as 'Covid-19' from symptoms alone or with a PCR test not testing for a 'virus'. Flu miraculously disappeared *worldwide* in 2020 and into 2021 as it was redesignated 'Covid-19'. It was really the same old flu with its 'flu-like' symptoms attributed to 'flu-like' 'Covid-19'. At the same time with very few exceptions the Chinese response of draconian lockdown and fascism was the chosen weapon to respond across the West as recommended by the Cult-owned Tedros at the Cult-owned World Health Organization run by the Cult-owned Gates. All was going according to plan. Chinese scientists – everything in China is controlled by the Cult-owned government – compared their contaminated RNA lung-fluid material with other RNA sequences and said it appeared to be just under 80 percent identical to the SARS-CoV-1 'virus' claimed to be the cause of the SARS (severe acute respiratory syndrome) 'outbreak' in 2003. They decreed that because of this the 'new virus' had to be related and they called it SARS-CoV-2. There are some serious problems with this assumption and *assumption* was all it was. Most 'factual' science turns out to be assumptions repeated into everyone-knows-that. A match of under 80-percent is meaningless. Dr Kaufman makes the point that there's a *96 percent* genetic correlation between humans and chimpanzees, but 'no one would say our genetic material is part of the chimpanzee family’. Yet the Chinese authorities were claiming that a much lower percentage, less than 80 percent, proved the existence of a new ‘coronavirus’. For goodness sake human DNA is 60 percent similar to a *banana*.

You are feeling sleepy

The entire ‘Covid’ hoax is a global Psyop, a psychological operation to program the human mind into believing and fearing a complete fantasy. A crucial aspect of this was what *appeared* to happen in Italy. It was all very well streaming out daily images of an alleged catastrophe in Wuhan, but to the Western mind it was still on the other side of the world in a very different culture and setting. A reaction of ‘this could happen to me and my family’ was still nothing like as intense enough for the mind-doctors. The Cult needed a Western example to push people over that edge and it chose Italy, one of its major global locations going back to the Roman Empire. An Italian ‘Covid’ crisis was manufactured in a particular area called Lombardy which just happens to be notorious for its toxic air and therefore respiratory disease. Wuhan, China, déjà vu. An hysterical media told horror stories of Italians dying from ‘Covid’ in their droves and how Lombardy hospitals were being overrun by a tidal wave of desperately ill people needing treatment after being struck down by the ‘deadly virus’. Here was the psychological turning point the Cult had planned. Wow, if this is happening in Italy, the Western mind concluded, this indeed could happen to me and my family. Another point is that Italian authorities responded by following the Chinese blueprint so vehemently recommended by the Cult-owned World Health Organization. They imposed fascistic lockdowns on the whole country viciously policed with the help of surveillance drones sweeping through the streets seeking out anyone who escaped from mass house arrest. Livelihoods were destroyed and psychology unravelled in the way we have witnessed since in all lockdown countries. Crucial to the plan was that Italy responded in this way to set the precedent of suspending freedom and imposing fascism in a ‘Western liberal democracy’. I emphasised in an animated video explanation on davidicke.com posted in the summer of 2020 how important it was to the Cult to expand the Chinese lockdown model across the West. Without this, and the bare-faced lie that non-symptomatic people could still transmit a ‘disease’ they didn’t have, there was no way locking down the whole population, sick and not sick, could be pulled off. At just the right time and with no evidence Cult operatives and gofers claimed that people without symptoms could pass on the ‘disease’. In the name of protecting the ‘vulnerable’ like elderly people, who lockdowns would kill by the tens of thousands, we had for the first time healthy people told to isolate as well as the sick. The great majority of people who tested positive had no symptoms because there was nothing wrong with them. It was just a trick made possible by a test not testing for the ‘virus’.

Months after my animated video the Gates-funded Professor Neil Ferguson at the Gates-funded Imperial College confirmed that I was right. He didn’t say it in those terms, naturally, but he did say it. Ferguson will enter the story shortly for his outrageously crazy ‘computer models’ that led to Britain, the United States and many other countries following the Chinese and now Italian methods of response. Put another way, following the Cult script. Ferguson said that SAGE, the UK government’s scientific advisory group which has controlled ‘Covid’ policy from the start, wanted to follow the Chinese lockdown model (while they all continued to work and be paid), but they wondered if they could possibly, in Ferguson’s words, ‘get away with it in Europe’. ‘Get away with it’? Who the hell do these moronic, arrogant people think they are? This appalling man Ferguson said that once Italy went into national lockdown they realised they, too, could mimic China:

> It’s a communist one-party state, we said. We couldn’t get away with it in Europe, we thought … and then Italy did it. And we realised we could. Behind this garbage from Ferguson is a simple fact: Doing the same as China in every country was the plan from the start and Ferguson’s ‘models’ would play a central role in achieving that. It’s just a coincidence, of course, and absolutely nothing to worry your little head about.

Oops, sorry, our mistake

Once the Italian segment of the Psyop had done the job it was designed to do a very different story emerged. Italian authorities revealed that 99 percent of those who had 'died from Covid-19' in Italy had one, two, three, or more 'co-morbidities' or illnesses and health problems that could have ended their life. The US Centers for Disease Control and Prevention (CDC) published a figure of 94 percent for Americans dying of 'Covid' while having other serious medical conditions – on average two to three (some five or six) other potential causes of death. In terms of death from an unproven 'virus' I say it is 100 percent. The other one percent in Italy and six percent in the US would presumably have died from 'Covid's' flu-like symptoms with a range of other possible causes in conjunction with a test not testing for the 'virus'. Fox News reported that even more startling figures had emerged in one US county in which 410 of 422 deaths attributed to 'Covid-19' had other potentially deadly health conditions. The Italian National Health Institute said later that the average age of people dying with a 'Covid-19' diagnosis in Italy was about 81. Ninety percent were over 70 with ten percent over 90. In terms of other reasons to die some 80 percent had two or more chronic diseases with half having three or more including cardiovascular problems, diabetes, respiratory problems and cancer. Why is the phantom 'Covid-19' said to kill overwhelmingly old people and hardly affect the young? Old people continually die of many causes and especially respiratory disease which you can re-diagnose 'Covid-19' while young people die in tiny numbers by comparison and rarely of respiratory disease. Old people 'die of Covid' because they die of other things that can be redesignated 'Covid' and it really is that simple.

Flu has flown

The blueprint was in place. Get your illusory 'cases' from a test not testing for the 'virus' and redesignate other causes of death as 'Covid-19'. You have an instant 'pandemic' from something that is nothing more than a computer-generated fiction. With near-on a billion people having ‘flu-like’ symptoms every year the potential was limitless and we can see why flu quickly and apparently miraculously disappeared *worldwide* by being diagnosed ‘Covid-19’. The painfully bloody obvious was explained away by the childlike media in headlines like this in the UK ‘*Independent*’: ‘Not a single case of flu detected by Public Health England this year as Covid restrictions suppress virus’. I kid you not. The masking, social distancing and house arrest that did not make the ‘Covid virus’ disappear somehow did so with the ‘flu virus’. Even worse the article, by a bloke called Samuel Lovett, suggested that maybe the masking, sanitising and other ‘Covid’ measures should continue to keep the flu away. With a ridiculousness that disturbs your breathing (it’s ‘Covid-19’) the said Lovett wrote: ‘With widespread social distancing and mask-wearing measures in place throughout the UK, the usual routes of transmission for influenza have been blocked.’ He had absolutely no evidence to support that statement, but look at the consequences of him acknowledging the obvious. With flu not disappearing at all and only being relabelled ‘Covid-19’ he would have to contemplate that ‘Covid’ was a hoax on a scale that is hard to imagine. You need guts and commitment to truth to even go there and that’s clearly something Samuel Lovett does not have in abundance. He would never have got it through the editors anyway.

Tens of thousands die in the United States alone every winter from flu including many with pneumonia complications. CDC figures record *45 million* Americans diagnosed with flu in 2017-2018 of which 61,000 died and some reports claim 80,000. Where was the same hysteria then that we have seen with ‘Covid-19’? Some 250,000 Americans are admitted to hospital with pneumonia every year with about 50,000 cases proving fatal. About 65 million suffer respiratory disease every year and three million deaths makes this the third biggest cause of death worldwide. You only have to redesignate a portion of all these people ‘Covid-19’ and you have an instant global pandemic or the *appearance* of one. Why would doctors do this? They are told to do this and all but a few dare not refuse those who must be obeyed. Doctors in general are not researching their own knowledge and instead take it direct and unquestioned from the authorities that own them and their careers. The authorities say they must now diagnose these symptoms 'Covid-19' and not flu, or whatever, and they do it. Dark suits say put 'Covid-19' on death certificates no matter what the cause of death and the doctors do it. Renegade Minds don't fall for the illusion that doctors and medical staff are all highly-intelligent, highly-principled, seekers of medical truth. *Some are*, but not the majority. They are repeaters, gofers, and yes sir, no sir, purveyors of what the system demands they purvey. The 'Covid' con is not merely confined to diseases of the lungs. Instructions to doctors to put 'Covid-19' on death certificates for anyone dying of *anything* within 28 days (or much more) of a positive test not testing for the 'virus' opened the floodgates. The term dying *with* 'Covid' and not *of* 'Covid' was coined to cover the truth. Whether it was a *with* or an *of* they were all added to the death numbers attributed to the 'deadly virus' compiled by national governments and globally by the Gates-funded Johns Hopkins operation in the United States that was so involved in those 'pandemic' simulations. Fraudulent deaths were added to the ever-growing list of fraudulent 'cases' from false positives from a false test. No wonder Professor Walter Ricciardi, scientific advisor to the Italian minister of health, said after the Lombardy hysteria had done its job that 'Covid' death rates were due to Italy having the second oldest population in the world and to *how hospitals record deaths*:

> The way in which we code deaths in our country is very generous in the sense that all the people who die in hospitals with the coronavirus are deemed to be dying of the coronavirus. On re-evaluation by the National Institute of Health, only 12 per cent of death certificates have shown a direct causality from coronavirus, while 88 per cent of patients who have died have at least one pre-morbidity – many had two or three.

This is extraordinary enough when you consider the propaganda campaign to use Italy to terrify the world, but how can they even say twelve percent were genuine when the 'virus' has not been shown to exist, its 'code' is a computer program, and diagnosis comes from a test not testing for it? As in China, and soon the world, 'Covid-19' in

Italy was a redesignation of diagnosis. Lies and corruption were to become the real 'pandemic' fuelled by a pathetically-compliant medical system taking its orders from the tiny few at the top of their national hierarchy who answered to the World Health Organization which answers to Gates and the Cult. Doctors were told – ordered – to diagnose a particular set of symptoms 'Covid-19' and put that on the death certificate for any cause of death if the patient had tested positive with a test not testing for the virus or had 'Covid' symptoms like the flu. The United States even introduced big financial incentives to manipulate the figures with hospitals receiving £4,600 from the Medicare system for diagnosing someone with regular pneumonia, $13,000 if they made the diagnosis from the same symptoms 'Covid-19' pneumonia, and $39, 000 if they put a 'Covid' diagnosed patient on a ventilator that would almost certainly kill them. A few – painfully and pathetically few – medical whistleblowers revealed (before Cult-owned YouTube deleted their videos) that they had been instructed to 'let the patient crash' and put them straight on a ventilator instead of going through a series of far less intrusive and dangerous methods as they would have done before the pandemic hoax began and the financial incentives kicked in. We are talking cold-blooded murder given that ventilators are so damaging to respiratory systems they are usually the last step before heaven awaits. Renegade Minds never fall for the belief that people in white coats are all angels of mercy and cannot be full-on psychopaths. I have explained in detail in *The Answer* how what I am describing here played out across the world coordinated by the World Health Organization through the medical hierarchies in almost every country.

Medical scientist calls it

Information about the non-existence of the 'virus' began to emerge for me in late March, 2020, and mushroomed after that. I was sent an email by Sir Julian Rose, a writer, researcher, and organic farming promotor, from a medical scientist friend of his in the United States. Even at that early stage in March the scientist was able to explain how the ‘Covid’ hoax was being manipulated. He said there were no reliable tests for a specific ‘Covid-19 virus’ and nor were there any reliable agencies or media outlets for reporting numbers of actual ‘Covid-19’ cases. We have seen in the long period since then that he was absolutely right. ‘Every action and reaction to Covid-19 is based on totally flawed data and we simply cannot make accurate assessments,’ he said. Most people diagnosed with ‘Covid-19’ were showing nothing more than cold and flu-like symptoms ‘because most coronavirus strains *are* nothing more than cold/flu-like symptoms’. We had farcical situations like an 84-year-old German man testing positive for ‘Covid-19’ and his nursing home ordered to quarantine only for him to be found to have a common cold. The scientist described back then why PCR tests and what he called the ‘Mickey Mouse test kits’ were useless for what they were claimed to be identifying. ‘The idea these kits can isolate a specific virus like Covid-19 is nonsense,’ he said. Significantly, he pointed out that ‘if you want to create a totally false panic about a totally false pandemic – pick a coronavirus’. This is exactly what the Cult-owned Gates, World Economic Forum and Johns Hopkins University did with their Event 201 ‘simulation’ followed by their real-life simulation called the ‘pandemic’. The scientist said that all you had to do was select the sickest of people with respiratory-type diseases in a single location – ‘say Wuhan’ – and administer PCR tests to them. You can then claim that anyone showing ‘viral sequences’ similar to a coronavirus ‘which will inevitably be quite a few’ is suffering from a ‘new’ disease:

> Since you already selected the sickest flu cases a fairly high proportion of your sample will go on to die. You can then say this ‘new’ virus has a CFR [case fatality rate] higher than the flu and use this to infuse more concern and do more tests which will of course produce more ‘cases’, which expands the testing, which produces yet more ‘cases’ and so on and so on. Before long you have your ‘pandemic’, and all you have done is use a simple test kit trick to convert the worst flu and pneumonia cases into something new that doesn’t ACTUALLY EXIST [my emphasis].

He said that you then ‘just run the same scam in other countries’ and make sure to keep the fear message running high ‘so that people will feel panicky and less able to think critically'. The only problem to overcome was the fact *there is no* actual new deadly pathogen and only regular sick people. This meant that deaths from the 'new deadly pathogen' were going to be way too low for a real new deadly virus pandemic, but he said this could be overcome in the following ways – all of which would go on to happen:

> 1. You can claim this is just the beginning and more deaths are imminent [you underpin this with fantasy 'computer projections']. Use this as an excuse to quarantine everyone and then claim the quarantine prevented the expected millions of dead.
>
> 2. You can [say that people] 'minimizing' the dangers are irresponsible and bully them into not talking about numbers.
>
> 3. You can talk crap about made up numbers hoping to blind people with pseudoscience.
>
> 4. You can start testing well people (who, of course, will also likely have shreds of coronavirus [RNA] in them) and thus inflate your 'case figures' with 'asymptomatic carriers' (you will of course have to spin that to sound deadly even though any virologist knows the more symptom-less cases you have the less deadly is your pathogen).

The scientist said that if you take these simple steps 'you can have your own entirely manufactured pandemic up and running in weeks'. His analysis made so early in the hoax was brilliantly prophetic of what would actually unfold. Pulling all the information together in these recent chapters we have this is simple 1, 2, 3, of how you can delude virtually the entire human population into believing in a 'virus' that doesn't exist:

- A 'Covid case' is someone who tests positive with a test not testing for the 'virus'.
- A 'Covid death' is someone who dies of *any cause* within 28 days (or much longer) of testing positive with a test not testing for the 'virus.
- Asymptomatic means there is nothing wrong with you, but they claim you can pass on what you don't have to justify locking

down (quarantining) healthy people in totality.

The foundations of the hoax are that simple. A study involving ten million people in Wuhan, published in November, 2020, demolished the whole lie about those without symptoms passing on the 'virus'. They found '300 asymptomatic cases' and traced their contacts to find that not one of them was detected with the 'virus'. 'Asymptomatic' patients and their contacts were isolated for no less than two weeks and nothing changed. I know it's all crap, but if you are going to claim that those without symptoms can transmit 'the virus' then you must produce evidence for that and they never have. Even World Health Organization official Dr Maria Van Kerkhove, head of the emerging diseases and zoonosis unit, said as early as June, 2020, that she doubted the validity of asymptomatic transmission. She said that 'from the data we have, it still seems to be rare that an asymptomatic person actually transmits onward to a secondary individual' and by 'rare' she meant that she couldn't cite any case of asymptomatic transmission.

The Ferguson factor

The problem for the Cult as it headed into March, 2020, when the script had lockdown due to start, was that despite all the manipulation of the case and death figures they still did not have enough people alleged to have died from 'Covid' to justify mass house arrest. This was overcome in the way the scientist described: 'You can claim this is just the beginning and more deaths are imminent … Use this as an excuse to quarantine everyone and then claim the quarantine prevented the expected millions of dead.' Enter one Professor Neil Ferguson, the Gates-funded 'epidemiologist' at the Gates-funded Imperial College in London. Ferguson is Britain's Christian Drosten in that he has a dire record of predicting health outcomes, but is still called upon to advise government on the next health outcome when another 'crisis' comes along. This may seem to be a strange and ridiculous thing to do. Why would you keep turning for policy guidance to people who have a history of being monumentally wrong? Ah, but it makes sense from the Cult point of view. These ‘experts’ keep on producing predictions that suit the Cult agenda for societal transformation and so it was with Neil Ferguson as he revealed his horrific (and clearly insane) computer model predictions that allowed lockdowns to be imposed in Britain, the United States and many other countries. Ferguson does not have even an A-level in biology and would appear to have no formal training in computer modelling, medicine or epidemiology, according to Derek Winton, an MSc in Computational Intelligence. He wrote an article somewhat aghast at what Ferguson did which included taking no account of respiratory disease ‘seasonality’ which means it is far worse in the winter months. Who would have thought that respiratory disease could be worse in the winter? Well, certainly not Ferguson.

The massively China-connected Imperial College and its bizarre professor provided the excuse for the long-incubated Chinese model of human control to travel westward at lightning speed. Imperial College confirms on its website that it collaborates with the Chinese Research Institute; publishes more than 600 research papers every year with Chinese research institutions; has 225 Chinese staff; 2,600 Chinese students – the biggest international group; 7,000 former students living in China which is the largest group outside the UK; and was selected for a tour by China’s President Xi Jinping during his state visit to the UK in 2015. The college takes major donations from China and describes itself as the UK’s number one university collaborator with Chinese research institutions. The China communist/fascist government did not appear phased by the woeful predictions of Ferguson and Imperial when during the lockdown that Ferguson induced the college signed a five-year collaboration deal with China tech giant Huawei that will have Huawei’s indoor 5G network equipment installed at the college’s West London tech campus along with an ‘AI cloud platform’. The deal includes Chinese sponsorship of Imperial’s Venture Catalyst entrepreneurship competition. Imperial is an example of the enormous influence the Chinese government has within British and North American universities and research centres – and further afield. Up to 200 academics from more than a dozen UK universities are being investigated on suspicion of ‘unintentionally’ helping the Chinese government build weapons of mass destruction by ‘transferring world-leading research in advanced military technology such as aircraft, missile designs and cyberweapons’. Similar scandals have broken in the United States, but it’s all a coincidence. Imperial College serves the agenda in many other ways including the promotion of every aspect of the United Nations Agenda 21/2030 (the Great Reset) and produced computer models to show that human-caused ‘climate change’ is happening when in the real world it isn’t. Imperial College is driving the climate agenda as it drives the ‘Covid’ agenda (both Cult hoaxes) while Patrick Vallance, the UK government’s Chief Scientific Adviser on ‘Covid’, was named Chief Scientific Adviser to the UN ‘climate change’ conference known as COP26 hosted by the government in Glasgow, Scotland. ‘Covid’ and ‘climate’ are fundamentally connected.

Professor Woeful

From Imperial’s bosom came Neil Ferguson still advising government despite his previous disasters and it was announced early on that he and other key people like UK Chief Medical Adviser Chris Whitty had caught the ‘virus’ as the propaganda story was being sold. Somehow they managed to survive and we had Prime Minister Boris Johnson admitted to hospital with what was said to be a severe version of the ‘virus’ in this same period. His whole policy and demeanour changed when he returned to Downing Street. It’s a small world with these government advisors – especially in their communal connections to Gates – and Ferguson had partnered with Whitty to write a paper called ‘Infectious disease: Tough choices to reduce Ebola transmission’ which involved another scare-story that didn’t happen. Ferguson’s ‘models’ predicted that up to150, 000 could die from ‘mad cow disease’, or BSE, and its version in sheep if it was transmitted to humans. BSE was not transmitted and instead triggered by an organophosphate pesticide used to treat a pest on cows. Fewer than 200 deaths followed from the human form. Models by Ferguson and his fellow incompetents led to the unnecessary culling of millions of pigs, cattle and sheep in the foot and mouth outbreak in 2001 which destroyed the lives and livelihoods of farmers and their families who had often spent decades building their herds and flocks. Vast numbers of these animals did not have foot and mouth and had no contact with the infection. Another 'expert' behind the cull was Professor Roy Anderson, a computer modeller at Imperial College specialising in the epidemiology of *human*, not animal, disease. Anderson has served on the Bill and Melinda Gates Grand Challenges in Global Health advisory board and chairs another Gates-funded organisation. Gates is everywhere.

In a precursor to the 'Covid' script Ferguson backed closing schools 'for prolonged periods' over the swine flu 'pandemic' in 2009 and said it would affect a third of the world population if it continued to spread at the speed he claimed to be happening. His mates at Imperial College said much the same and a news report said: 'One of the authors, the epidemiologist and disease modeller Neil Ferguson, who sits on the World Health Organisation's emergency committee for the outbreak, said the virus had "full pandemic potential".' Professor Liam Donaldson, the Chris Whitty of his day as Chief Medical Officer, said the worst case could see 30 percent of the British people infected by swine flu with 65,000 dying. Ferguson and Donaldson were indeed proved correct when at the end of the year the number of deaths attributed to swine flu was 392. The term 'expert' is rather liberally applied unfortunately, not least to complete idiots. Swine flu 'projections' were great for GlaxoSmithKline (GSK) as millions rolled in for its Pandemrix influenza vaccine which led to brain damage with children most affected. The British government (taxpayers) paid out more than £60 million in compensation after GSK was given immunity from prosecution. Yet another 'Covid' déjà vu. Swine flu was supposed to have broken out in Mexico, but Dr Wolfgang Wodarg, a German doctor, former member of parliament and critic of the 'Covid' hoax, observed 'the spread of swine flu' in Mexico City at the time. He said: 'What we experienced in Mexico City was a very mild flu which did not kill more than usual – which killed even fewer people than usual.' Hyping the fear against all the facts is not unique to 'Covid' and has happened many times before. Ferguson is reported to have over-estimated the projected death toll of bird flu (H5N1) by some three million-fold, but bird flu vaccine makers again made a killing from the scare. This is some of the background to the Neil Ferguson who produced the perfectly-timed computer models in early 2020 predicting that half a million people would die in Britain without draconian lockdown and 2.2 million in the United States. Politicians panicked, people panicked, and lockdowns of alleged short duration were instigated to 'flatten the curve' of cases gleaned from a test not testing for the 'virus'. I said at the time that the public could forget the 'short duration' bit. This was an agenda to destroy the livelihoods of the population and force them into mass control through dependency and there was going to be nothing 'short' about it. American researcher Daniel Horowitz described the consequences of the 'models' spewed out by Gates-funded Ferguson and Imperial College:

> What led our government and the governments of many other countries into panic was a single Imperial College of UK study, funded by global warming activists, that predicted 2.2 million deaths if we didn't lock down the country. In addition, the reported 8-9% death rate in Italy scared us into thinking there was some other mutation of this virus that they got, which might have come here.
>
> Together with the fact that we were finally testing and had the ability to actually report new cases, we thought we were headed for a death spiral. But again … we can't flatten a curve if we don't know when the curve started.

How about it *never* started?

Giving them what they want

An investigation by German news outlet *Welt Am Sonntag (World on Sunday)* revealed how in March, 2020, the German government gathered together 'leading scientists from several research institutes and universities' and 'together, they were to produce a [modelling] paper that would serve as legitimization for further tough political measures'. The Cult agenda was justified by computer modelling not based on evidence or reality; it was specifically constructed to justify the Cult demand for lockdowns all over the world to destroy the independent livelihoods of the global population. All these modellers and everyone responsible for the 'Covid' hoax have a date with a trial like those in Nuremberg after World War Two when Nazis faced the consequences of their war crimes. These corrupt-beyond-belief 'modellers' wrote the paper according to government instructions and it said that that if lockdown measures were lifted then up to one million Germans would die from 'Covid-19' adding that some would die 'agonizingly at home, gasping for breath' unable to be treated by hospitals that couldn't cope. All lies. No matter – it gave the Cult all that it wanted. What did long-time government 'modeller' Neil Ferguson say? If the UK and the United States didn't lockdown half a million would die in Britain and 2.2 million Americans. Anyone see a theme here? 'Modellers' are such a crucial part of the lockdown strategy that we should look into their background and follow the money. Researcher Rosemary Frei produced an excellent article headlined 'The Modelling-paper Mafiosi'. She highlights a guy called John Edmunds, a British epidemiologist, and professor in the Faculty of Epidemiology and Population Health at the London School of Hygiene & Tropical Medicine. He studied at Imperial College. Edmunds is a member of government 'Covid' advisory bodies which have been dictating policy, the New and Emerging Respiratory Virus Threats Advisory Group (NERVTAG) and the Scientific Advisory Group for Emergencies (SAGE).

Ferguson, another member of NERVTAG and SAGE, led the way with the original 'virus' and Edmunds has followed in the 'variant' stage and especially the so-called UK or Kent variant known as the 'Variant of Concern' (VOC) B.1.1.7. He said in a co-written report for the Centre for Mathematical modelling of Infectious Diseases at the London School of Hygiene and Tropical Medicine, with input from the Centre's 'Covid-19' Working Group, that there was 'a realistic possibility that VOC B.1.1.7 is associated with an increased risk of death compared to non-VOC viruses'. Fear, fear, fear, get the vaccine, fear, fear, fear, get the vaccine. Rosemary Frei reveals that almost all the paper's authors and members of the modelling centre's 'Covid-19' Working Group receive funding from the Bill and Melinda Gates Foundation and/or the associated Gates-funded Wellcome Trust. The paper was published by e-journal *Medr χiv* which only publishes papers not peer-reviewed and the journal was established by an organisation headed by Facebook's Mark Zuckerberg and his missus. What a small world it is. Frei discovered that Edmunds is on the Scientific Advisory Board of the Coalition for Epidemic Preparedness Innovations (CEPI) which was established by the Bill and Melinda Gates Foundation, Klaus Schwab's Davos World Economic Forum and Big Pharma giant Wellcome. CEPI was 'launched in Davos [in 2017] to develop vaccines to stop future epidemics', according to its website. 'Our mission is to accelerate the development of vaccines against emerging infectious diseases and enable equitable access to these vaccines for people during outbreaks.' What kind people they are. Rosemary Frei reveals that Public Health England (PHE) director Susan Hopkins is an author of her organisation's non-peer-reviewed reports on 'new variants'. Hopkins is a professor of infectious diseases at London's Imperial College which is gifted tens of millions of dollars a year by the Bill and Melinda Gates Foundation. Gates-funded modelling disaster Neil Ferguson also co-authors Public Health England reports and he spoke in December, 2020, about the potential danger of the B.1.1.7. 'UK variant' promoted by Gates-funded modeller John Edmunds. When I come to the 'Covid vaccines' the 'new variants' will be shown for what they are – bollocks.

Connections, connections

All these people and modellers are lockdown-obsessed or, put another way, they demand what the Cult demands. Edmunds said in January, 2021, that to ease lockdowns too soon would be a disaster and they had to 'vaccinate much, much, much more widely than the elderly'. Rosemary Frei highlights that Edmunds is married to Jeanne Pimenta who is described in a LinkedIn profile as director of epidemiology at GlaxoSmithKline (GSK) and she held shares in the company. Patrick Vallance, co-chair of SAGE and the government's Chief Scientific Adviser, is a former executive of GSK and has a deferred bonus of shares in the company worth £600,000. GSK has serious business connections with Bill Gates and is collaborating with mRNA-'vaccine' company CureVac to make 'vaccines' for the new variants that Edmunds is talking about. GSK is planning a 'Covid vaccine' with drug giant Sanofi. Puppet Prime Minister Boris Johnson announced in the spring of 2021 that up to 60 million vaccine doses were to be made at the GSK facility at Barnard Castle in the English North East. Barnard Castle, with a population of just 6,000, was famously visited in breach of lockdown rules in April, 2020, by Johnson aide Dominic Cummings who said that he drove there 'to test his eyesight' before driving back to London. Cummings would be better advised to test his integrity – not that it would take long. The GSK facility had nothing to do with his visit then although I'm sure Patrick Vallance would have been happy to arrange an introduction and some tea and biscuits. Ruthless psychopath Gates has made yet another fortune from vaccines in collaboration with Big Pharma companies and gushes at the phenomenal profits to be made from vaccines – more than a 20-to-1 return as he told one interviewer. Gates also tweeted in December, 2019, with the foreknowledge of what was coming: 'What's next for our foundation? I'm particularly excited about what the next year could mean for one of the best buys in global health: vaccines.'

Modeller John Edmunds is a big promotor of vaccines as all these people appear to be. He's the dean of the London School of Hygiene & Tropical Medicine's Faculty of Epidemiology and Population Health which is primarily funded by the Bill and Melinda Gates Foundation and the Gates-established and funded GAVI vaccine alliance which is the Gates vehicle to vaccinate the world. The organisation Doctors Without Borders has described GAVI as being 'aimed more at supporting drug-industry desires to promote new products than at finding the most efficient and sustainable means for fighting the diseases of poverty'. But then that's why the psychopath Gates created it. John Edmunds said in a video that the London School of Hygiene & Tropical Medicine is involved in every aspect of vaccine development including large-scale clinical trials. He contends that mathematical modelling can show that vaccines protect individuals and society. That's on the basis of shit in and shit out, I take it. Edmunds serves on the UK Vaccine Network as does Ferguson and the government's foremost 'Covid' adviser, the grim-faced, dark-eyed Chris Whitty. The Vaccine Network says it works 'to support the government to identify and shortlist targeted investment opportunities for the most promising vaccines and vaccine technologies that will help combat infectious diseases with epidemic potential, and to address structural issues related to the UK's broader vaccine infrastructure'. Ferguson is acting Director of the Imperial College Vaccine Impact Modelling Consortium which has funding from the Bill and Melina Gates Foundation and the Gates-created GAVI 'vaccine alliance'. Anyone wonder why these characters see vaccines as the answer to every problem? Ferguson is wildly enthusiastic in his support for GAVI's campaign to vaccine children en masse in poor countries. You would expect someone like Gates who has constantly talked about the need to reduce the population to want to fund vaccines to keep more people alive. I'm sure that's why he does it. The John Edmunds London School of Hygiene & Tropical Medicine (LSHTM) has a Vaccines Manufacturing Innovation Centre which develops, tests and commercialises vaccines. Rosemary Frei writes:

> The vaccines centre also performs affiliated activities like combating 'vaccine hesitancy'. The latter includes the Vaccine Confidence Project. The project's stated purpose is, among other things, 'to provide analysis and guidance for early response and engagement with the public to ensure sustained confidence in vaccines and immunisation'. The Vaccine Confidence Project's director is LSHTM professor Heidi Larson. For more than a decade she's been researching how to combat vaccine hesitancy.

How the bloody hell can blokes like John Edmunds and Neil Ferguson with those connections and financial ties model 'virus' case and death projections for the government and especially in a way that gives their paymasters like Gates exactly what they want? It's insane, but this is what you find throughout the world.

'Covid' is not dangerous, oops, wait, yes it is

Only days before Ferguson's nightmare scenario made Jackboot Johnson take Britain into a China-style lockdown to save us from a deadly 'virus' the UK government website gov.uk was reporting something very different to Ferguson on a page of official government guidance for 'high consequence infectious diseases (HCID)'. It said this about 'Covid-19':

> As of 19 March 2020, COVID-19 *is no longer considered to be a high consequence infectious diseases (HCID) in the UK* [my emphasis]. The 4 nations public health HCID group made an interim recommendation in January 2020 to classify COVID-19 as an HCID. This was based on consideration of the UK HCID criteria about the virus and the disease with information available during the early stages of the outbreak.
>
> Now that more is known about COVID-19, the public health bodies in the UK have reviewed the most up to date information about COVID-19 against the UK HCID criteria. They have determined that several features have now changed; in particular, more information is available about mortality rates (low overall), and there is now greater clinical awareness and a specific and sensitive laboratory test, the availability of which continues to increase. The Advisory Committee on Dangerous Pathogens (ACDP) is also of the opinion that COVID-19 should no longer be classified as an HCID.

Soon after the government had been exposed for downgrading the risk they upgraded it again and everyone was back to singing from the same Cult hymn book. Ferguson and his fellow Gates clones indicated that lockdowns and restrictions would have to continue until a Gates-funded vaccine was developed. Gates said the same because Ferguson and his like were repeating the Gates script which is the Cult script. 'Flatten the curve' became an ongoing nightmare of continuing lockdowns with periods in between of severe restrictions in pursuit of destroying independent incomes and had nothing to do with protecting health about which the Cult gives not a shit. Why wouldn't Ferguson be pushing a vaccine 'solution' when he's owned by vaccine-obsessive Gates who makes a fortune from them and when Ferguson heads the Vaccine Impact Modelling Consortium at Imperial College funded by the Gates Foundation and GAVI, the ‘vaccine alliance’, created by Gates as his personal vaccine promotion operation? To compound the human catastrophe that Ferguson’s ‘models’ did so much to create he was later exposed for breaking his own lockdown rules by having sexual liaisons with his married girlfriend Antonia Staats at his home while she was living at another location with her husband and children. Staats was a ‘climate’ activist and senior campaigner at the Soros-funded Avaaz which I wouldn’t trust to tell me that grass is green. Ferguson had to resign as a government advisor over this hypocrisy in May, 2020, but after a period of quiet he was back being quoted by the ridiculous media on the need for more lockdowns and a vaccine rollout. Other government-advising ‘scientists’ from Imperial College’ held the fort in his absence and said lockdown could be indefinite until a vaccine was found. The Cult script was being sung by the payrolled choir. I said there was no intention of going back to ‘normal’ when the ‘vaccine’ came because the ‘vaccine’ is part of a very different agenda that I will discuss in Human 2.0. Why would the Cult want to let the world go back to normal when destroying that normal forever was the whole point of what was happening? House arrest, closing businesses and schools through lockdown, (un)social distancing and masks all followed the Ferguson fantasy models. Again as I predicted (these people are so predictable) when the ‘vaccine’ arrived we were told that house arrest, lockdown, (un)social distancing and masks would still have to continue. I will deal with the masks in the next chapter because they are of fundamental importance.

Where’s the ‘pandemic’?

Any mildly in-depth assessment of the figures revealed what was really going on. Cult-funded and controlled organisations still have genuine people working within them such is the number involved. So it is with Genevieve Briand, assistant program director of the Applied Economics master’s degree program at Johns Hopkins

University. She analysed the impact that 'Covid-19' had on deaths from *all* causes in the United States using official data from the CDC for the period from early February to early September, 2020. She found that allegedly 'Covid' *related*-deaths exceeded those from heart disease which she found strange with heart disease always the biggest cause of fatalities. Her research became even more significant when she noted the sudden decline in 2020 of *all* non-'Covid' deaths: 'This trend is completely contrary to the pattern observed in all previous years … the total decrease in deaths by other causes almost exactly equals the increase in deaths by Covid-19.' This was such a game, set and match in terms of what was happening that Johns Hopkins University deleted the article on the grounds that it 'was being used to support false and dangerous inaccuracies about the impact of the pandemic'. No – because it exposed the scam from official CDC figures and this was confirmed when those figures were published in January, 2021. Here we can see the effect of people dying from heart attacks, cancer, road accidents and gunshot wounds – *anything* – having 'Covid-19' on the death certificate along with those diagnosed from 'symptoms' who had even not tested positive with a test not testing for the 'virus'. I am not kidding with the gunshot wounds, by the way. Brenda Bock, coroner in Grand County, Colorado, revealed that two gunshot victims tested positive for the 'virus' within the previous 30 days and were therefore classified as 'Covid deaths'. Bock said: 'These two people had tested positive for Covid, but that's not what killed them. A gunshot wound is what killed them.' She said she had not even finished her investigation when the state listed the gunshot victims as deaths due to the 'virus'. The death and case figures for 'Covid-19' are an absolute joke and yet they are repeated like parrots by the media, politicians and alleged medical 'experts'. The official Cult narrative is the only show in town.

Genevieve Briand found that deaths from all causes were not exceptional in 2020 compared with previous years and a Spanish magazine published figures that said the same about Spain which was a 'Covid' propaganda hotspot at one point. *Discovery Salud*, a health and medicine magazine, quoted government figures which showed how 17,000 *fewer* people died in Spain in 2020 than in 2019 and more than 26,000 fewer than in 2018. The age-standardised mortality rate for England and Wales when age distribution is taken into account was significantly lower in 2020 than the 1970s, 80s and 90s, and was only the ninth highest since 2000. Where is the 'pandemic'?

Post mortems and autopsies virtually disappeared for 'Covid' deaths amid claims that 'virus-infected' bodily fluids posed a risk to those carrying out the autopsy. This was rejected by renowned German pathologist and forensic doctor Klaus Püschel who said that he and his staff had by then done 150 autopsies on 'Covid' patients with no problems at all. He said they were needed to know why some 'Covid' patients suffered blood clots and not severe respiratory infections. The 'virus' is, after all, called SARS or 'severe acute respiratory syndrome'. I highlighted in the spring of 2020 this phenomenon and quoted New York intensive care doctor Cameron Kyle-Sidell who posted a soon deleted YouTube video to say that they had been told to prepare to treat an infectious disease called 'Covid-19', but that was not what they were dealing with. Instead he likened the lung condition of the most severely ill patients to what you would expect with cabin depressurisation in a plane at 30,000 feet or someone dropped on the top of Everest without oxygen or acclimatisation. I have never said this is not happening to a small minority of alleged 'Covid' patients – I am saying this is not caused by a phantom 'contagious virus'. Indeed Kyle-Sidell said that 'Covid-19' was not the disease they were told was coming their way. 'We are operating under a medical paradigm that is untrue,' he said, and he believed they were treating the wrong disease: 'These people are being slowly starved of oxygen.' Patients would take off their oxygen masks in a state of fear and stress and while they were blue in the face on the brink of death. They did not look like patients dying of pneumonia. You can see why they don't want autopsies when their virus doesn't exist and there is another condition in some people that they don't wish to be uncovered. I should add here that the 5G system of millimetre waves was being rapidly introduced around the world in 2020 and even more so now as they fire 5G at the Earth from satellites. At 60 gigahertz within the 5G range that frequency interacts with the oxygen molecule and stops people breathing in sufficient oxygen to be absorbed into the bloodstream. They are installing 5G in schools and hospitals. The world is not mad or anything. 5G can cause major changes to the lungs and blood as I detail in *The Answer* and these consequences are labelled 'Covid-19', the alleged symptoms of which can be caused by 5G and other electromagnetic frequencies as cells respond to radiation poisoning.

The 'Covid death' scam

Dr Scott Jensen, a Minnesota state senator and medical doctor, exposed 'Covid' Medicare payment incentives to hospitals and death certificate manipulation. He said he was sent a seven-page document by the US Department of Health 'coaching' him on how to fill out death certificates which had never happened before. The document said that he didn't need to have a laboratory test for 'Covid-19' to put that on the death certificate and that shocked him when death certificates are supposed to be about facts. Jensen described how doctors had been 'encouraged, if not pressured' to make a diagnosis of 'Covid-19' if they thought it was probable or *'presumed'*. No positive test was necessary – not that this would have mattered anyway. He said doctors were told to diagnose 'Covid' by symptoms when these were the same as colds, allergies, other respiratory problems, and certainly with influenza which 'disappeared' in the 'Covid' era. A common sniffle was enough to get the dreaded verdict. Ontario authorities decreed that a single care home resident with *one* symptom from a long list must lead to the isolation of the entire home. Other courageous doctors like Jensen made the same point about death figure manipulation and how deaths by other causes were falling while 'Covid-19 deaths' were rising at the same rate due to re-diagnosis. Their videos rarely survive long on YouTube with its Cult-supporting algorithms courtesy of CEO Susan Wojcicki and her bosses at Google. Figure-tampering was so glaring and ubiquitous that even officials were letting it slip or outright saying it. UK chief scientific adviser Patrick Vallance said on one occasion that 'Covid' on the death certificate doesn't mean 'Covid' was the cause of death (so why the hell is it there?) and we had the rare sight of a BBC reporter telling the truth when she said: 'Someone could be successfully treated for Covid, in say April, discharged, and then in June, get run over by a bus and die … That person would still be counted as a Covid death in England.' Yet the BBC and the rest of the world media went on repeating the case and death figures as if they were real. Illinois Public Health Director Dr Ngozi Ezike revealed the deceit while her bosses must have been clenching their buttocks:

> If you were in a hospice and given a few weeks to live and you were then found to have Covid that would be counted as a Covid death. [There might be] a clear alternate cause, but it is still listed as a Covid death. So everyone listed as a Covid death doesn't mean that was the cause of the death, but that they had Covid at the time of death.

Yes, a 'Covid virus' never shown to exist and tested for with a test not testing for the 'virus'. In the first period of the pandemic hoax through the spring of 2020 the process began of designating almost everything a 'Covid' death and this has continued ever since. I sat in a restaurant one night listening to a loud conversation on the next table where a family was discussing in bewilderment how a relative who had no symptoms of 'Covid', and had died of a long-term problem, could have been diagnosed a death by the 'virus'. I could understand their bewilderment. If they read this book they will know why this medical fraud has been perpetrated the world over.

Some media truth shock

The media ignored the evidence of death certificate fraud until eventually one columnist did speak out when she saw it first-hand. Bel Mooney is a long-time national newspaper journalist in Britain currently working for the *Daily Mail*. Her article on February 19th, 2021, carried this headline: 'My dad Ted passed three Covid tests and died of a chronic illness yet he's officially one of Britain's 120,000 victims of the virus and is far from alone ... so how many more are there?' She told how her 99-year-old father was in a care home with a long-standing chronic obstructive pulmonary disease and vascular dementia. Maybe, but he was still aware enough to tell her from the start that there was no 'virus' and he refused the 'vaccine' for that reason. His death was not unexpected given his chronic health problems and Mooney said she was shocked to find that 'Covid-19' was declared the cause of death on his death certificate. She said this was a 'bizarre and unacceptable untruth' for a man with long-time health problems who had tested negative twice at the home for the 'virus'. I was also shocked by this story although not by what she said. I had been highlighting the death certificate manipulation for ten months. It was the confirmation that a professional full-time journalist only realised this was going on when it affected her directly and neither did she know that whether her dad tested positive or negative was irrelevant with the test not testing for the 'virus'. Where had she been? She said she did not believe in 'conspiracy theories' without knowing I'm sure that this and 'conspiracy theorists' were terms put into widespread circulation by the CIA in the 1960s to discredit those who did not accept the ridiculous official story of the Kennedy assassination. A blanket statement of 'I don't believe in conspiracy theories' is always bizarre. The dictionary definition of the term alone means the world is drowning in conspiracies. What she said was even more daft when her dad had just been affected by the 'Covid' conspiracy. Why else does she think that 'Covid-19' was going on the death certificates of people who died of something else?

To be fair once she saw from personal experience what was happening she didn't mince words. Mooney was called by the care home on the morning of February 9th to be told her father had died in his sleep. When she asked for the official cause of death what came back was 'Covid-19'. Mooney challenged this and was told there had been deaths from Covid on the dementia floor (confirmed by a test not testing for the 'virus') so they considered it 'reasonable to assume'. 'But doctor,' Mooney rightly protested, 'an assumption isn't a diagnosis.' She said she didn't blame the perfectly decent and sympathetic doctor – 'he was just doing his job'. Sorry, but that's *bullshit*. He wasn't doing his job at all. He was putting a false cause of death on the death certificate and that is a criminal offence for which he should be brought to account and the same with the millions of doctors worldwide who have done the same. They were not doing their job they were following orders and that must not wash at new Nuremberg trials any more than it did at the first ones. Mooney's doctor was 'assuming' (presuming) as he was told to, but 'just following orders' makes no difference to his actions. A doctor's job is to serve the patient and the truth, not follow orders, but that's what they have done all over the world and played a central part in making the 'Covid' hoax possible with all its catastrophic consequences for humanity. Shame on them and they must answer for their actions. Mooney said her disquiet worsened when she registered her father's death by telephone and was told by the registrar there had been very many other cases like hers where 'the deceased' had not tested positive for 'Covid' yet it was recorded as the cause of death. The test may not matter, but those involved at their level *think* it matters and it shows a callous disregard for accurate diagnosis. The pressure to do this is coming from the top of the national 'health' pyramids which in turn obey the World Health Organization which obeys Gates and the Cult. Mooney said the registrar agreed that this must distort the national figures adding that 'the strangest thing is that every winter we record countless deaths from flu, and this winter there have been none. Not one!' She asked if the registrar thought deaths from flu were being misdiagnosed and lumped together with 'Covid' deaths. The answer was a 'puzzled yes'. Mooney said that the funeral director said the same about 'Covid' deaths which had nothing to do with 'Covid'. They had lost count of the number of families upset by this and other funeral companies in different countries have had the same experience. Mooney wrote:

> The nightly shroud-waving and shocking close-ups of pain imposed on us by the TV news bewildered and terrified the population into eager compliance with lockdowns. We were invited to 'save the NHS' and to grieve for strangers – the real-life loved ones behind those shocking death counts. Why would the public imagine what I now fear, namely that the way Covid-19 death statistics are compiled might make the numbers seem greater than they are?

Oh, just a little bit – like 100 percent.

Do the maths

Mooney asked why a country would wish to skew its mortality figures by wrongly certifying deaths? What had been going on? Well, if you don't believe in conspiracies you will never find the answer which is that *it's a conspiracy*. She did, however, describe what she had discovered as a 'national scandal'. In reality it's a global scandal and happening everywhere. Pillars of this conspiracy were all put into place before the button was pressed with the Drosten PCR protocol and high amplifications to produce the cases and death certificate changes to secure illusory 'Covid' deaths. Mooney notes that normally two doctors were needed to certify a death, with one having to know the patient, and how the rules were changed in the spring of 2020 to allow one doctor to do this. In the same period 'Covid deaths' were decreed to be all cases where Covid-19 was put on the death certificate even without a positive test or any symptoms. Mooney asked: 'How many of the 30,851 (as of January 15) care home resident deaths with Covid-19 on the certificate (32.4 per cent of all deaths so far) were based on an assumption, like that of my father? And what has that done to our national psyche?'All of them is the answer to the first question and it has devastated and dismantled the national psyche, actually the global psyche, on a colossal scale. In the UK case and death data is compiled by organisations like Public Health England (PHE) and the Office for National Statistics (ONS). Mooney highlights the insane policy of counting a death from any cause as 'Covid-19' if this happens within 28 days of a positive test (with a test not testing for the 'virus') and she points out that ONS statistics reflect deaths 'involving Covid' 'or due to Covid' which meant in practice any death where ‘Covid-19’ was mentioned on the death certificate. She described the consequences of this fraud:

> Most people will accept the narrative they are fed, so panicky governments here and in Europe witnessed the harsh measures enacted in totalitarian China and jumped into lockdown. Headlines about Covid deaths tolled like the knell that would bring doomsday to us all. Fear stalked our empty streets. Politicians parroted the frankly ridiculous aim of ‘zero Covid’ and shut down the economy, while most British people agreed that lockdown was essential and (astonishingly to me, as a patriotic Brit) even wanted more restrictions.
>
> For what? Lies on death certificates? Never mind the grim toll of lives ruined, suicides, schools closed, rising inequality, depression, cancelled hospital treatments, cancer patients in a torture of waiting, poverty, economic devastation, loneliness, families kept apart, and so on. How many lives have been lost as a direct result of lockdown?

She said that we could join in a national chorus of shock and horror at reaching the 120,000 death toll which was surely certain to have been totally skewed all along, but what about the human cost of lockdown justified by these ‘death figures’? *The British Medical Journal* had reported a 1,493 percent increase in cases of children taken to Great Ormond Street Hospital with abusive head injuries alone and then there was the effect on families:

> Perhaps the most shocking thing about all this is that families have been kept apart – and obeyed the most irrational, changing rules at the whim of government – because they believed in the statistics. They succumbed to fear, which his generation rejected in that war fought for freedom. Dad (God rest his soul) would be angry. And so am I.

Another theme to watch is that in the winter months when there are more deaths from all causes they focus on ‘Covid’ deaths and in the summer when the British Lung Foundation says respiratory disease plummets by 80 percent they rage on about ‘cases’. Either way fascism on population is always the answer.

Nazi eugenics in the 21st century

Elderly people in care homes have been isolated from their families month after lonely month with no contact with relatives and grandchildren who were banned from seeing them. We were told that lockdown fascism was to ‘protect the vulnerable’ like elderly people. At the same time Do Not Resuscitate (DNR) orders were placed on their medical files so that if they needed resuscitation it wasn’t done and ‘Covid-19’ went on their death certificates. Old people were not being ‘protected’ they were being culled – murdered in truth. DNR orders were being decreed for disabled and young people with learning difficulties or psychological problems. The UK Care Quality Commission, a non-departmental body of the Department of Health and Social Care, found that 34 percent of those working in health and social care were pressured into placing ‘do not attempt cardiopulmonary resuscitation’ orders on ‘Covid’ patients who suffered from disabilities and learning difficulties without involving the patient or their families in the decision. UK judges ruled that an elderly woman with dementia should have the DNA-manipulating ‘Covid vaccine’ against her son’s wishes and that a man with severe learning difficulties should have the jab despite his family’s objections. Never mind that many had already died. The judiciary always supports doctors and government in fascist dictatorships. They wouldn’t dare do otherwise. A horrific video was posted showing fascist officers from Los Angeles police forcibly giving the ‘Covid’ shot to women with special needs who were screaming that they didn’t want it. The same fascists are seen giving the jab to a sleeping elderly woman in a care home. This is straight out of the Nazi playbook. Hitler’s Nazis committed mass murder of the mentally ill and physically disabled throughout Germany and occupied territories in the programme that became known as Aktion T4, or just T4. Sabbatian-controlled Hitler and his grotesque crazies set out to kill those they considered useless and unnecessary. The Reich Committee for the Scientific Registering of Hereditary and Congenital Illnesses registered the births of babies identified by physicians to have ‘defects’. By 1941 alone more than 5,000 children were murdered by the state and it is estimated that in total the number of innocent people killed in Aktion T4 was between 275,000 and 300,000. Parents were told their children had been sent away for ‘special treatment’ never to return. It is rather pathetic to see claims about plans for new extermination camps being dismissed today when the same force behind current events did precisely that 80 years ago. Margaret Sanger was a Cult operative who used 'birth control' to sanitise her programme of eugenics. Organisations she founded became what is now Planned Parenthood. Sanger proposed that 'the whole dysgenic population would have its choice of segregation or sterilization'. These included epileptics, 'feeble-minded', and prostitutes. Sanger opposed charity because it perpetuated 'human waste'. She reveals the Cult mentality and if anyone thinks that extermination camps are a 'conspiracy theory' their naivety is touching if breathtakingly stupid.

If you don't believe that doctors can act with callous disregard for their patients it is worth considering that doctors and medical staff agreed to put government-decreed DNR orders on medical files and do nothing when resuscitation is called for. I don't know what you call such people in your house. In mine they are Nazis from the Josef Mengele School of Medicine. Phenomenal numbers of old people have died worldwide from the effects of lockdown, depression, lack of treatment, the 'vaccine' (more later) and losing the will to live. A common response at the start of the manufactured pandemic was to remove old people from hospital beds and transfer them to nursing homes. The decision would result in a mass cull of elderly people in those homes through lack of treatment – *not* 'Covid'. Care home whistleblowers have told how once the 'Covid' era began doctors would not come to their homes to treat patients and they were begging for drugs like antibiotics that often never came. The most infamous example was ordered by New York governor Andrew Cuomo, brother of a moronic CNN host, who amazingly was given an Emmy Award for his handling of the 'Covid crisis' by the ridiculous Wokers that hand them out. Just how ridiculous could be seen in February, 2021, when a Department of Justice and FBI investigation began into how thousands of old people in New York died in nursing homes after being discharged from hospital to make way for 'Covid' patients on Cuomo's say-so – and how he and his staff covered up these facts. This couldn't have happened to a nicer psychopath. Even then there was a 'Covid' spin. Reports said that thousands of old people who tested positive for 'Covid' in hospital were transferred to nursing homes to both die of 'Covid' and transmit it to others. No – they were in hospital because they were ill and the fact that they tested positive with a test not testing for the 'virus' is irrelevant. They were ill often with respiratory diseases ubiquitous in old people near the end of their lives. Their transfer out of hospital meant that their treatment stopped and many would go on to die.

They're old. Who gives a damn?

I have exposed in the books for decades the Cult plan to cull the world's old people and even to introduce at some point what they call a 'demise pill' which at a certain age everyone would take and be out of here by law. In March, 2021, Spain legalised euthanasia and assisted suicide following the Netherlands, Belgium, Luxembourg and Canada on the Tiptoe to the demise pill. Treatment of old people by many 'care' homes has been a disgrace in the 'Covid' era. There are many, many, caring staff – I know some. There have, however, been legions of stories about callous treatment of old people and their families. Police were called when families came to take their loved ones home in the light of isolation that was killing them. They became prisoners of the state. Care home residents in insane, fascist Ontario, Canada, were not allowed to leave their *room* once the 'Covid' hoax began. UK staff have even wheeled elderly people away from windows where family members were talking with them. Oriana Criscuolo from Stockport in the English North West dropped off some things for her 80-year-old father who has Parkinson's disease and dementia and she wanted to wave to him through a ground-floor window. She was told that was 'illegal'. When she went anyway they closed the curtains in the middle of the day. Oriana said:

> It's just unbelievable. I cannot understand how care home staff – people who are being paid to care – have become so uncaring. Their behaviour is inhumane and cruel. It's beyond belief.

She was right and this was not a one-off. What a way to end your life in such loveless circumstances. UK registered nurse Nicky Millen, a proper old school nurse for 40 years, said that when she started her career care was based on dignity, choice, compassion and empathy. Now she said 'the things that are important to me have gone out of the window.' She was appalled that people were dying without their loved ones and saying goodbye on iPads. Nicky described how a distressed 89-year-old lady stroked her face and asked her 'how many paracetamol would it take to finish me off'. Life was no longer worth living while not seeing her family. Nicky said she was humiliated in front of the ward staff and patients for letting the lady stroke her face and giving her a cuddle. Such is the dehumanisation that the 'Covid' hoax has brought to the surface. Nicky worked in care homes where patients told her they were being held prisoner. 'I want to live until I die', one said to her. 'I had a lady in tears because she hadn't seen her great-grandson.' Nicky was compassionate old school meeting psychopathic New Normal. She also said she had worked on a 'Covid' ward with no 'Covid' patients. Jewish writer Shai Held wrote an article in March, 2020, which was headlined 'The Staggering, Heartless Cruelty Toward the Elderly'. What he described was happening from the earliest days of lockdown. He said 'the elderly' were considered a group and not unique individuals (the way of the Woke). Shai Held said:

> Notice how the all-too-familiar rhetoric of dehumanization works: 'The elderly' are bunched together as a faceless mass, all of them considered culprits and thus effectively deserving of the suffering the pandemic will inflict upon them. Lost entirely is the fact that the elderly are individual human beings, each with a distinctive face and voice, each with hopes and dreams, memories and regrets, friendships and marriages, loves lost and loves sustained.

'The elderly' have become another dehumanised group for which anything goes and for many that has resulted in cold disregard for their rights and their life. The distinctive face that Held talks about is designed to be deleted by masks until everyone is part of a faceless mass.

'War-zone' hospitals myth

Again and again medical professionals have told me what was really going on and how hospitals 'overrun like war zones' according to the media were virtually empty. The mantra from medical whistleblowers was please don't use my name or my career is over. Citizen journalists around the world sneaked into hospitals to film evidence exposing the 'war-zone' lie. They really *were* largely empty with closed wards and operating theatres. I met a hospital worker in my town on the Isle of Wight during the first lockdown in 2020 who said the only island hospital had never been so quiet. Lockdown was justified by the psychopaths to stop hospitals being overrun. At the same time that the island hospital was near-empty the military arrived here to provide *extra beds*. It was all propaganda to ramp up the fear to ensure compliance with fascism as were never-used temporary hospitals with thousands of beds known as Nightingales and never-used make-shift mortuaries opened by the criminal UK government. A man who helped to install those extra island beds attributed to the army said they were never used and the hospital was empty. Doctors and nurses 'stood around talking or on their phones, wandering down to us to see what we were doing'. There were no masks or social distancing. He accused the useless local island paper, the *County Press*, of 'pumping the fear as if our hospital was overrun and we only have one so it should have been'. He described ambulances parked up with crews outside in deck chairs. When his brother called an ambulance he was told there was a two-hour backlog which he called 'bullshit'. An old lady on the island fell 'and was in a bad way', but a caller who rang for an ambulance was told the situation wasn't urgent enough. Ambulance stations were working under capacity while people would hear ambulances with sirens blaring driving through the streets. When those living near the stations realised what was going on they would follow them as they left, circulated around an urban area with the sirens going, and then came back without stopping. All this was to increase levels of fear and the same goes for the 'ventilator shortage crisis' that cost tens of millions for hastily produced ventilators never to be used.

Ambulance crews that agreed to be exploited in this way for fear propaganda might find themselves a mirror. I wish them well with that. Empty hospitals were the obvious consequence of treatment and diagnoses of non-'Covid' conditions cancelled and those involved handed a death sentence. People have been dying at home from undiagnosed and untreated cancer, heart disease and other life-threatening conditions to allow empty hospitals to deal with a 'pandemic' that wasn't happening.

Death of the innocent

'War-zones' have been laying off nursing staff, even doctors where they can. There was no work for them. Lockdown was justified by saving lives and protecting the vulnerable they were actually killing with DNR orders and preventing empty hospitals being 'overrun'. In Britain the mantra of stay at home to 'save the NHS' was everywhere and across the world the same story was being sold when it was all lies. Two California doctors, Dan Erickson and Artin Massihi at Accelerated Urgent Care in Bakersfield, held a news conference in April, 2020, to say that intensive care units in California were 'empty, essentially', with hospitals shutting floors, not treating patients and laying off doctors. The California health system was working at minimum capacity 'getting rid of doctors because we just don't have the volume'. They said that people with conditions such as heart disease and cancer were not coming to hospital out of fear of 'Covid-19'. Their video was deleted by Susan Wojcicki's Cult-owned YouTube after reaching five million views. Florida governor Ron Desantis, who rejected the severe lockdowns of other states and is being targeted for doing so, said that in March, 2020, every US governor was given models claiming they would run out of hospital beds in days. That was never going to happen and the 'modellers' knew it. Deceit can be found at every level of the system. Urgent children's operations were cancelled including fracture repairs and biopsies to spot cancer. Eric Nicholls, a consultant paediatrician, said 'this is obviously concerning and we need to return to normal operating and to increase capacity as soon as possible'. Psychopaths in power were rather less concerned *because* they are psychopaths. Deletion of urgent care and diagnosis has been happening all over the world and how many kids and others have died as a result of the actions of these cold and heartless lunatics dictating 'health' policy? The number must be stratospheric. Richard Sullivan, professor of cancer and global health at King's College London, said people feared 'Covid' more than cancer such was the campaign of fear. 'Years of lost life will be quite dramatic', Sullivan said, with 'a huge amount of avoidable mortality'. Sarah Woolnough, executive director for policy at Cancer Research UK, said there had been a 75 percent drop in urgent referrals to hospitals by family doctors of people with suspected cancer. Sullivan said that 'a lot of services have had to scale back – we've seen a dramatic decrease in the amount of elective cancer surgery'. Lockdown deaths worldwide has been absolutely fantastic with the *New York Post* reporting how data confirmed that 'lockdowns end more lives than they save':

> There was a sharp decline in visits to emergency rooms and an increase in fatal heart attacks because patients didn't receive prompt treatment. Many fewer people were screened for cancer. Social isolation contributed to excess deaths from dementia and Alzheimer's.
>
> Researchers predicted that the social and economic upheaval would lead to tens of thousands of "deaths of despair" from drug overdoses, alcoholism and suicide. As unemployment surged and mental-health and substance-abuse treatment programs were interrupted, the reported levels of anxiety, depression and suicidal thoughts increased dramatically, as did alcohol sales and fatal drug overdoses.

This has been happening while nurses and other staff had so much time on their hands in the 'war-zones' that Tic-Tok dancing videos began appearing across the Internet with medical staff dancing around in empty wards and corridors as people died at home from causes that would normally have been treated in hospital.

Mentions in dispatches

One brave and truth-committed whistleblower was Louise Hampton, a call handler with the UK NHS who made a viral Internet video saying she had done 'fuck all' during the 'pandemic' which was ‘a load of bollocks’. She said that ‘Covid-19’ was rebranded flu and of course she lost her job. This is what happens in the medical and endless other professions now when you tell the truth. Louise filmed inside ‘war-zone’ accident and emergency departments to show they were empty and I mean *empty* as in no one there. The mainstream media could have done the same and blown the gaff on the whole conspiracy. They haven’t to their eternal shame. Not that most ‘journalists’ seem capable of manifesting shame as with the psychopaths they slavishly repeat without question. The relative few who were admitted with serious health problems were left to die alone with no loved ones allowed to see them because of ‘Covid’ rules and they included kids dying without the comfort of mum and dad at their bedside while the evil behind this couldn’t give a damn. It was all good fun to them. A Scottish NHS staff nurse publicly quit in the spring of 2021 saying: ‘I can no longer be part of the lies and the corruption by the government.’ She said hospitals ‘aren’t full, the beds aren’t full, beds have been shut, wards have been shut’. Hospitals were never busy throughout ‘Covid’. The staff nurse said that Nicola Sturgeon, tragically the leader of the Scottish government, was on television saying save the hospitals and the NHS – ‘but the beds are empty’ and ‘we’ve not seen flu, we always see flu every year’. She wrote to government and spoke with her union Unison (the unions are Cult-compromised and *useless*, but nothing changed. Many of her colleagues were scared of losing their jobs if they spoke out as they wanted to. She said nursing staff were being affected by wearing masks all day and ‘my head is splitting every shift from wearing a mask’. The NHS is part of the fascist tyranny and must be dismantled so we can start again with human beings in charge. (Ironically, hospitals were reported to be busier again when official ‘Covid’ cases *fell* in spring/summer of 2021 and many other conditions required treatment at the same time as *the fake vaccine rollout*.)

I will cover the ‘Covid vaccine’ scam in detail later, but it is another indicator of the sickening disregard for human life that I am highlighting here. The DNA-manipulating concoctions do not fulfil the definition of a ‘vaccine’, have never been used on humans before and were given only emergency approval because trials were not completed and they continued using the unknowing public. The result was what a NHS senior nurse with responsibility for ‘vaccine’ procedure said was ‘genocide’. She said the ‘vaccines’ were not ‘vaccines’. They had not been shown to be safe and claims about their effectiveness by drug companies were ‘poetic licence’. She described what was happening as a ‘horrid act of human annihilation’. The nurse said that management had instigated a policy of not providing a Patient Information Leaflet (PIL) before people were ‘vaccinated’ even though health care professionals are supposed to do this according to protocol. Patients should also be told that they are taking part in an ongoing clinical trial. Her challenges to what is happening had seen her excluded from meetings and ridiculed in others. She said she was told to ‘watch my step … or I would find myself surplus to requirements’. The nurse, who spoke anonymously in fear of her career, said she asked her NHS manager why he/she was content with taking part in genocide against those having the ‘vaccines’. The reply was that everyone had to play their part and to ‘put up, shut up, and get it done’. Government was ‘leaning heavily’ on NHS management which was clearly leaning heavily on staff. This is how the global ‘medical’ hierarchy operates and it starts with the Cult and its World Health Organization.

She told the story of a doctor who had the Pfizer jab and when questioned had no idea what was in it. The doctor had never read the literature. We have to stop treating doctors as intellectual giants when so many are moral and medical pygmies. The doctor did not even know that the ‘vaccines’ were not fully approved or that their trials were ongoing. They were, however, asking their patients if they minded taking part in follow-ups for research purposes – yes, the *ongoing clinical trial*. The nurse said the doctor’s ignorance was not rare and she had spoken to a hospital consultant who had the jab without any idea of the background or that the ‘trials’ had not been completed. Nurses and pharmacists had shown the same ignorance.

‘My NHS colleagues have forsaken their duty of care, broken their code of conduct – Hippocratic Oath – and have been brainwashed just the same as the majority of the UK public through propaganda …’ She said she had not been able to recruit a single NHS colleague, doctor, nurse or pharmacist to stand with her and speak out. Her union had refused to help. She said that if the genocide came to light she would not hesitate to give evidence at a Nuremberg-type trial against those in power who could have affected the outcomes but didn’t.

And all for what?

To put the nonsense into perspective let’s say the ‘virus’ does exist and let’s go completely crazy and accept that the official manipulated figures for cases and deaths are accurate. *Even then* a study by Stanford University epidemiologist Dr John Ioannidis published on the World Health Organization website produced an average infection to fatality rate of … *0.23 percent!* Ioannidis said: ‘If one could sample equally from all locations globally, the median infection fatality rate might even be substantially lower than the 0.23% observed in my analysis.’ For healthy people under 70 it was … *0.05 percent*! This compares with the 3.4 percent claimed by the Cult-owned World Health Organization when the hoax was first played and maximum fear needed to be generated. An updated Stanford study in April, 2021, put the ‘infection’ to ‘fatality’ rate at just 0.15 percent. Another team of scientists led by Megan O’Driscoll and Henrik Salje studied data from 45 countries and published their findings on the Nature website. For children and young people the figure is so small it virtually does not register although authorities will be hyping dangers to the young when they introduce DNA-manipulating ‘vaccines’ for children. The O’Driscoll study produced an average infection-fatality figure of 0.003 for children from birth to four; 0.001 for 5 to 14; 0.003 for 15 to 19; and it was still only 0.456 up to 64. To claim that children must be ‘vaccinated’ to protect them from ‘Covid’ is an obvious lie and so there must be another reason and there is. What’s more the average age of a ‘Covid’ death is akin to the average age that people die in general. The average age of death in England is about 80 for men and 83 for women. The average age of death from alleged 'Covid' is between 82 and 83. California doctors, Dan Erickson and Artin Massihi, said at their April media conference that projection models of millions of deaths had been 'woefully inaccurate'. They produced detailed figures showing that Californians had a 0.03 chance of dying from 'Covid' based on the number of people who tested positive (with a test not testing for the 'virus'). Erickson said there was a 0.1 percent chance of dying from 'Covid' in the *state* of New York, not just the city, and a 0.05 percent chance in Spain, a centre of 'Covid-19' hysteria at one stage. The Stanford studies supported the doctors' data with fatality rate estimates of 0.23 and 0.15 percent. How close are these figures to my estimate of *zero*? Death-rate figures claimed by the World Health Organization at the start of the hoax were some 15 times higher. The California doctors said there was no justification for lockdowns and the economic devastation they caused. Everything they had ever learned about quarantine was that you quarantine the *sick* and not the healthy. They had never seen this before and it made no medical sense.

Why in the in the light of all this would governments and medical systems the world over say that billions must go under house arrest; lose their livelihood; in many cases lose their mind, their health and their life; force people to wear masks dangerous to health and psychology; make human interaction and even family interaction a criminal offence; ban travel; close restaurants, bars, watching live sport, concerts, theatre, and any activity involving human togetherness and discourse; and closing schools to isolate children from their friends and cause many to commit suicide in acts of hopelessness and despair? The California doctors said lockdown consequences included increased child abuse, partner abuse, alcoholism, depression, and other impacts they were seeing every day. Who would do that to the entire human race if not mentally-ill psychopaths of almost unimaginable extremes like Bill Gates? We must face the reality of what we are dealing with and come out of denial. Fascism and tyranny are made possible only by the target population submitting and acquiescing to fascism and tyranny. The whole of human history shows that to be true. Most people naively and unquestioning believed what they were told about a 'deadly virus' and meekly and weakly submitted to house arrest. Those who didn't believe it – at least in total – still submitted in fear of the consequences of not doing so. For the rest who wouldn't submit draconian fines have been imposed, brutal policing by psychopaths *for* psychopaths, and condemnation from the meek and weak who condemn the Pushbackers on behalf of the very force that has them, too, in its gunsights. 'Pathetic' does not even begin to suffice. Britain's brainless 'Health' Secretary Matt Hancock warned anyone lying to border officials about returning from a list of 'hotspot' countries could face a jail sentence of up to ten years which is more than for racially-aggravated assault, incest and attempting to have sex with a child under 13. Hancock is a lunatic, but he has the state apparatus behind him in a Cult-led chain reaction and the same with UK 'Vaccine Minister' Nadhim Zahawi, a prominent member of the mega-Cult secret society, Le Cercle, which featured in my earlier books. The Cult enforces its will on governments and medical systems; government and medical systems enforce their will on business and police; business enforces its will on staff who enforce it on customers; police enforce the will of the Cult on the population and play their essential part in creating a world of fascist control that their own children and grandchildren will have to live in their entire lives. It is a hierarchical pyramid of imposition and acquiescence and, yes indeedy, of clinical insanity.

Does anyone bright enough to read this book have to ask what the answer is? I think not, but I will reveal it anyway in the fewest of syllables: Tell the psychos and their moronic lackeys to fuck off and let's get on with our lives. We are many – They are few.

CHAPTER SEVEN

War on your mind

One believes things because one has been conditioned to believe them
Aldous Huxley, *Brave New World*

I have described the 'Covid' hoax as a 'Psyop' and that is true in every sense and on every level in accordance with the definition of that term which is psychological warfare. Break down the 'Covid pandemic' to the foundation themes and it is psychological warfare on the human individual and collective mind.

The same can be said for the entire human belief system involving every subject you can imagine. Huxley was right in his contention that people believe what they are conditioned to believe and this comes from the repetition throughout their lives of the same falsehoods. They spew from government, corporations, media and endless streams of 'experts' telling you what the Cult wants you to believe and often believing it themselves (although *far* from always). 'Experts' are rewarded with 'prestigious' jobs and titles and as agents of perceptual programming with regular access to the media. The Cult has to control the narrative – control *information* – or they lose control of the vital, crucial, without-which-they-cannot-prevail public perception of reality. The foundation of that control today is the Internet made possible by the Defense Advanced Research Projects Agency (DARPA), the incredibly sinister technological arm of the Pentagon. The Internet is the result of military technology.

DARPA openly brags about establishing the Internet which has been a long-term project to lasso the minds of the global population. I have said for decades the plan is to control information to such an extreme that eventually no one would see or hear anything that the Cult does not approve. We are closing in on that end with ferocious censorship since the 'Covid' hoax began and in my case it started back in the 1990s in terms of books and speaking venues. I had to create my own publishing company in 1995 precisely because no one else would publish my books even then. I think they're all still running.

Cult Internet

To secure total control of information they needed the Internet in which pre-programmed algorithms can seek out 'unclean' content for deletion and even stop it being posted in the first place. The Cult had to dismantle print and non-Internet broadcast media to ensure the transfer of information to the appropriate-named 'Web' – a critical expression of the *Cult* web. We've seen the ever-quickening demise of traditional media and control of what is left by a tiny number of corporations operating worldwide. Independent journalism in the mainstream is already dead and never was that more obvious than since the turn of 2020. The Cult wants all information communicated via the Internet to globally censor and allow the plug to be pulled any time. Lockdowns and forced isolation has meant that communication between people has been through electronic means and no longer through face-to-face discourse and discussion. Cult psychopaths have targeted the bars, restaurants, sport, venues and meeting places in general for this reason. None of this is by chance and it's to stop people gathering in any kind of privacy or number while being able to track and monitor all Internet communications and block them as necessary. Even private messages between individuals have been censored by these fascists that control Cult fronts like Facebook, Twitter, Google and YouTube which are all officially run by Sabbatian place-people and from the background by higher-level Sabbatian place people.

Facebook, Google, Amazon and their like were seed-funded and supported into existence with money-no-object infusions of funds either directly or indirectly from DARPA and CIA technology arm In-Q-Tel. The Cult plays the long game and prepares very carefully for big plays like 'Covid'. Amazon is another front in the psychological war and pretty much controls the global market in book sales and increasingly publishing. Amazon's limitless funds have deleted fantastic numbers of independent publishers to seize global domination on the way to deciding which books can be sold and circulated and which cannot. Moves in that direction are already happening. Amazon's leading light Jeff Bezos is the grandson of Lawrence Preston Gise who worked with DARPA predecessor ARPA. Amazon has big connections to the CIA and the Pentagon. The plan I have long described went like this:

1. Employ military technology to establish the Internet.

2. Sell the Internet as a place where people can freely communicate without censorship and allow that to happen until the Net becomes the central and irreversible pillar of human society. If the Internet had been highly censored from the start many would have rejected it.

3. Fund and manipulate major corporations into being to control the circulation of information on your Internet using cover stories about geeks in garages to explain how they came about. Give them unlimited funds to expand rapidly with no need to make a profit for years while non-Cult companies who need to balance the books cannot compete. You know that in these circumstances your Googles, YouTubes, Facebooks and Amazons are going to secure near monopolies by either crushing or buying up the opposition.

4. Allow freedom of expression on both the Internet and communication platforms to draw people in until the Internet is the central and irreversible pillar of human society and your communication corporations have reached a stage of near monopoly domination.

5. Then unleash your always-planned frenzy of censorship on the basis of 'where else are you going to go?' and continue to expand that until nothing remains that the Cult does not want its human targets to see.

The process was timed to hit the 'Covid' hoax to ensure the best chance possible of controlling the narrative which they knew they had to do at all costs. They were, after all, about to unleash a 'deadly virus' that didn't really exist. If you do that in an environment of free-flowing information and opinion you would be dead in the water before you could say Gates is a psychopath. The network was in place through which the Cult-created-and-owned World Health Organization could dictate the 'Covid' narrative and response policy slavishly supported by Cult-owned Internet communication giants and mainstream media while those telling a different story were censored. Google, YouTube, Facebook and Twitter openly announced that they would do this. What else would we expect from Cult-owned operations like Facebook which former executives have confirmed set out to make the platform more addictive than cigarettes and coldly manipulates emotions of its users to sow division between people and groups and scramble the minds of the young? If Zuckerberg lives out the rest of his life without going to jail for crimes against humanity, and most emphatically against the young, it will be a travesty of justice. Still, no matter, cause and effect will catch up with him eventually and the same with Sergey Brin and Larry Page at Google with its CEO Sundar Pichai who fix the Google search results to promote Cult narratives and hide the opposition. Put the same key words into Google and other search engines like DuckDuckGo and you will see how different results can be. Wikipedia is another intensely biased 'encyclopaedia' which skews its content to the Cult agenda. YouTube links to Wikipedia's version of 'Covid' and 'climate change' on video pages in which experts in their field offer a different opinion (even that is increasingly rare with Wojcicki censorship). Into this 'Covid' silence-them network must be added government media censors, sorry 'regulators', such as Ofcom in the UK which imposed tyrannical restrictions on British broadcasters that had the effect of banning me from ever appearing. Just to debate with me about my evidence and views on 'Covid' would mean breaking the fascistic impositions of Ofcom and its CEO career government bureaucrat Melanie Dawes. Gutless British broadcasters tremble at the very thought of fascist Ofcom.

Psychos behind 'Covid'

The reason for the ‘Covid’ catastrophe in all its facets and forms can be seen by whom and what is driving the policies worldwide in such a coordinated way. Decisions are not being made to protect health, but to target psychology. The dominant group guiding and ‘advising’ government policy are not medical professionals. They are psychologists and behavioural scientists. Every major country has its own version of this phenomenon and I’ll use the British example to show how it works. In many ways the British version has been affecting the wider world in the form of the huge behaviour manipulation network in the UK which operates in other countries. The network involves private companies, government, intelligence and military. The Cabinet Office is at the centre of the government ‘Covid’ Psyop and part-owns, with ‘innovation charity’ Nesta, the Behavioural Insights Team (BIT) which claims to be independent of government but patently isn’t. The BIT was established in 2010 and its job is to manipulate the psyche of the population to acquiesce to government demands and so much more. It is also known as the ‘Nudge Unit’, a name inspired by the 2009 book by two ultra-Zionists, Cass Sunstein and Richard Thaler, called *Nudge: Improving Decisions About Health, Wealth, and Happiness*. The book, as with the Behavioural Insights Team, seeks to ‘nudge’ behaviour (manipulate it) to make the public follow patterns of action and perception that suit those in authority (the Cult). Sunstein is so skilled at this that he advises the World Health Organization and the UK Behavioural Insights Team and was Administrator of the White House Office of Information and Regulatory Affairs in the Obama administration. Biden appointed him to the Department of Homeland Security – another ultra-Zionist in the fold to oversee new immigration laws which is another policy the Cult wants to control. Sunstein is desperate to silence anyone exposing conspiracies and co-authored a 2008 report on the subject in which suggestions were offered to ban ‘conspiracy theorizing’ or impose ‘some kind of tax, financial or otherwise, on those who disseminate such theories’. I guess a psychiatrist’s chair is out of the question?

Sunstein's mate Richard Thaler, an 'academic affiliate' of the UK Behavioural Insights Team, is a proponent of 'behavioural economics' which is defined as the study of 'the effects of psychological, cognitive, emotional, cultural and social factors on the decisions of individuals and institutions'. Study the effects so they can be manipulated to be what you want them to be. Other leading names in the development of behavioural economics are ultra-Zionists Daniel Kahneman and Robert J. Shiller and they, with Thaler, won the Nobel Memorial Prize in Economic Sciences for their work in this field. The Behavioural Insights Team is operating at the heart of the UK government and has expanded globally through partnerships with several universities including Harvard, Oxford, Cambridge, University College London (UCL) and Pennsylvania. They claim to have 'trained' (reframed) 20,000 civil servants and run more than 750 projects involving 400 randomised controlled trials in dozens of countries' as another version of mind reframers Common Purpose. BIT works from its office in New York with cities and their agencies, as well as other partners, across the United States and Canada – this is a company part-owned by the British government Cabinet Office. An executive order by President Cult-servant Obama established a US Social and Behavioral Sciences Team in 2015. They all have the same reason for being and that's to brainwash the population directly and by brainwashing those in positions of authority.

'Covid' mind game

Another prime aspect of the UK mind-control network is the 'independent' [joke] Scientific Pandemic Insights Group on Behaviours (SPI-B) which 'provides behavioural science advice aimed at anticipating and helping people adhere to interventions that are recommended by medical or epidemiological experts'. That means manipulating public perception and behaviour to do whatever government tells them to do. It's disgusting and if they really want the public to be 'safe' this lot should all be under lock and key. According to the government website SPI-B consists of ‘behavioural scientists, health and social psychologists, anthropologists and historians’ and advises the Whitty-Vallance-led Scientific Advisory Group for Emergencies (SAGE) which in turn advises the government on ‘the science’ (it doesn’t) and ‘Covid’ policy. When politicians say they are being guided by ‘the science’ this is the rabble in each country they are talking about and that ‘science’ is dominated by behaviour manipulators to enforce government fascism through public compliance. The Behaviour Insight Team is headed by psychologist David Solomon Halpern, a visiting professor at King’s College London, and connects with a national and global web of other civilian and military organisations as the Cult moves towards its goal of fusing them into one fascistic whole in every country through its ‘Fusion Doctrine’. The behaviour manipulation network involves, but is not confined to, the Foreign Office; National Security Council; government communications headquarters (GCHQ); MI5; MI6; the Cabinet Office-based Media Monitoring Unit; and the Rapid Response Unit which ‘monitors digital trends to spot emerging issues; including misinformation and disinformation; and identifies the best way to respond’.

There is also the 77th Brigade of the UK military which operates like the notorious Israeli military’s Unit 8200 in manipulating information and discussion on the Internet by posing as members of the public to promote the narrative and discredit those who challenge it. Here we have the military seeking to manipulate *domestic* public opinion while the Nazis in government are fine with that. Conservative Member of Parliament Tobias Ellwood, an advocate of lockdown and control through ‘vaccine passports’, is a Lieutenant Colonel reservist in the 77th Brigade which connects with the military operation jHub, the ‘innovation centre’ for the Ministry of Defence and Strategic Command. jHub has also been involved with the civilian National Health Service (NHS) in ‘symptom tracing’ the population. The NHS is a key part of this mind control network and produced a document in December, 2020, explaining to staff how to use psychological manipulation with different groups and ages to get them to have the DNA-manipulating ‘Covid vaccine’ that's designed to cumulatively rewrite human genetics. The document, called 'Optimising Vaccination Roll Out – Do's and Dont's for all messaging, documents and "communications" in the widest sense', was published by NHS England and the NHS Improvement *Behaviour Change Unit* in partnership with Public Health England and Warwick Business School. I hear the mantra about 'save the NHS' and 'protect the NHS' when we need to scrap the NHS and start again. The current version is far too corrupt, far too anti-human and totally compromised by Cult operatives and their assets. UK government broadcast media censor Ofcom will connect into this web – as will the BBC with its tremendous Ofcom influence – to control what the public see and hear and dictate mass perception. Nuremberg trials must include personnel from all these organisations.

The fear factor

The 'Covid' hoax has led to the creation of the UK Cabinet Office-connected Joint Biosecurity Centre (JBC) which is officially described as providing 'expert advice on pandemics' using its independent [all Cult operations are 'independent'] analytical function to provide real-time analysis about infection outbreaks to identify and respond to outbreaks of Covid-19'. Another role is to advise the government on a response to spikes in infections – 'for example by closing schools or workplaces in local areas where infection levels have risen'. Put another way, promoting the Cult agenda. The Joint Biosecurity Centre is modelled on the Joint Terrorism Analysis Centre which analyses intelligence to set 'terrorism threat levels' and here again you see the fusion of civilian and military operations and intelligence that has led to military intelligence producing documents about 'vaccine hesitancy' and how it can be combated. Domestic civilian matters and opinions should not be the business of the military. The Joint Biosecurity Centre is headed by Tom Hurd, director general of the Office for Security and Counter-Terrorism from the establishment-to-its-fingertips Hurd family. His father is former Foreign Secretary Douglas Hurd. How coincidental that Tom

Hurd went to the elite Eton College and Oxford University with Boris Johnson. Imperial College with its ridiculous computer modeller Neil Ferguson will connect with this gigantic web that will itself interconnect with similar set-ups in other major and not so major countries. Compared with this Cult network the politicians, be they Boris Johnson, Donald Trump or Joe Biden, are bit-part players 'following the science'. The network of psychologists was on the 'Covid' case from the start with the aim of generating maximum fear of the 'virus' to ensure compliance by the population. A government behavioural science group known as SPI-B produced a paper in March, 2020, for discussion by the main government science advisory group known as SAGE. It was headed 'Options for increasing adherence to social distancing measures' and it said the following in a section headed 'Persuasion':

- A substantial number of people still do not feel sufficiently personally threatened; it could be that they are reassured by the low death rate in their demographic group, although levels of concern may be rising. Having a good understanding of the risk has been found to be positively associated with adoption of COVID-19 social distancing measures in Hong Kong.

- The perceived level of personal threat needs to be increased among those who are complacent, using hard-hitting evaluation of options for increasing social distancing emotional messaging. To be effective this must also empower people by making clear the actions they can take to reduce the threat.

- Responsibility to others: There seems to be insufficient understanding of, or feelings of responsibility about, people's role in transmitting the infection to others … Messaging about actions need to be framed positively in terms of protecting oneself and the community, and increase confidence that they will be effective.

- Some people will be more persuaded by appeals to play by the rules, some by duty to the community, and some to personal risk.

> All these different approaches are needed. The messaging also needs to take account of the realities of different people's lives. Messaging needs to take account of the different motivational levers and circumstances of different people.

All this could be achieved the SPI-B psychologists said by *using the media to increase the sense of personal threat* which translates as terrify the shit out of the population, including children, so they all do what we want. That's not happened has it? Those excuses for 'journalists' who wouldn't know journalism if it bit them on the arse (the great majority) have played their crucial part in serving this Cult-government Psyop to enslave their own kids and grandkids. How they live with themselves I have no idea. The psychological war has been underpinned by constant government 'Covid' propaganda in almost every television and radio ad break, plus the Internet and print media, which has pounded out the fear with taxpayers footing the bill for their own programming. The result has been people terrified of a 'virus' that doesn't exist or one with a tiny fatality rate even if you believe it does. People walk down the street and around the shops wearing face-nappies damaging their health and psychology while others report those who refuse to be that naïve to the police who turn up in their own face-nappies. I had a cameraman come to my flat and he was so frightened of 'Covid' he came in wearing a mask and refused to shake my hand in case he caught something. He had – naïveitis – and the thought that he worked in the mainstream media was both depressing and made his behaviour perfectly explainable. The fear which has gripped the minds of so many and frozen them into compliance has been carefully cultivated by these psychologists who are really psychopaths. If lives get destroyed and a lot of young people commit suicide it shows our plan is working. SPI-B then turned to compulsion on the public to comply. 'With adequate preparation, rapid change can be achieved', it said. Some countries had introduced mandatory self-isolation on a wide scale without evidence of major public unrest and a large majority of the UK's population appeared to be supportive of more coercive measures with 64 percent of adults saying they would support putting London under a lockdown (watch the 'polls' which are designed to make people believe that public opinion is in favour or against whatever the subject in hand).

For 'aggressive protective measures' to be effective, the SPI-B paper said, special attention should be devoted to those population groups that are more at risk. Translated from the Orwellian this means making the rest of population feel guilty for not protecting the 'vulnerable' such as old people which the Cult and its agencies were about to kill on an industrial scale with lockdown, lack of treatment and the Gates 'vaccine'. Psychopath psychologists sold their guilt-trip so comprehensively that Los Angeles County Supervisor Hilda Solis reported that children were apologising (from a distance) to their parents and grandparents for bringing 'Covid' into their homes and getting them sick. '… These apologies are just some of the last words that loved ones will ever hear as they die alone,' she said. Gut-wrenchingly Solis then used this childhood tragedy to tell children to stay at home and 'keep your loved ones alive'. Imagine heaping such potentially life-long guilt on a kid when it has absolutely nothing to do with them. These people are deeply disturbed and the psychologists behind this even more so.

Uncivil war – divide and rule

Professional mind-controllers at SPI-B wanted the media to increase a sense of responsibility to others (do as you're told) and promote 'positive messaging' for those actions while in contrast to invoke 'social disapproval' by the unquestioning, obedient, community of anyone with a mind of their own. Again the compliant Goebbels-like media obliged. This is an old, old, trick employed by tyrannies the world over throughout human history. You get the target population to keep the target population in line – *your* line. SPI-B said this could 'play an important role in preventing anti-social behaviour or discouraging failure to enact pro-social behaviour'. For 'anti-social' in the Orwellian parlance of SPI-B see any behaviour that government doesn't approve. SPI-B recommendations said that 'social disapproval' should be accompanied by clear messaging and promotion of strong collective identity – hence the government and celebrity mantra of 'we're all in this together'. Sure we are. The mind doctors have such contempt for their targets that they think some clueless comedian, actor or singer telling them to do what the government wants will be enough to win them over. We have had UK comedian Lenny Henry, actor Michael Caine and singer Elton John wheeled out to serve the propagandists by urging people to have the DNA-manipulating 'Covid' non-'vaccine'. The role of Henry and fellow black celebrities in seeking to coax a 'vaccine' reluctant black community into doing the government's will was especially stomach-turning. An emotion-manipulating script and carefully edited video featuring these black 'celebs' was such an insult to the intelligence of black people and where's the self-respect of those involved selling their souls to a fascist government agenda? Henry said he heard black people's 'legitimate worries and concerns', but people must 'trust the facts' when they were doing exactly that by not having the 'vaccine'. They had to include the obligatory reference to Black Lives Matter with the line … 'Don't let coronavirus cost even more black lives – because we matter'. My god, it was pathetic. 'I know the vaccine is safe and what it does.' How? 'I'm a comedian and it says so in my script.'

SPI-B said social disapproval needed to be carefully managed to avoid victimisation, scapegoating and misdirected criticism, but they knew that their 'recommendations' would lead to exactly that and the media were specifically used to stir-up the divide-and-conquer hostility. Those who conform like good little baa, baas, are praised while those who have seen through the tidal wave of lies are 'Covidiots'. The awake have been abused by the fast asleep for not conforming to fascism and impositions that the awake know are designed to endanger their health, dehumanise them, and tear asunder the very fabric of human society. We have had the curtain-twitchers and morons reporting neighbours and others to the face-nappied police for breaking 'Covid rules' with fascist police delighting in posting links and phone numbers where this could be done. The Cult cannot impose its will without a compliant police and military or a compliant population willing to play their part in enslaving themselves and their kids. The words of a pastor in Nazi Germany are so appropriate today:

> First they came for the socialists and I did not speak out because I was not a socialist.
>
> Then they came for the trade unionists and I did not speak out because I was not a trade unionist.
>
> Then they came for the Jews and I did not speak out because I was not a Jew.
>
> Then they came for me and there was no one left to speak for me.

Those who don't learn from history are destined to repeat it and so many are.

'Covid' rules: Rewiring the mind

With the background laid out to this gigantic national and global web of psychological manipulation we can put 'Covid' rules into a clear and sinister perspective. Forget the claims about protecting health. 'Covid' rules are about dismantling the human mind, breaking the human spirit, destroying self-respect, and then putting Humpty Dumpty together again as a servile, submissive slave. Social isolation through lockdown and distancing have devastating effects on the human psyche as the psychological psychopaths well know and that's the real reason for them. Humans need contact with each other, discourse, closeness and touch, or they eventually, and literarily, go crazy. Masks, which I will address at some length, fundamentally add to the effects of isolation and the Cult agenda to dehumanise and de-individualise the population. To do this while knowing – in fact *seeking* – this outcome is the very epitome of evil and psychologists involved in this *are* the epitome of evil. They must like all the rest of the Cult demons and their assets stand trial for crimes against humanity on a scale that defies the imagination. Psychopaths in uniform use isolation to break enemy troops and agents and make them subservient and submissive to tell what they know. The technique is rightly considered a form of torture and torture is most certainly what has been imposed on the human population.

Clinically-insane American psychologist Harry Harlow became famous for his isolation experiments in the 1950s in which he separated baby monkeys from their mothers and imprisoned them for months on end in a metal container or 'pit of despair'. They soon began to show mental distress and depression as any idiot could have predicted. Harlow put other monkeys in steel chambers for three, six or twelve months while denying them any contact with animals or humans. He said that the effects of total social isolation for six months were 'so devastating and debilitating that we had assumed initially that twelve months of isolation would not produce any additional decrement'; but twelve months of isolation 'almost obliterated the animals socially'. This is what the Cult and its psychopaths are doing to you and your children. Even monkeys in partial isolation in which they were not allowed to form relationships with other monkeys became 'aggressive and hostile, not only to others, but also towards their own bodies'. We have seen this in the young as a consequence of lockdown. UK government psychopaths launched a public relations campaign telling people not to hug each other even after they received the 'Covid-19 vaccine' which we were told with more lies would allow a return to 'normal life'. A government source told *The Telegraph*: 'It will be along the lines that it is great that you have been vaccinated, but if you are going to visit your family and hug your grandchildren there is a chance you are going to infect people you love.' The source was apparently speaking from a secure psychiatric facility. Janet Lord, director of Birmingham University's Institute of Inflammation and Ageing, said that parents and grandparents should avoid hugging their children. Well, how can I put it, Ms Lord? Fuck off. Yep, that'll do.

Destroying the kids – where are the parents?

Observe what has happened to people enslaved and isolated by lockdown as suicide and self-harm has soared worldwide, particularly among the young denied the freedom to associate with their friends. A study of 49,000 people in English-speaking countries concluded that almost half of young adults are at clinical risk of mental health disorders. A national survey in America of 1,000 currently enrolled high school and college students found that 5 percent reported attempting suicide during the pandemic. Data from the US CDC's National Syndromic Surveillance Program from January 1st to October 17th, 2020, revealed a *31 percent* increase in mental health issues among adolescents aged 12 to 17 compared with 2019. The CDC reported that America in general suffered the biggest drop in life expectancy since World War Two as it fell by a year in the first half of 2020 as a result of 'deaths of despair' – overdoses and suicides. Deaths of despair have leapt by more than 20 percent during lockdown and include the highest number of fatal overdoses ever recorded in a single year – 81,000. Internet addiction is another consequence of being isolated at home which lowers interest in physical activities as kids fall into inertia and what's the point? Children and young people are losing hope and giving up on life, sometimes literally. A 14-year-old boy killed himself in Maryland because he had 'given up' when his school district didn't reopen; an 11-year-old boy shot himself during a zoom class; a teenager in Maine succumbed to the isolation of the 'pandemic' when he ended his life after experiencing a disrupted senior year at school. Children as young as nine have taken their life and all these stories can be repeated around the world. Careers are being destroyed before they start and that includes those in sport in which promising youngsters have not been able to take part. The plan of the psycho-psychologists is working all right. Researchers at Cambridge University found that lockdowns cause significant harm to children's mental health. Their study was published in the *Archives of Disease in Childhood*, and followed 168 children aged between 7 and 11. The researchers concluded:

> During the UK lockdown, children's depression symptoms have increased substantially, relative to before lockdown. The scale of this effect has direct relevance for the continuation of different elements of lockdown policy, such as complete or partial school closures …

> … Specifically, we observed a statistically significant increase in ratings of depression, with a medium-to-large effect size. Our findings emphasise the need to incorporate the potential impact of lockdown on child mental health in planning the ongoing response to the global pandemic and the recovery from it.

Not a chance when the Cult's psycho-psychologists were getting exactly what they wanted. The UK's Royal College of Paediatrics and Child Health has urged parents to look for signs of eating disorders in children and young people after a three to four fold increase. Specialists say the 'pandemic' is a major reason behind the rise. You don't say. The College said isolation from friends during school closures, exam cancellations, loss of extra-curricular activities like sport, and an increased use of social media were all contributory factors along with fears about the virus (psycho-psychologists again), family finances, and students being forced to quarantine. Doctors said young people were becoming severely ill by the time they were seen with 'Covid' regulations reducing face-to-face consultations. Nor is it only the young that have been devastated by the psychopaths. Like all bullies and cowards the Cult is targeting the young, elderly, weak and infirm. A typical story was told by a British lady called Lynn Parker who was not allowed to visit her husband in 2020 for the last ten and half months of his life 'when he needed me most' between March 20th and when he died on December 19th. This vacates the criminal and enters the territory of evil. The emotional impact on the immune system alone is immense as are the number of people of all ages worldwide who have died as a result of Cult-demanded, Gates-demanded, lockdowns.

Isolation is torture

The experience of imposing solitary confinement on millions of prisoners around the world has shown how a large percentage become 'actively psychotic and/or acutely suicidal'. Social isolation has been found to trigger 'a specific psychiatric syndrome, characterized by hallucinations; panic attacks; overt paranoia; diminished impulse control; hypersensitivity to external stimuli; and difficulties with thinking, concentration and memory'. Juan Mendez, a United Nations rapporteur (investigator), said that isolation is a form of torture. Research has shown that even after isolation prisoners find it far more difficult to make social connections and I remember chatting to a shop assistant after one lockdown who told me that when her young son met another child again he had no idea how to act or what to do. Hannah Flanagan, Director of Emergency Services at Journey Mental Health Center in Dane County, Wisconsin, said: 'The specificity about Covid social distancing and isolation that we've come across as contributing factors to the suicides are really new to us this year.' But they are not new to those that devised them. They are getting the effect they want as the population is psychologically dismantled to be rebuilt in a totally different way. Children and the young are particularly targeted. They will be the adults when the full-on fascist AI-controlled technocracy is planned to be imposed and they are being prepared to meekly submit. At the same time older people who still have a memory of what life was like before – and how fascist the new normal really is – are being deleted. You are going to see efforts to turn the young against the old to support this geriatric genocide. Hannah Flanagan said the big increase in suicide in her county proved that social isolation is not only harmful, but deadly. Studies have shown that isolation from others is one of the main risk factors in suicide and even more so with women. Warnings that lockdown could create a 'perfect storm' for suicide were ignored. After all this was one of the *reasons* for lockdown. Suicide, however, is only the most extreme of isolation consequences. There are many others. Dr Dhruv Khullar, assistant professor of healthcare policy at Weill Cornell Medical College, said in a *New York Times* article in 2016 long before the fake 'pandemic':

> A wave of new research suggests social separation is bad for us. Individuals with less social connection have disrupted sleep patterns, altered immune systems, more inflammation and higher levels of stress hormones. One recent study found that isolation increases the risk of heart disease by 29 percent and stroke by 32 percent. Another analysis that pooled data from 70 studies and 3.4 million people found that socially isolated individuals had a 30 percent higher risk of dying in the next seven years, and that this effect was largest in middle age.

> Loneliness can accelerate cognitive decline in older adults, and isolated individuals are twice as likely to die prematurely as those with more robust social interactions. These effects start early: Socially isolated children have significantly poorer health 20 years later, even after controlling for other factors. All told, loneliness is as important a risk factor for early death as obesity and smoking.

There you have proof from that one article alone four years before 2020 that those who have enforced lockdown, social distancing and isolation knew what the effect would be and that is even more so with professional psychologists that have been driving the policy across the globe. We can go back even further to the years 2000 and 2003 and the start of a major study on the effects of isolation on health by Dr Janine Gronewold and Professor Dirk M. Hermann at the University Hospital in Essen, Germany, who analysed data on 4,316 people with an average age of 59 who were recruited for the long-term research project. They found that socially isolated people are more than 40 percent more likely to have a heart attack, stroke, or other major cardiovascular event and nearly 50 percent more likely to die from any cause. Given the financial Armageddon unleashed by lockdown we should note that the study found a relationship between increased cardiovascular risk and lack of financial support. After excluding other factors social isolation was still connected to a 44 percent increased risk of cardiovascular problems and a 47 percent increased risk of death by any cause. Lack of financial support was associated with a 30 percent increase in the risk of cardiovascular health events. Dr Gronewold said it had been known for some time that feeling lonely or lacking contact with close friends and family can have an impact on physical health and the study had shown that having strong social relationships is of high importance for heart health. Gronewold said they didn't understand yet why people who are socially isolated have such poor health outcomes, but this was obviously a worrying finding, particularly during these times of prolonged social distancing. Well, it can be explained on many levels. You only have to identify the point in the body where people feel loneliness and missing people they are parted from – it's in the centre of the chest where they feel the ache of loneliness and the ache of missing people. 'My heart aches for you' … 'My heart aches for some company.' I will explain this more in the chapter Escaping Wetiko, but when you realise that the body is the mind – they are expressions of each other – the reason why state of the mind dictates state of the body becomes clear.

American psychologist Ranjit Powar was highlighting the effects of lockdown isolation as early as April, 2020. She said humans have evolved to be social creatures and are wired to live in interactive groups. Being isolated from family, friends and colleagues could be unbalancing and traumatic for most people and could result in short or even long-term psychological and physical health problems. An increase in levels of anxiety, aggression, depression, forgetfulness and hallucinations were possible psychological effects of isolation. 'Mental conditions may be precipitated for those with underlying pre-existing susceptibilities and show up in many others without any pre-condition.' Powar said personal relationships helped us cope with stress and if we lost this outlet for letting off steam the result can be a big emotional void which, for an average person, was difficult to deal with. 'Just a few days of isolation can cause increased levels of anxiety and depression' – so what the hell has been the effect on the global population of *18 months* of this at the time of writing? Powar said: 'Add to it the looming threat of a dreadful disease being repeatedly hammered in through the media and you have a recipe for many shades of mental and physical distress.' For those with a house and a garden it is easy to forget that billions have had to endure lockdown isolation in tiny overcrowded flats and apartments with nowhere to go outside. The psychological and physical consequences of this are unimaginable and with lunatic and abusive partners and parents the consequences have led to tremendous increases in domestic and child abuse and alcoholism as people seek to shut out the horror. Ranjit Powar said:

> Staying in a confined space with family is not all a rosy picture for everyone. It can be extremely oppressive and claustrophobic for large low-income families huddled together in small single-room houses. Children here are not lucky enough to have many board/electronic games or books to keep them occupied.

> Add to it the deep insecurity of running out of funds for food and basic necessities. On the other hand, there are people with dysfunctional family dynamics, such as domineering, abusive or alcoholic partners, siblings or parents which makes staying home a period of trial. Incidence of suicide and physical abuse against women has shown a worldwide increase. Heightened anxiety and depression also affect a person's immune system, making them more susceptible to illness.

To think that Powar's article was published on April 11th, 2020.

Six-feet fantasy

Social (unsocial) distancing demanded that people stay six feet or two metres apart. UK government advisor Robert Dingwall from the New and Emerging Respiratory Virus Threats Advisory Group said in a radio interview that the two-metre rule was 'conjured up out of nowhere' and was not based on science. No, it was not based on *medical* science, but it didn't come out of nowhere. The distance related to *psychological* science. Six feet/two metres was adopted in many countries and we were told by people like the criminal Anthony Fauci and his ilk that it was founded on science. Many schools could not reopen because they did not have the space for six-feet distancing. Then in March, 2021, after a year of six-feet 'science', a study published in the *Journal of Infectious Diseases* involving more than 500,000 students and almost 100,000 staff over 16 weeks revealed no significant difference in 'Covid' cases between six feet and three feet and Fauci changed his tune. Now three feet was okay. There is no difference between six feet and three *inches* when there is no 'virus' and they got away with six feet for psychological reasons for as long as they could. I hear journalists and others talk about 'unintended consequences' of lockdown. They are not *unintended* at all; they have been coldly-calculated for a specific outcome of human control and that's why super-psychopaths like Gates have called for them so vehemently. Super-psychopath psychologists have demanded them and psychopathic or clueless, spineless, politicians have gone along with them by 'following the science'. But it's not science at all. 'Science' is not what is; it's only what people can be manipulated to believe it is. The whole 'Covid' catastrophe is founded on mind control. Three word or three statement mantras issued by the UK government are a well-known mind control technique and so we've had 'Stay home/protect the NHS/save lives', 'Stay alert/control the virus/save lives' and 'hands/face/space'. One of the most vocal proponents of extreme 'Covid' rules in the UK has been Professor Susan Michie, a member of the British Communist Party, who is not a medical professional. Michie is the director of the Centre for Behaviour Change at University College London. She is a *behavioural psychologist* and another filthy rich 'Marxist' who praised China's draconian lockdown. She was known by fellow students at Oxford University as 'Stalin's nanny' for her extreme Marxism. Michie is an influential member of the UK government's Scientific Advisory Group for Emergencies (SAGE) and behavioural manipulation groups which have dominated 'Covid' policy. She is a consultant adviser to the World Health Organization on 'Covid-19' and behaviour. Why the hell are lockdowns anything to do with her when they are claimed to be about health? Why does a behavioural psychologist from a group charged with changing the behaviour of the public want lockdown, human isolation and mandatory masks? Does that question really need an answer? Michie *absolutely* has to explain herself before a Nuremberg court when humanity takes back its world again and even more so when you see the consequences of masks that she demands are compulsory. This is a Michie classic:

> The benefits of getting primary school children to wear masks is that regardless of what little degree of transmission is occurring in those age groups it could help normalise the practice. Young children wearing masks may be more likely to get their families to accept masks.

Those words alone should carry a prison sentence when you ponder on the callous disregard for children involved and what a statement it makes about the mind and motivations of Susan Michie. What a lovely lady and what she said there encapsulates the mentality of the psychopaths behind the 'Covid' horror. Let us compare what Michie said with a countrywide study in Germany published at researchsquare.com involving 25,000 school children and 17,854 health complaints submitted by parents. Researchers found that masks are harming children physically, psychologically, and behaviourally with 24 health issues associated with mask wearing. They include: shortness of breath (29.7%); dizziness (26.4%); increased headaches (53%); difficulty concentrating (50%); drowsiness or fatigue (37%); and malaise (42%). Nearly a third of children experienced more sleep issues than before and a quarter developed new fears. Researchers found health issues and other impairments in 68 percent of masked children covering their faces for an average of 4.5 hours a day. Hundreds of those taking part experienced accelerated respiration, tightness in the chest, weakness, and short-term impairment of consciousness. A reminder of what Michie said again:

> The benefits of getting primary school children to wear masks is that regardless of what little degree of transmission is occurring in those age groups it could help normalise the practice. Young children wearing masks may be more likely to get their families to accept masks.

Psychopaths in government and psychology now have children and young people – plus all the adults – wearing masks for hours on end while clueless teachers impose the will of the psychopaths on the young they should be protecting. What the hell are parents doing?

Cult lab rats

We have some schools already imposing on students microchipped buzzers that activate when they get ‘too close’ to their pals in the way they do with lab rats. How apt. To the Cult and its brain-dead servants our children *are* lab rats being conditioned to be unquestioning, dehumanised slaves for the rest of their lives. Children and young people are being weaned and frightened away from the most natural human instincts including closeness and touch. I have tracked in the books over the years how schools were banning pupils from greeting each other with a hug and the whole Cult-induced Me Too movement has terrified men and boys from a relaxed and natural interaction with female friends and work colleagues to the point where many men try never to be in a room alone with a woman that's not their partner. Airhead celebrities have as always played their virtue-signalling part in making this happen with their gross exaggeration. For every monster like Harvey Weinstein there are at least tens of thousands of men that don't treat women like that; but everyone must be branded the same and policy changed for them as well as the monster. I am going to be using the word 'dehumanise' many times in this chapter because that is what the Cult is seeking to do and it goes very deep as we shall see. Don't let them kid you that social distancing is planned to end one day. That's not the idea. We are seeing more governments and companies funding and producing wearable gadgets to keep people apart and they would not be doing that if this was meant to be short-term. A tech start-up company backed by GCHQ, the British Intelligence and military surveillance headquarters, has created a social distancing wrist sensor that alerts people when they get too close to others. The CIA has also supported tech companies developing similar devices. The wearable sensor was developed by Tended, one of a number of start-up companies supported by GCHQ (see the CIA and DARPA). The device can be worn on the wrist or as a tag on the waistband and will vibrate whenever someone wearing the device breaches social distancing and gets anywhere near natural human contact. The company had a lucky break in that it was developing a distancing sensor when the 'Covid' hoax arrived which immediately provided a potentially enormous market. How fortunate. The government in big-time Cult-controlled Ontario in Canada is investing $2.5 million in wearable contact tracing technology that 'will alert users if they may have been exposed to the Covid-19 in the workplace and will beep or vibrate if they are within six feet of another person'. Facedrive Inc., the technology company behind this, was founded in 2016 with funding from the Ontario Together Fund and obviously they, too, had a prophet on the board of directors. The human surveillance and control technology is called TraceSCAN and would be worn by the human cyborgs in places such as airports, workplaces, construction sites, care homes and … *schools*.

I emphasise schools with children and young people the prime targets. You know what is planned for society as a whole if you keep your eyes on the schools. They have always been places where the state program the next generation of slaves to be its compliant worker-ants – or Woker-ants these days; but in the mist of the 'Covid' madness they have been transformed into mind laboratories on a scale never seen before. Teachers and head teachers are just as programmed as the kids – often more so. Children are kept apart from human interaction by walk lanes, classroom distancing, staggered meal times, masks, and the rolling-out of buzzer systems. Schools are now physically laid out as a laboratory maze for lab-rats. Lunatics at a school in Anchorage, Alaska, who should be prosecuted for child abuse, took away desks and forced children to kneel (know your place) on a mat for five hours a day while wearing a mask and using their chairs as a desk. How this was supposed to impact on a 'virus' only these clinically insane people can tell you and even then it would be clap-trap. The school banned recess (interaction), art classes (creativity), and physical exercise (getting body and mind moving out of inertia). Everyone behind this outrage should be in jail or better still a mental institution. The behavioural manipulators are all for this dystopian approach to schools. Professor Susan Michie, the mind-doctor and British Communist Party member, said it was wrong to say that schools were safe. They had to be made so by 'distancing', masks and ventilation (sitting all day in the cold). I must ask this lady round for dinner on a night I know I am going to be out and not back for weeks. She probably wouldn't be able to make it, anyway, with all the visits to her own psychologist she must have block-booked.

Masking identity

I know how shocking it must be for you that a behaviour manipulator like Michie wants everyone to wear masks which have long been a feature of mind-control programs like the infamous MKUltra in the United States, but, there we are. We live and learn. I spent many years from 1996 to right across the millennium researching mind control in detail on both sides of the Atlantic and elsewhere. I met a large number of mind-control survivors and many had been held captive in body and mind by MKUltra. MK stands for mind-control, but employs the German spelling in deference to the Nazis spirited out of Germany at the end of World War Two by Operation Paperclip in which the US authorities, with help from the Vatican, transported Nazi mind-controllers and engineers to America to continue their work. Many of them were behind the creation of NASA and they included Nazi scientist and SS officer Wernher von Braun who swapped designing V-2 rockets to bombard London with designing the Saturn V rockets that powered the NASA moon programme's Apollo craft. I think I may have mentioned that the Cult has no borders. Among Paperclip escapees was Josef Mengele, the Angel of Death in the Nazi concentration camps where he conducted mind and genetic experiments on children often using twins to provide a control twin to measure the impact of his 'work' on the other. If you want to observe the Cult mentality in all its extremes of evil then look into the life of Mengele. I have met many people who suffered mercilessly under Mengele in the United States where he operated under the name Dr Greene and became a stalwart of MKUltra programming and torture. Among his locations was the underground facility in the Mojave Desert in California called the China Lake Naval Weapons Station which is almost entirely below the surface. My books *The Biggest Secret*, *Children of the Matrix* and *The Perception Deception* have the detailed background to MKUltra.

The best-known MKUltra survivor is American Cathy O'Brien. I first met her and her late partner Mark Phillips at a conference in Colorado in 1996. Mark helped her escape and deprogram from decades of captivity in an offshoot of MKUltra known as Project Monarch in which 'sex slaves' were provided for the rich and famous including Father George Bush, Dick Cheney and the Clintons. Read Cathy and Mark's book *Trance-Formation of America* and if you are new to this you will be shocked to the core. I read it in 1996 shortly before, with the usual synchronicity of my life, I found myself given a book table at the conference right next to hers. MKUltra never ended despite being very publicly exposed (only a small part of it) in the 1970s and continues in other guises. I am still in touch with Cathy. She contacted me during 2020 after masks became compulsory in many countries to tell me how they were used as part of MKUltra programming. I had been observing 'Covid regulations' and the relationship between authority and public for months. I saw techniques that I knew were employed on individuals in MKUltra being used on the global population. I had read many books and manuals on mind control including one called *Silent Weapons for Quiet Wars* which came to light in the 1980s and was a guide on how to perceptually program on a mass scale. 'Silent Weapons' refers to mind-control. I remembered a line from the manual as governments, medical authorities and law enforcement agencies have so obviously talked to – or rather at – the adult population since the 'Covid' hoax began as if they are children. The document said:

> If a person is spoken to by a T.V. advertiser as if he were a twelve-year-old, then, due to suggestibility, he will, with a certain probability, respond or react to that suggestion with the uncritical response of a twelve-year-old and will reach in to his economic reservoir and deliver its energy to buy that product on impulse when he passes it in the store.

That's why authority has spoken to adults like children since all this began.

Why did Michael Jackson wear masks?

Every aspect of the 'Covid' narrative has mind-control as its central theme. Cathy O'Brien wrote an article for davidicke.com about the connection between masks and mind control. Her daughter Kelly who I first met in the 1990s was born while Cathy was still held captive in MKUltra. Kelly was forced to wear a mask as part of her programming from the age of *two* to dehumanise her, target her sense of individuality and reduce the amount of oxygen her brain and body received. *Bingo*. This is the real reason for compulsory masks, why they have been enforced en masse, and why they seek to increase the number they demand you wear. First one, then two, with one disgraceful alleged 'doctor' recommending four which is nothing less than a death sentence. Where and how often they must be worn is being expanded for the purpose of mass mind control and damaging respiratory health which they can call 'Covid-19'. Canada's government headed by the man-child Justin Trudeau, says it's fine for children of two and older to wear masks. An insane 'study' in Italy involving just 47 children concluded there was no problem for babies as young as *four months* wearing them. Even after people were 'vaccinated' they were still told to wear masks by the criminal that is Anthony Fauci. Cathy wrote that mandating masks is allowing the authorities literally to control the air we breathe which is what was done in MKUltra. You might recall how the singer Michael Jackson wore masks and there is a reason for that. He was subjected to MKUltra mind control through Project Monarch and his psyche was scrambled by these simpletons. Cathy wrote:

> In MKUltra Project Monarch mind control, Michael Jackson had to wear a mask to silence his voice so he could not reach out for help. Remember how he developed that whisper voice when he wasn't singing? Masks control the mind from the outside in, like the redefining of words is doing. By controlling what we can and cannot say for fear of being labeled racist or beaten, for example, it ultimately controls thought that drives our words and ultimately actions (or lack thereof).
>
> Likewise, a mask muffles our speech so that we are not heard, which controls voice … words … mind. This is Mind Control. Masks are an obvious mind control device, and I am disturbed so many people are complying on a global scale. Masks depersonalize while making a person feel as though they have no voice. It is a barrier to others. People who would never choose to comply but are forced to wear a mask in order to keep their job, and ultimately their family fed, are compromised. They often feel shame and are subdued. People have stopped talking with each other while media controls the narrative.

The 'no voice' theme has often become literal with train passengers told not to speak to each other in case they pass on the 'virus', singing banned for the same reason and bonkers California officials telling people riding roller coasters that they cannot shout and scream. Cathy said she heard every day from healed MKUltra survivors who cannot wear a mask without flashing back on ways their breathing was controlled – 'from ball gags and penises to water boarding'. She said that through the years when she saw images of people in China wearing masks 'due to pollution' that it was really to control their oxygen levels. 'I knew it was as much of a population control mechanism of depersonalisation as are burkas', she said. Masks are another Chinese communist/fascist method of control that has been swept across the West as the West becomes China at lightning speed since we entered 2020.

Mask-19

There are other reasons for mandatory masks and these include destroying respiratory health to call it 'Covid-19' and stunting brain development of children and the young. Dr Margarite Griesz-Brisson MD, PhD, is a Consultant Neurologist and Neurophysiologist and the Founder and Medical Director of the London Neurology and Pain Clinic. Her CV goes down the street and round the corner. She is clearly someone who cares about people and won't parrot the propaganda. Griesz-Brisson has a PhD in pharmacology, with special interest in neurotoxicology, environmental medicine, neuroregeneration and neuroplasticity (the way the brain can change in the light of information received). She went public in October, 2020, with a passionate warning about the effects of mask-wearing laws:

> The reinhalation of our exhaled air will without a doubt create oxygen deficiency and a flooding of carbon dioxide. We know that the human brain is very sensitive to oxygen deprivation. There are nerve cells for example in the hippocampus that can't be longer than 3 minutes without oxygen – they cannot survive. The acute warning symptoms are headaches, drowsiness, dizziness, issues in concentration, slowing down of reaction time – reactions of the cognitive system.

Oh, I know, let's tell bus, truck and taxi drivers to wear them and people working machinery. How about pilots, doctors and police? Griesz-Brisson makes the important point that while the symptoms she mentions may fade as the body readjusts this does not alter the fact that people continue to operate in oxygen deficit with long list of potential consequences. She said it was well known that neurodegenerative diseases take years or decades to develop. 'If today you forget your phone number, the breakdown in your brain would have already started 20 or 30 years ago.' She said degenerative processes in your brain are getting amplified as your oxygen deprivation continues through wearing a mask. Nerve cells in the brain are unable to divide themselves normally in these circumstances and lost nerve cells will no longer be regenerated. 'What is gone is gone.' Now consider that people like shop workers and *schoolchildren* are wearing masks for hours every day. What in the name of sanity is going to be happening to them? 'I do not wear a mask, I need my brain to think', Griesz-Brisson said, 'I want to have a clear head when I deal with my patients and not be in a carbon dioxide-induced anaesthesia'. If you are told to wear a mask anywhere ask the organisation, police, store, whatever, for their risk assessment on the dangers and negative effects on mind and body of enforcing mask-wearing. They won't have one because it has never been done not even by government. All of them must be subject to class-action lawsuits as the consequences come to light. They don't do mask risk assessments for an obvious reason. They know what the conclusions would be and independent scientific studies that *have* been done tell a horror story of consequences.

'Masks are criminal'

Dr Griesz-Brisson said that for children and adolescents, masks are an absolute no-no. They had an extremely active and adaptive immune system and their brain was incredibly active with so much to learn. 'The child's brain, or the youth's brain, is thirsting for oxygen.' The more metabolically active an organ was, the more oxygen it required; and in children and adolescents every organ was metabolically active. Griesz-Brisson said that to deprive a child's or adolescent's brain of oxygen, or to restrict it in any way, was not only dangerous to their health, it was absolutely criminal. 'Oxygen deficiency inhibits the development of the brain, and the damage that has taken place as a result CANNOT be reversed.' Mind manipulators of MKUltra put masks on two-year-olds they wanted to neurologically rewire and you can see why. Griesz-Brisson said a child needs the brain to learn and the brain needs oxygen to function. 'We don't need a clinical study for that. This is simple, indisputable physiology.' Consciously and purposely induced oxygen deficiency was an absolutely deliberate health hazard, and an absolute medical contraindication which means that 'this drug, this therapy, this method or measure should not be used, and is not allowed to be used'. To coerce an entire population to use an absolute medical contraindication by force, she said, there had to be definite and serious reasons and the reasons must be presented to competent interdisciplinary and independent bodies to be verified and authorised. She had this warning of the consequences that were coming if mask wearing continued:

> When, in ten years, dementia is going to increase exponentially, and the younger generations couldn't reach their god-given potential, it won't help to say 'we didn't need the masks'. I know how damaging oxygen deprivation is for the brain, cardiologists know how damaging it is for the heart, pulmonologists know how damaging it is for the lungs. Oxygen deprivation damages every single organ. Where are our health departments, our health insurance, our medical associations? It would have been their duty to be vehemently against the lockdown and to stop it and stop it from the very beginning.
>
> Why do the medical boards issue punishments to doctors who give people exemptions? Does the person or the doctor seriously have to prove that oxygen deprivation harms people? What kind of medicine are our doctors and medical associations representing? Who is responsible for this crime? The ones who want to enforce it? The ones who let it happen and play along, or the ones who don't prevent it?

All of the organisations and people she mentions there either answer directly to the Cult or do whatever hierarchical levels above them tell them to do. The outcome of both is the same. 'It's not about masks, it's not about viruses, it's certainly not about your health', Griesz-Brisson said. 'It is about much, much more. I am not participating. I am not afraid.' They were taking our air to breathe and there was no unfounded medical exemption from face masks. Oxygen deprivation was dangerous for every single brain. It had to be the free decision of every human being whether they want to wear a mask that was absolutely ineffective to protect themselves from a virus. She ended by rightly identifying where the responsibility lies for all this:

> The imperative of the hour is personal responsibility. We are responsible for what we think, not the media. We are responsible for what we do, not our superiors. We are responsible for our health, not the World Health Organization. And we are responsible for what happens in our country, not the government.

Halle-bloody-lujah.

But surgeons wear masks, right?

Independent studies of mask-wearing have produced a long list of reports detailing mental, emotional and physical dangers. What a definition of insanity to see police officers imposing mask-wearing on the public which will cumulatively damage their health while the police themselves wear masks that will cumulatively damage *their* health. It's utter madness and both public and police do this because 'the government says so' – yes a government of brain-donor idiots like UK Health Secretary Matt Hancock reading the 'follow the science' scripts of psychopathic, lunatic psychologists. The response you get from Stockholm syndrome sufferers defending the very authorities that are destroying them and their families is that 'surgeons wear masks'. This is considered the game, set and match that they must work and don't cause oxygen deficit. Well, actually, scientific studies have shown that they *do* and oxygen levels are monitored in operating theatres to compensate. Surgeons wear masks to stop spittle and such like dropping into open wounds – not to stop 'viral particles' which are so miniscule they can only be seen through an electron microscope. Holes in the masks are significantly bigger than 'viral particles' and if you sneeze or cough they will breach the mask. I watched an incredibly disingenuous 'experiment' that claimed to prove that masks work in catching 'virus' material from the mouth and nose. They did this with a slow motion camera and the mask did block big stuff which stayed inside the mask and against the face to be breathed in or cause infections on the face as we have seen with many children. ‘Viral particles’, however, would never have been picked up by the camera as they came through the mask when they are far too small to be seen. The ‘experiment’ was therefore disingenuous *and* useless.

Studies have concluded that wearing masks in operating theatres (and thus elsewhere) make no difference to preventing infection while the opposite is true with toxic shite building up in the mask and this had led to an explosion in tooth decay and gum disease dubbed by dentists ‘mask mouth’. You might have seen the Internet video of a furious American doctor urging people to take off their masks after a four-year-old patient had been rushed to hospital the night before and nearly died with a lung infection that doctors sourced to mask wearing. A study in the journal *Cancer Discovery* found that inhalation of harmful microbes can contribute to advanced stage lung cancer in adults and long-term use of masks can help breed dangerous pathogens. Microbiologists have said frequent mask wearing creates a moist environment in which microbes can grow and proliferate before entering the lungs. The Canadian Agency for Drugs and Technologies in Health, or CADTH, a Canadian national organisation that provides research and analysis to healthcare decision-makers, said this as long ago as 2013 in a report entitled ‘Use of Surgical Masks in the Operating Room: A Review of the Clinical Effectiveness and Guidelines’. It said:

- No evidence was found to support the use of surgical face masks to reduce the frequency of surgical site infections
- No evidence was found on the effectiveness of wearing surgical face masks to protect staff from infectious material in the operating room.
- Guidelines recommend the use of surgical face masks by staff in the operating room to protect both operating room staff and patients (despite the lack of evidence).

We were told that the world could go back to 'normal' with the arrival of the 'vaccines'. When they came, fraudulent as they are, the story changed as I knew that it would. We are in the midst of transforming 'normal', not going back to it. Mary Ramsay, head of immunisation at Public Health England, echoed the words of US criminal Anthony Fauci who said masks and other regulations must stay no matter if people are vaccinated. The Fauci idiot continued to wear two masks – different colours so both could be clearly seen – after he *claimed* to have been vaccinated. Senator Rand Paul told Fauci in one exchange that his double-masks were 'theatre' and he was right. It's all theatre. Mary Ramsay back-tracked on the vaccine-return-to-normal theme when she said the public may need to wear masks and social-distance for years despite the jabs. 'People have got used to those lower-level restrictions now, and [they] can live with them', she said telling us what the idea has been all along. 'The vaccine does not give you a pass, even if you have had it, you must continue to follow all the guidelines' said a Public Health England statement which reneged on what we had been told before and made having the 'vaccine' irrelevant to 'normality' even by the official story. Spain's fascist government trumped everyone by passing a law mandating the wearing of masks on the beach and even when swimming in the sea. The move would have devastated what's left of the Spanish tourist industry, posed potential breathing dangers to swimmers and had Northern European sunbathers walking around with their forehead brown and the rest of their face white as a sheet. The ruling was so crazy that it had to be retracted after pressure from public and tourist industry, but it confirmed where the Cult wants to go with masks and how clinically insane authority has become. The determination to make masks permanent and hide the serious dangers to body and mind can be seen in the censorship of scientist Professor Denis Rancourt by Bill Gates-funded academic publishing website ResearchGate over his papers exposing the dangers and uselessness of masks. Rancourt said:

> ResearchGate today has permanently locked my account, which I have had since 2015. Their reasons graphically show the nature of their attack against democracy, and their corruption of

science … By their obscene non-logic, a scientific review of science articles reporting on harms caused by face masks has a 'potential to cause harm'. No criticism of the psychological device (face masks) is tolerated, if the said criticism shows potential to influence public policy.

This is what happens in a fascist world.

Where are the 'greens' (again)?

Other dangers of wearing masks especially regularly relate to the inhalation of minute plastic fibres into the lungs and the deluge of discarded masks in the environment and oceans. Estimates predicted that more than 1.5 billion disposable masks will end up in the world's oceans every year polluting the water with tons of plastic and endangering marine wildlife. Studies project that humans are using 129 billion face masks each month worldwide – about three million a minute. Most are disposable and made from plastic, non-biodegradable microfibers that break down into smaller plastic particles that become widespread in ecosystems. They are littering cities, clogging sewage channels and turning up in bodies of water. I have written in other books about the immense amounts of microplastics from endless sources now being absorbed into the body. Rolf Halden, director of the Arizona State University (ASU) Biodesign Center for Environmental Health Engineering, was the senior researcher in a 2020 study that analysed 47 human tissue samples and found microplastics in all of them. 'We have detected these chemicals of plastics in every single organ that we have investigated', he said. I wrote in *The Answer* about the world being deluged with microplastics. A study by the Worldwide Fund for Nature (WWF) found that people are consuming on average every week some 2,000 tiny pieces of plastic mostly through water and also through marine life and the air. Every year humans are ingesting enough microplastics to fill a heaped dinner plate and in a life-time of 79 years it is enough to fill two large waste bins. Marco Lambertini, WWF International director general said: 'Not only are plastics polluting our oceans and waterways and killing marine life – it's in all of us and we can't escape consuming plastics,' American geologists found tiny plastic fibres, beads and shards in rainwater samples collected from the remote slopes of the Rocky Mountain National Park near Denver, Colorado. Their report was headed: 'It is raining plastic.' Rachel Adams, senior lecturer in Biomedical Science at Cardiff Metropolitan University, said that among health consequences are internal inflammation and immune responses to a 'foreign body'. She further pointed out that microplastics become carriers of toxins including mercury, pesticides and dioxins (a known cause of cancer and reproductive and developmental problems). These toxins accumulate in the fatty tissues once they enter the body through microplastics. Now this is being compounded massively by people putting plastic on their face and throwing it away.

Workers exposed to polypropylene plastic fibres known as 'flock' have developed 'flock worker's lung' from inhaling small pieces of the flock fibres which can damage lung tissue, reduce breathing capacity and exacerbate other respiratory problems. *Now* … commonly used surgical masks have three layers of melt-blown textiles made of … polypropylene. We have billions of people putting these microplastics against their mouth, nose and face for hours at a time day after day in the form of masks. How does anyone think that will work out? I mean – what could possibly go wrong? We posted a number of scientific studies on this at davidicke.com, but when I went back to them as I was writing this book the links to the science research website where they were hosted were dead. Anything that challenges the official narrative in any way is either censored or vilified. The official narrative is so unsupportable by the evidence that only deleting the truth can protect it. A study by Chinese scientists still survived – with the usual twist which it why it was still active, I guess. Yes, they found that virtually all the masks they tested increased the daily intake of microplastic fibres, but people should still wear them because the danger from the 'virus' was worse said the crazy 'team' from the Institute of Hydrobiology in Wuhan. Scientists first discovered microplastics in lung tissue of some patients who died of lung cancer in the 1990s. Subsequent studies have confirmed the potential health damage with the plastic degrading slowly and remaining in the lungs to accumulate in volume. Wuhan researchers used a machine simulating human breathing to establish that masks shed up to nearly 4,000 microplastic fibres in a month with reused masks producing more. Scientists said some masks are laced with toxic chemicals and a variety of compounds seriously restricted for both health and environmental reasons. They include cobalt (used in blue dye) and formaldehyde known to cause watery eyes, burning sensations in the eyes, nose, and throat, plus coughing, wheezing and nausea. No – that must be 'Covid-19'.

Mask 'worms'

There is another and potentially even more sinister content of masks. Mostly new masks of different makes filmed under a microscope around the world have been found to contain strange black fibres or 'worms' that appear to move or 'crawl' by themselves and react to heat and water. The nearest I have seen to them are the self-replicating fibres that are pulled out through the skin of those suffering from Morgellons disease which has been connected to the phenomena of 'chemtrails' which I will bring into the story later on. Morgellons fibres continue to grow outside the body and have a form of artificial intelligence. Black 'worm' fibres in masks have that kind of feel to them and there is a nanotechnology technique called 'worm micelles' which carry and release drugs or anything else you want to deliver to the body. For sure the suppression of humanity by mind altering drugs is the Cult agenda big time and the more excuses they can find to gain access to the body the more opportunities there are to make that happen whether through 'vaccines' or masks pushed against the mouth and nose for hours on end.

So let us summarise the pros and cons of masks:

Against masks: Breathing in your own carbon dioxide; depriving the body and brain of sufficient oxygen; build-up of toxins in the mask that can be breathed into the lungs and cause rashes on the face and 'mask-mouth'; breathing microplastic fibres and toxic chemicals into the lungs; dehumanisation and deleting individualisation by literally making people faceless; destroying human emotional interaction through facial expression and deleting parental connection with their babies which look for guidance to their facial expression.

For masks: They don't protect you from a 'virus' that doesn't exist and even if it did 'viral' particles are so minute they are smaller than the holes in the mask.

Governments, police, supermarkets, businesses, transport companies, and all the rest who seek to impose masks have done no risk assessment on their consequences for health and psychology and are now open to group lawsuits when the impact becomes clear with a cumulative epidemic of respiratory and other disease. Authorities will try to exploit these effects and hide the real cause by dubbing them 'Covid-19'. Can you imagine setting out to force the population to wear health-destroying masks without doing any assessment of the risks? It is criminal and it is evil, but then how many people targeted in this way, who see their children told to wear them all day at school, have asked for a risk assessment? Billions can't be imposed upon by the few unless the billions allow it. Oh, yes, with just a tinge of irony, 85 percent of all masks made worldwide come from *China*.

Wash your hands in toxic shite

'Covid' rules include the use of toxic sanitisers and again the health consequences of constantly applying toxins to be absorbed through the skin is obvious to any level of Renegade Mind. America's Food and Drug Administration (FDA) said that sanitisers are drugs and issued a warning about 75 dangerous brands which contain methanol used in antifreeze and can cause death, kidney damage and blindness. The FDA circulated the following warning even for those brands that it claims to be safe:

> Store hand sanitizer out of the reach of pets and children, and children should use it only with adult supervision. Do not drink hand sanitizer. This is particularly important for young children, especially toddlers, who may be attracted by the pleasant smell or brightly colored bottles of hand sanitizer.
>
> Drinking even a small amount of hand sanitizer can cause alcohol poisoning in children. (However, there is no need to be concerned if your children eat with or lick their hands after using hand sanitizer.) During this coronavirus pandemic, poison control centers have had an increase in calls about accidental ingestion of hand sanitizer, so it is important that adults monitor young children's use.
>
> Do not allow pets to swallow hand sanitizer. If you think your pet has eaten something potentially dangerous, call your veterinarian or a pet poison control center right away. Hand sanitizer is flammable and should be stored away from heat and flames. When using hand sanitizer, rub your hands until they feel completely dry before performing activities that may involve heat, sparks, static electricity, or open flames.

There you go, perfectly safe, then, and that's without even a mention of the toxins absorbed through the skin. Come on kids – sanitise your hands everywhere you go. It will save you from the 'virus'. Put all these elements together of the 'Covid' normal and see how much health and psychology is being cumulatively damaged, even devastated, to 'protect your health'. Makes sense, right? They are only imposing these things because they care, right? *Right?*

Submitting to insanity

Psychological reframing of the population goes very deep and is done in many less obvious ways. I hear people say how contradictory and crazy 'Covid' rules are and how they are ever changing. This is explained away by dismissing those involved as idiots. It is a big mistake. The Cult is delighted if its cold calculation is perceived as incompetence and idiocy when it is anything but. Oh, yes, there are idiots within the system – lots of them – but they are *administering* the Cult agenda, mostly unknowingly. They are not deciding and dictating it. The bulwark against tyranny is self-respect, always has been, always will be. It is self-respect that has broken every tyranny in history. By its very nature self-respect will not bow to oppression and its perpetrators. There is so little self-respect that it's always the few that overturn dictators. Many may eventually follow, but the few with the iron spines (self-respect) kick it off and generate the momentum. The Cult targets self-respect in the knowledge that once this has gone only submission remains. Crazy, contradictory, ever-changing 'Covid' rules are systematically applied by psychologists to delete self-respect. They *want* you to see that the rules make no sense. It is one thing to decide to do something when *you* have made the choice based on evidence and logic. You still retain your self-respect. It is quite another when you can see what you are being told to do is insane, ridiculous and makes no sense, and *yet you still do it*. Your self-respect is extinguished and this has been happening as ever more obviously stupid and nonsensical things have been demanded and the great majority have complied even when they can see they are stupid and nonsensical.

People walk around in face-nappies knowing they are damaging their health and make no difference to a 'virus'. They do it in fear of not doing it. I know it's daft, but I'll do it anyway. When that happens something dies inside of you and submissive reframing has begun. Next there's a need to hide from yourself that you have conceded your self-respect and you convince yourself that you have not really submitted to fear and intimidation. You begin to believe that you are complying with craziness because it's the right thing to do. When first you concede your self-respect of 2+2 = 4 to 2+2 = 5 you *know* you are compromising your self-respect. Gradually to avoid facing that fact you begin to *believe* that 2+2=5. You have been reframed and I have been watching this process happening in the human psyche on an industrial scale. The Cult is working to break your spirit and one of its major tools in that war is humiliation. I read how former American soldier Bradley Manning (later Chelsea Manning after a sex-change) was treated after being jailed for supplying WikiLeaks with documents exposing the enormity of government and elite mendacity. Manning was isolated in solitary confinement for eight months, put under 24-hour surveillance, forced to hand over clothing before going to bed, and stand naked for every roll call. This is systematic humiliation. The introduction of anal swab 'Covid' tests in China has been done for the same reason to delete self-respect and induce compliant submission. Anal swabs are mandatory for incoming passengers in parts of China and American diplomats have said they were forced to undergo the indignity which would have been calculated humiliation by the Cult-owned Chinese government that has America in its sights.

Government-people: An abusive relationship

Spirit-breaking psychological techniques include giving people hope and apparent respite from tyranny only to take it away again. This happened in the UK during Christmas, 2020, when the psycho-psychologists and their political lackeys announced an easing of restrictions over the holiday only to reimpose them almost immediately on the basis of yet another lie. There is a big psychological difference between getting used to oppression and being given hope of relief only to have that dashed. Psychologists know this and we have seen the technique used repeatedly. Then there is traumatising people before you introduce more extreme regulations that require compliance. A perfect case was the announcement by the dark and sinister Whitty and Vallance in the UK that 'new data' predicted that 4,000 could die every day over the winter of 2020/2021 if we did not lockdown again. I think they call it lying and after traumatising people with that claim out came Jackboot Johnson the next day with new curbs on human freedom. Psychologists know that a frightened and traumatised mind becomes suggestable to submission and behaviour reframing. Underpinning all this has been to make people fearful and suspicious of each other and see themselves as a potential danger to others. In league with deleted self-respect you have the perfect psychological recipe for self-loathing. The relationship between authority and public is now demonstrably the same as that of subservience to an abusive partner. These are signs of an abusive relationship explained by psychologist Leslie Becker-Phelps:

Psychological and emotional abuse: Undermining a partner's self-worth with verbal attacks, name-calling, and belittling. Humiliating the partner in public, unjustly accusing them of having an affair, or interrogating them about their every behavior. Keeping partner confused or off balance by saying they were just kidding or blaming the partner for 'making' them act this way … Feigning in public that they care while turning against them in private. This leads to victims frequently feeling confused, incompetent, unworthy, hopeless, and chronically self-doubting. [Apply these techniques to how governments have treated the population since New Year, 2020, and the parallels are obvious.]

Physical abuse: The abuser might physically harm their partner in a range of ways, such as grabbing, hitting, punching, or shoving them. They might throw objects at them or harm them with a weapon. [Observe the physical harm imposed by masks, lockdown, and so on.]

Threats and intimidation: One way abusers keep their partners in line is by instilling fear. They might be verbally threatening, or give threatening looks or gestures. Abusers often make it known that they are tracking their partner's every move. They might destroy their partner's possessions, threaten to harm them, or threaten to harm their family members. Not surprisingly, victims of this abuse often feel anxiety, fear, and panic. [No words necessary.]

Isolation: Abusers often limit their partner's activities, forbidding them to talk or interact with friends or family. They might limit access to a car or even turn off their phone. All of this might be done by physically holding them against their will, but is often accomplished through psychological abuse and intimidation. The more isolated a person feels, the fewer resources they have to help gain perspective on their situation and to escape from it. [No words necessary.]

Economic abuse: Abusers often make their partners beholden to them for money by controlling access to funds of any kind. They might prevent their partner from getting a job or withhold access to money they earn from a job. This creates financial dependency that makes leaving the relationship very difficult. [See destruction of livelihoods and the proposed meagre 'guaranteed income' so long as you do whatever you are told.]

Using children: An abuser might disparage their partner's parenting skills, tell their children lies about their partner, threaten to take custody of their children, or threaten to harm their children. These tactics instil fear and often elicit compliance. [See reframed social service mafia and how children are being mercilessly abused by the state over 'Covid' while their parents look on too frightened to do anything.]

A further recurring trait in an abusive relationship is the abused blaming themselves for their abuse and making excuses for the abuser. We have the public blaming each other for lockdown abuse by government and many making excuses for the government while attacking those who challenge the government. How often we have heard authorities say that rules are being imposed or reimposed only because people have refused to 'behave' and follow the rules. We don't want to do it – it's *you*.

Renegade Minds are an antidote to all of these things. They will never concede their self-respect no matter what the circumstances. Even when apparent humiliation is heaped upon them they laugh in its face and reflect back the humiliation on the abuser where it belongs. Renegade Minds will never wear masks they know are only imposed to humiliate, suppress and damage both physically and psychologically. Consequences will take care of themselves and they will never break their spirit or cause them to concede to tyranny. UK newspaper columnist Peter Hitchens was one of the few in the mainstream media to speak out against lockdowns and forced vaccinations. He then announced he had taken the jab. He wanted to see family members abroad and he believed vaccine passports were inevitable even though they had not yet been introduced. Hitchens has a questioning and critical mind, but not a Renegade one. If he had no amount of pressure would have made him concede. Hitchens excused his action by saying that the battle has been lost. Renegade Minds never accept defeat when freedom is at stake and even if they are the last one standing the self-respect of not submitting to tyranny is more important than any outcome or any consequence.

That's why Renegade Minds are the only minds that ever changed anything worth changing.

CHAPTER EIGHT

'Reframing' insanity

Insanity is relative. It depends on who has who locked in what cage
Ray Bradbury

'Reframing' a mind means simply to change its perception and behaviour. This can be done subconsciously to such an extent that subjects have no idea they have been 'reframed' while to any observer changes in behaviour and attitudes are obvious.

Human society is being reframed on a ginormous scale since the start of 2020 and here we have the reason why psychologists rather than doctors have been calling the shots. Ask most people who have succumbed to 'Covid' reframing if they have changed and most will say 'no'; but they *have* and fundamentally. The Cult's long-game has been preparing for these times since way back and crucial to that has been to prepare both population and officialdom mentally and emotionally. To use the mind-control parlance they had to reframe the population with a mentality that would submit to fascism and reframe those in government and law enforcement to impose fascism or at least go along with it. The result has been the fact-deleted mindlessness of 'Wokeness' and officialdom that has either enthusiastically or unquestioningly imposed global tyranny demanded by reframed politicians on behalf of psychopathic and deeply evil cultists. 'Cognitive reframing' identifies and challenges the way someone sees the world in the form of situations, experiences and emotions and then restructures those perceptions to view the same set of circumstances in a different way. This can have benefits if the attitudes are personally destructive while on the other side it has the potential for individual and collective mind control which the subject has no idea has even happened.

Cognitive therapy was developed in the 1960s by Aaron T. Beck who was born in Rhode Island in 1921 as the son of Jewish immigrants from the Ukraine. He became interested in the techniques as a treatment for depression. Beck's daughter Judith S. Beck is prominent in the same field and they founded the Beck Institute for Cognitive Behavior Therapy in Philadelphia in 1994. Cognitive reframing, however, began to be used worldwide by those with a very dark agenda. The Cult reframes politicians to change their attitudes and actions until they are completely at odds with what they once appeared to stand for. The same has been happening to government administrators at all levels, law enforcement, military and the human population. Cultists love mind control for two main reasons: It allows them to control what people think, do and say to secure agenda advancement and, by definition, it calms their legendary insecurity and fear of the unexpected. I have studied mind control since the time I travelled America in 1996. I may have been talking to next to no one in terms of an audience in those years, but my goodness did I gather a phenomenal amount of information and knowledge about so many things including the techniques of mind control. I have described this in detail in other books going back to *The Biggest Secret* in 1998. I met a very large number of people recovering from MKUltra and its offshoots and successors and I began to see how these same techniques were being used on the population in general. This was never more obvious than since the 'Covid' hoax began.

Reframing the enforcers

I have observed over the last two decades and more the very clear transformation in the dynamic between the police, officialdom and the public. I tracked this in the books as the relationship mutated from one of serving the public to seeing them as almost the enemy and certainly a lower caste. There has always been a class divide based on income and always been some psychopathic, corrupt, and big-I-am police officers. This was different. Wholesale change was unfolding in the collective dynamic; it was less about money and far more about position and perceived power. An us-and-them was emerging. Noses were lifted skyward by government administration and law enforcement and their attitude to the public they were *supposed* to be serving changed to one of increasing contempt, superiority and control. The transformation was so clear and widespread that it had to be planned. Collective attitudes and dynamics do not change naturally and organically that quickly on that scale. I then came across an organisation in Britain called Common Purpose created in the late 1980s by Julia Middleton who would work in the office of Deputy Prime Minister John Prescott during the long and disastrous premiership of war criminal Tony Blair. When Blair speaks the Cult is speaking and the man should have been in jail a long time ago. Common Purpose proclaims itself to be one of the biggest 'leadership development' organisations in the world while functioning as a *charity* with all the financial benefits which come from that. It hosts 'leadership development' courses and programmes all over the world and claims to have 'brought together' what it calls 'leaders' from more than 100 countries on six continents. The modus operandi of Common Purpose can be compared with the work of the UK government's reframing network that includes the Behavioural Insights Team 'nudge unit' and 'Covid' reframing specialists at SPI-B. WikiLeaks described Common Purpose long ago as 'a hidden virus in our government and schools' which is unknown to the general public: 'It recruits and trains "leaders" to be loyal to the directives of Common Purpose and the EU, instead of to their own departments, which they then undermine or subvert, the NHS [National Health Service] being an example.' This is a vital point to understand the 'Covid' hoax. The NHS, and its equivalent around the world, has been utterly reframed in terms of administrators and much of the medical personnel with the transformation underpinned by recruitment policies. The outcome has been the criminal and psychopathic behaviour of the

NHS over ‘Covid’ and we have seen the same in every other major country. WikiLeaks said Common Purpose trainees are ‘learning to rule without regard to democracy’ and to usher in a police state (current events explained). Common Purpose operated like a ‘glue’ and had members in the NHS, BBC, police, legal profession, church, many of Britain’s 7,000 quangos, local councils, the Civil Service, government ministries and Parliament, and controlled many RDA’s (Regional Development Agencies). Here we have one answer for how and why British institutions and their like in other countries have changed so negatively in relation to the public. This further explains how and why the beyond-disgraceful reframed BBC has become a propaganda arm of ‘Covid’ fascism. They are all part of a network pursuing the same goal.

By 2019 Common Purpose was quoting a figure of 85,000 ‘leaders’ that had attended its programmes. These ‘students’ of all ages are known as Common Purpose ‘graduates’ and they consist of government, state and local government officials and administrators, police chiefs and officers, and a whole range of others operating within the national, local and global establishment. Cressida Dick, Commissioner of the London Metropolitan Police, is the Common Purpose graduate who was the ‘Gold Commander’ that oversaw what can only be described as the murder of Brazilian electrician Jean Charles de Menezes in 2005. He was held down by psychopathic police and shot seven times in the head by a psychopathic lunatic after being mistaken for a terrorist when he was just a bloke going about his day. Dick authorised officers to pursue and keep surveillance on de Menezes and ordered that he be stopped from entering the underground train system. Police psychopaths took her at her word clearly. She was ‘disciplined’ for this outrage by being *promoted* – eventually to the top of the ‘Met’ police where she has been a disaster. Many Chief Constables controlling the police in different parts of the UK are and have been Common Purpose graduates. I have heard the ‘graduate’ network described as a sort of Mafia or secret society operating within the fabric of government at all levels pursuing a collective policy ingrained at Common Purpose training events. Founder Julia Middleton herself has said:

> Locally and internationally, Common Purpose graduates will be 'lighting small fires' to create change in their organisations and communities … The Common Purpose effect is best illustrated by the many stories of small changes brought about by leaders, who themselves have changed.

A Common Purpose mission statement declared:

> Common Purpose aims to improve the way society works by expanding the vision, decision-making ability and influence of all kinds of leaders. The organisation runs a variety of educational programmes for leaders of all ages, backgrounds and sectors, in order to provide them with the inspirational, information and opportunities they need to change the world.

Yes, but into what? Since 2020 the answer has become clear.

NLP and the Delphi technique

Common Purpose would seem to be a perfect name or would common programming be better? One of the foundation methods of reaching 'consensus' (group think) is by setting the agenda theme and then encouraging, cajoling or pressuring everyone to agree a 'consensus' in line with the core theme promoted by Common Purpose. The methodology involves the 'Delphi technique', or an adaption of it, in which opinions are expressed that are summarised by a 'facilitator or change agent' at each stage. Participants are 'encouraged' to modify their views in the light of what others have said. Stage by stage the former individual opinions are merged into group consensus which just happens to be what Common Purpose wants them to believe. A key part of this is to marginalise anyone refusing to concede to group think and turn the group against them to apply pressure to conform. We are seeing this very technique used on the general population to make 'Covid' group-thinkers hostile to those who have seen through the bullshit. People can be reframed by using perception manipulation methods such as Neuro-Linguistic Programming (NLP) in which you change perception with the use of carefully constructed language. An NLP website described the technique this way:

> … A method of influencing brain behaviour (the 'neuro' part of the phrase) through the use of language (the 'linguistic' part) and other types of communication to enable a person to 'recode' the way the brain responds to stimuli (that's the 'programming') and manifest new and better behaviours. Neuro-Linguistic Programming often incorporates hypnosis and self-hypnosis to help achieve the change (or 'programming') that is wanted.

British alternative media operation UKColumn has done very detailed research into Common Purpose over a long period. I quoted co-founder and former naval officer Brian Gerrish in my book *Remember Who You Are*, published in 2011, as saying the following years before current times:

> It is interesting that many of the mothers who have had children taken by the State speak of the Social Services people being icily cool, emotionless and, as two ladies said in slightly different words, '… like little robots'. We know that NLP is cumulative, so people can be given small imperceptible doses of NLP in a course here, another in a few months, next year etc. In this way, major changes are accrued in their personality, but the day by day change is almost unnoticeable.

In these and other ways 'graduates' have had their perceptions uniformly reframed and they return to their roles in the institutions of government, law enforcement, legal profession, military, 'education', the UK National Health Service and the whole swathe of the establishment structure to pursue a common agenda preparing for the 'post-industrial', 'post-democratic' society. I say 'preparing' but we are now there. 'Post-industrial' is code for the Great Reset and 'post-democratic' is 'Covid' fascism. UKColumn has spoken to partners of those who have attended Common Purpose 'training'. They have described how personalities and attitudes of 'graduates' changed very noticeably for the worse by the time they had completed the course. They had been 'reframed' and told they are the 'leaders' – the special ones – who know better than the population. There has also been the very demonstrable recruitment of psychopaths and narcissists into government administration at all levels and law enforcement. If you want psychopathy hire psychopaths and you get a simple cause and effect. If you want administrators, police officers and 'leaders' to perceive the public as lesser beings who don't matter then employ narcissists. These personalities are identified using 'psychometrics' that identifies knowledge, abilities, attitudes and personality traits, mostly through carefully-designed questionnaires and tests. As this policy has passed through the decades we have had power-crazy, power-trippers appointed into law enforcement, security and government administration in preparation for current times and the dynamic between public and law enforcement/officialdom has been transformed. UKColumn's Brian Gerrish said of the narcissistic personality:

> Their love of themselves and power automatically means that they will crush others who get in their way. I received a major piece of the puzzle when a friend pointed out that when they made public officials re-apply for their own jobs several years ago they were also required to do psychometric tests. This was undoubtedly the start of the screening process to get 'their' sort of people in post.

How obvious that has been since 2020 although it was clear what was happening long before if people paid attention to the changing public-establishment dynamic.

Change agents

At the centre of events in 'Covid' Britain is the National Health Service (NHS) which has behaved disgracefully in slavishly following the Cult agenda. The NHS management structure is awash with Common Purpose graduates or 'change agents' working to a common cause. Helen Bevan, a Chief of Service Transformation at the NHS Institute for Innovation and Improvement, co-authored a document called 'Towards a million change agents, a review of the social movements literature: implications for large scale change in the NHS'. The document compared a project management approach to that of change and social movements where 'people change themselves and each other – peer to peer'. Two definitions given for a 'social movement' were:

A group of people who consciously attempt to build a radically new social order; involves people of a broad range of social backgrounds; and deploys politically confrontational and socially disruptive tactics – Cyrus Zirakzadeh 1997

Collective challenges, based on common purposes and social solidarities, in sustained interaction with elites, opponents, and authorities – Sidney Tarrow 1994

Helen Bevan wrote another NHS document in which she defined 'framing' as 'the process by which leaders construct, articulate and put across their message in a powerful and compelling way in order to win people to their cause and call them to action'. I think I could come up with another definition that would be rather more accurate. The National Health Service and institutions of Britain and the wider world have been taken over by reframed 'change agents' and that includes everything from the United Nations to national governments, local councils and social services which have been kidnapping children from loving parents on an extraordinary and gathering scale on the road to the end of parenthood altogether. Children from loving homes are stolen and kidnapped by the state and put into the 'care' (inversion) of the local authority through council homes, foster parents and forced adoption. At the same time children are allowed to be abused without response while many are under council 'care'. UKColumn highlighted the Common Purpose connection between South Yorkshire Police and Rotherham council officers in the case of the scandal in that area of the sexual exploitation of children to which the authorities turned not one blind eye, but both:

> We were alarmed to discover that the Chief Executive, the Strategic Director of Children and Young People's Services, the Manager for the Local Strategic Partnership, the Community Cohesion Manager, the Cabinet Member for Cohesion, the Chief Constable and his predecessor had all attended Leadership training courses provided by the pseudo-charity Common Purpose.

Once 'change agents' have secured positions of hire and fire within any organisation things start to move very quickly. Personnel are then hired and fired on the basis of whether they will work towards the agenda the change agent represents. If they do they are rapidly promoted even though they may be incompetent. Those more qualified and skilled who are pre-Common Purpose 'old school' see their careers stall and even disappear. This has been happening for decades in every institution of state, police, 'health' and social services and all of them have been transformed as a result in their attitudes to their jobs and the public. Medical professions, including nursing, which were once vocations for the caring now employ many cold, callous and couldn't give a shit personality types. The UKColumn investigation concluded:

> By blurring the boundaries between people, professions, public and private sectors, responsibility and accountability, Common Purpose encourages 'graduates' to believe that as new selected leaders, they can work together, outside of the established political and social structures, to achieve a paradigm shift or CHANGE – so called 'Leading Beyond Authority'. In doing so, the allegiance of the individual becomes 'reframed' on CP colleagues and their NETWORK.

Reframing the Face-Nappies

Nowhere has this process been more obvious than in the police where recruitment of psychopaths and development of unquestioning mind-controlled group-thinkers have transformed law enforcement into a politically-correct 'Woke' joke and a travesty of what should be public service. Today they wear their face-nappies like good little gofers and enforce 'Covid' rules which are fascism under another name. Alongside the specifically-recruited psychopaths we have software minds incapable of free thought. Brian Gerrish again:

An example is the policeman who would not get on a bike for a press photo because he had not done the cycling proficiency course. Normal people say this is political correctness gone mad. Nothing could be further from the truth. The policeman has been reframed, and in his reality it is perfect common sense not to get on the bike 'because he hasn't done the cycling course'.

Another example of this is where the police would not rescue a boy from a pond until they had taken advice from above on the 'risk assessment'. A normal person would have arrived, perhaps thought of the risk for a moment, and dived in. To the police now 'reframed', they followed 'normal' procedure.

There are shocking cases of reframed ambulance crews doing the same. Sheer unthinking stupidity of London Face-Nappies headed by Common Purpose graduate Cressida Dick can be seen in their behaviour at a vigil in March, 2021, for a murdered woman, Sarah Everard. A police officer had been charged with the crime. Anyone with a brain would have left the vigil alone in the circumstances. Instead they 'manhandled' women to stop them breaking 'Covid rules' to betray classic reframing. Minds in the thrall of perception control have no capacity for seeing a situation on its merits and acting accordingly. 'Rules is rules' is their only mind-set. My father used to say that rules and regulations are for the guidance of the intelligent and the blind obedience of the idiot. Most of the intelligent, decent, coppers have gone leaving only the other kind and a few old school for whom the job must be a daily nightmare. The combination of psychopaths and rule-book software minds has been clearly on public display in the 'Covid' era with automaton robots in uniform imposing fascistic 'Covid' regulations on the population without any personal initiative or judging situations on their merits. There are thousands of examples around the world, but I'll make my point with the infamous Derbyshire police in the English East Midlands – the ones who think pouring dye into beauty spots and using drones to track people walking in the countryside away from anyone is called 'policing'. To them there are rules decreed by the government which they have to enforce and in their bewildered state a group gathering in a closed space and someone walking alone in the countryside are the same thing. It is beyond idiocy and enters the realm of clinical insanity.

Police officers in Derbyshire said they were 'horrified' – *horrified* – to find 15 to 20 'irresponsible' kids playing a football match at a closed leisure centre 'in breach of coronavirus restrictions'. When they saw the police the kids ran away leaving their belongings behind and the reframed men and women of Derbyshire police were seeking to establish their identities with a view to fining their parents. The most natural thing for youngsters to do – kicking a ball about – is turned into a criminal activity and enforced by the moronic software programs of Derbyshire police. You find the same mentality in every country. These barely conscious 'horrified' officers said they had to take action because 'we need to ensure these rules are being followed' and 'it is of the utmost importance that you ensure your children are following the rules and regulations for Covid-19'. Had any of them done ten seconds of research to see if this parroting of their masters' script could be supported by any evidence? Nope. Reframed people don't think – others think for them and that's the whole idea of reframing. I have seen police officers one after the other repeating without question word for word what officialdom tells them just as I have seen great swathes of the public doing the same. Ask either for 'their' opinion and out spews what they have been told to think by the official narrative. Police and public may seem to be in different groups, but their mentality is the same. Most people do whatever they are told in fear not doing so or because they believe what officialdom tells them; almost the entirety of the police do what they are told for the same reason. Ultimately it's the tiny inner core of the global Cult that's telling both what to do.

So Derbyshire police were 'horrified'. Oh, really? Why did they think those kids were playing football? It was to relieve the psychological consequences of lockdown and being denied human contact with their friends and interaction, touch and discourse vital to human psychological health. Being denied this month after month has dismantled the psyche of many children and young people as depression and suicide have exploded. Were Derbyshire police *horrified by that*? Are you kidding? Reframed people don't have those mental and emotional processes that can see how the impact on the psychological health of youngsters is far more dangerous than any ‘virus’ even if you take the mendacious official figures to be true. The reframed are told (programmed) how to act and so they do. The Derbyshire Chief Constable in the first period of lockdown when the black dye and drones nonsense was going on was Peter Goodman. He was the man who severed the connection between his force and the Derbyshire Constabulary *Male Voice* Choir when he decided that it was not inclusive enough to allow women to join. The fact it was a male voice choir making a particular sound produced by male voices seemed to elude a guy who terrifyingly ran policing in Derbyshire. He retired weeks after his force was condemned as disgraceful by former Supreme Court Justice Jonathan Sumption for their behaviour over extreme lockdown impositions. Goodman was replaced by his deputy Rachel Swann who was in charge when her officers were ‘horrified’. The police statement over the boys committing the hanging-offence of playing football included the line about the youngsters being ‘irresponsible in the times we are all living through’ missing the point that the real relevance of the ‘times we are all living through’ is the imposition of fascism enforced by psychopaths and reframed minds of police officers playing such a vital part in establishing the fascist tyranny that their own children and grandchildren will have to live in their entire lives. As a definition of insanity that is hard to beat although it might be run close by imposing masks on people that can have a serious effect on their health while wearing a face nappy all day themselves. Once again public and police do it for the same reason – the authorities tell them to and who are they to have the self-respect to say no?

Wokers in uniform

How reframed do you have to be to arrest a *six-year-old* and take him to court for *picking a flower* while waiting for a bus? Brain dead police and officialdom did just that in North Carolina where criminal proceedings happen regularly for children under nine. Attorney Julie Boyer gave the six-year-old crayons and a colouring book during the ‘flower’ hearing while the ‘adults’ decided his fate. County Chief District Court Judge Jay Corpening asked: ‘Should a child that believes in Santa Claus, the Easter Bunny and the tooth fairy be making life-altering decisions?’ Well, of course not, but common sense has no meaning when you have a common purpose and a reframed mind. Treating children in this way, and police operating in American schools, is all part of the psychological preparation for children to accept a police state as normal all their adult lives. The same goes for all the cameras and biometric tracking technology in schools. Police training is focused on reframing them as snowflake Wokers and this is happening in the military. Pentagon top brass said that ‘training sessions on extremism’ were needed for troops who asked why they were so focused on the Capitol Building riot when Black Lives Matter riots were ignored. What’s the difference between them some apparently and rightly asked. Actually, there is a difference. Five people died in the Capitol riot, only one through violence, and that was a police officer shooting an unarmed protestor. BLM riots killed at least 25 people and cost billions. Asking the question prompted the psychopaths and reframed minds that run the Pentagon to say that more ‘education’ (programming) was needed. Troop training is all based on psychological programming to make them fodder for the Cult – ‘Military men are just dumb, stupid animals to be used as pawns in foreign policy’ as Cult-to-his-DNA former Secretary of State Henry Kissinger famously said. Governments see the police in similar terms and it’s time for those among them who can see this to defend the people and stop being enforcers of the Cult agenda upon the people.

The US military, like the country itself, is being targeted for destruction through a long list of Woke impositions. Cult-owned gaga ‘President’ Biden signed an executive order when he took office to allow taxpayer money to pay for transgender surgery for active military personnel and veterans. Are you a man soldier? No, I’m a LGBTQIA+ with a hint of Skoliosexual and Spectrasexual. Oh, good man. Bad choice of words you bigot. The Pentagon announced in March, 2021, the appointment of the first ‘diversity and inclusion officer' for US Special Forces. Richard Torres-Estrada arrived with the publication of a 'D&I Strategic Plan which will guide the enterprise-wide effort to institutionalize and sustain D&I'. If you think a Special Forces 'Strategic Plan' should have something to do with defending America you haven't been paying attention. Defending Woke is now the military's new role. Torres-Estrada has posted images comparing Donald Trump with Adolf Hitler and we can expect no bias from him as a representative of the supposedly non-political Pentagon. Cable news host Tucker Carlson said: 'The Pentagon is now the Yale faculty lounge but with cruise missiles.' Meanwhile Secretary of Defense Lloyd Austin, a board member of weapons-maker Raytheon with stock and compensation interests in October, 2020, worth $1.4 million, said he was purging the military of the 'enemy within' – anyone who isn't Woke and supports Donald Trump. Austin refers to his targets as 'racist extremists' while in true Woke fashion being himself a racist extremist. Pentagon documents pledge to 'eradicate, eliminate and conquer all forms of racism, sexism and homophobia'. The definitions of these are decided by 'diversity and inclusion committees' peopled by those who see racism, sexism and homophobia in every situation and opinion. Woke (the Cult) is dismantling the US military and purging testosterone as China expands its military and gives its troops 'masculinity training'. How do we think that is going to end when this is all Cult coordinated? The US military, like the British military, is controlled by Woke and spineless top brass who just go along with it out of personal career interests.

'Woke' means fast asleep

Mind control and perception manipulation techniques used on individuals to create group-think have been unleashed on the global population in general. As a result many have no capacity to see the obvious fascist agenda being installed all around them or what 'Covid' is really all about. Their brains are firewalled like a computer system not to process certain concepts, thoughts and realisations that are bad for the Cult. The young are most targeted as the adults they will be when the whole fascist global state is planned to be fully implemented. They need to be prepared for total compliance to eliminate all pushback from entire generations. The Cult has been pouring billions into taking complete control of ‘education’ from schools to universities via its operatives and corporations and not least Bill Gates as always. The plan has been to transform ‘education’ institutions into programming centres for the mentality of ‘Woke’. James McConnell, professor of psychology at the University of Michigan, wrote in *Psychology Today* in 1970:

> The day has come when we can combine sensory deprivation with drugs, hypnosis, and astute manipulation of reward and punishment, to gain almost absolute control over an individual’s behaviour. It should then be possible to achieve a very rapid and highly effective type of brainwashing that would allow us to make dramatic changes in a person’s behaviour and personality ...
>
> … We should reshape society so that we all would be trained from birth to want to do what society wants us to do. We have the techniques to do it... no-one owns his own personality you acquired, and there’s no reason to believe you should have the right to refuse to acquire a new personality if your old one is anti-social.

This was the potential for mass brainwashing in 1970 and the mentality there displayed captures the arrogant psychopathy that drives it forward. I emphasise that not all young people have succumbed to Woke programming and those that haven’t are incredibly impressive people given that today’s young are the most perceptually-targeted generations in history with all the technology now involved. Vast swathes of the young generations, however, have fallen into the spell – and that’s what it is – of Woke. The Woke mentality and perceptual program is founded on *inversion* and you will appreciate later why that is so significant. Everything with Woke is inverted and the opposite of what it is claimed to be. Woke was a term used in African-American culture from the 1900s and referred to an awareness of social and racial justice. This is not the meaning of the modern version or ‘New Woke’ as I call it in *The Answer*. Oh, no, Woke today means something very different no matter how much Wokers may seek to hide that and insist Old Woke and New

Woke are the same. See if you find any ‘awareness of social justice’ here in the modern variety:

- Woke demands ‘inclusivity’ while excluding anyone with a different opinion and calls for mass censorship to silence other views.
- Woke claims to stand against oppression when imposing oppression is the foundation of all that it does. It is the driver of political correctness which is nothing more than a Cult invention to manipulate the population to silence itself.
- Woke believes itself to be ‘liberal’ while pursuing a global society that can only be described as fascist (see ‘anti-fascist’ fascist Antifa).
- Woke calls for ‘social justice’ while spreading injustice wherever it goes against the common ‘enemy’ which can be easily identified as a differing view.
- Woke is supposed to be a metaphor for ‘awake’ when it is solid-gold asleep and deep in a Cult-induced coma that meets the criteria for ‘off with the fairies’.

I state these points as obvious facts if people only care to look. I don’t do this with a sense of condemnation. We need to appreciate that the onslaught of perceptual programming on the young has been incessant and merciless. I can understand why so many have been reframed, or, given their youth, framed from the start to see the world as the Cult demands. The Cult has had access to their minds day after day in its ‘education’ system for their entire formative years. Perception is formed from information received and the Cult-created system is a life-long download of information delivered to elicit a particular perception, thus behaviour. The more this has expanded into still new extremes in recent decades and ever-increasing censorship has deleted other opinions and information why wouldn’t that lead to a perceptual reframing on a mass scale? I have described already cradle-to-grave programming and in more recent times the targeting of young minds from birth to adulthood has entered the stratosphere. This has taken the form of skewing what is ‘taught’ to fit the Cult agenda and the omnipresent techniques of group-think to isolate non-believers and pressure them into line. There has always been a tendency to follow the herd, but we really are in a new world now in relation to that. We have parents who can see the ‘Covid’ hoax told by their children not to stop them wearing masks at school, being ‘Covid’ tested or having the ‘vaccine’ in fear of the peer-pressure consequences of being different. What is ‘peer-pressure’ if not pressure to conform to group-think? Renegade Minds never group-think and always retain a set of perceptions that are unique to them. Group-think is always underpinned by consequences for not group-thinking. Abuse now aimed at those refusing DNA-manipulating ‘Covid vaccines’ are a potent example of this. The biggest pressure to conform comes from the very group which is itself being manipulated. ‘I am programmed to be part of a hive mind and so you must be.’

Woke control structures in ‘education’ now apply to every mainstream organisation. Those at the top of the ‘education’ hierarchy (the Cult) decide the policy. This is imposed on governments through the Cult network; governments impose it on schools, colleges and universities; their leadership impose the policy on teachers and academics and they impose it on children and students. At any level where there is resistance, perhaps from a teacher or university lecturer, they are targeted by the authorities and often fired. Students themselves regularly demand the dismissal of academics (increasingly few) at odds with the narrative that the students have been programmed to believe in. It is quite a thought that students who are being targeted by the Cult become so consumed by programmed group-think that they launch protests and demand the removal of those who are trying to push back against those targeting the students. Such is the scale of perceptual inversion. We see this with ‘Covid’ programming as the Cult imposes the rules via psycho-psychologists and governments on shops, transport companies and businesses which impose them on their staff who impose them on their customers who pressure Pushbackers to conform to the will of the Cult which is in the process of destroying them and their families. Scan all aspects of society and you will see the same sequence every time.

Fact free Woke and hijacking the 'left'

There is no more potent example of this than 'Woke', a mentality only made possible by the deletion of factual evidence by an 'education' system seeking to produce an ever more uniform society. Why would you bother with facts when you don't know any? Deletion of credible history both in volume and type is highly relevant. Orwell said: 'Who controls the past controls the future: who controls the present controls the past.' They who control the perception of the past control the perception of the future and they who control the present control the perception of the past through the writing and deleting of history. Why would you oppose the imposition of Marxism in the name of Wokeism when you don't know that Marxism cost at least 100 million lives in the 20th century alone? Watch videos and read reports in which Woker generations are asked basic historical questions – it's mind-blowing. A survey of 2,000 people found that six percent of millennials (born approximately early1980s to early 2000s) believed the Second World War (1939-1945) broke out with the assassination of President Kennedy (in 1963) and one in ten thought Margaret Thatcher was British Prime Minister at the time. She was in office between 1979 and 1990. We are in a post-fact society. Provable facts are no defence against the fascism of political correctness or Silicon Valley censorship. Facts don't matter anymore as we have witnessed with the 'Covid' hoax. Sacrificing uniqueness to the Woke group-think religion is all you are required to do and that means thinking for yourself is the biggest Woke no, no. All religions are an expression of group-think and censorship and Woke is just another religion with an orthodoxy defended by group-think and censorship. Burned at the stake becomes burned on Twitter which leads back eventually to burned at the stake as Woke humanity regresses to ages past.

The biggest Woke inversion of all is its creators and funders. I grew up in a traditional left of centre political household on a council estate in Leicester in the 1950s and 60s – you know, the left that challenged the power of wealth-hoarding elites and threats to freedom of speech and opinion. In those days students went on marches defending freedom of speech while today's Wokers march for its deletion. What on earth could have happened? Those very elites (collectively the Cult) that we opposed in my youth and early life have funded into existence the antithesis of that former left and hijacked the 'brand' while inverting everything it ever stood for. We have a mentality that calls itself 'liberal' and 'progressive' while acting like fascists. Cult billionaires and their corporations have funded themselves into control of 'education' to ensure that Woke programming is unceasing throughout the formative years of children and young people and that non-Wokers are isolated (that word again) whether they be students, teachers or college professors. The Cult has funded into existence the now colossal global network of Woke organisations that have spawned and promoted all the 'causes' on the Cult wish-list for global transformation and turned Wokers into demanders of them. Does anyone really think it's a coincidence that the Cult agenda for humanity is a carbon (sorry) copy of the societal transformations desired by Woke?? These are only some of them:

Political correctness: The means by which the Cult deletes all public debates that it knows it cannot win if we had the free-flow of information and evidence.

Human-caused 'climate change': The means by which the Cult seeks to transform society into a globally-controlled dictatorship imposing its will over the fine detail of everyone's lives 'to save the planet' which doesn't actually need saving.

Transgender obsession: Preparing collective perception to accept the 'new human' which would not have genders because it would be created technologically and not through procreation. I'll have much more on this in Human 2.0.

Race obsession: The means by which the Cult seeks to divide and rule the population by triggering racial division through the perception that society is more racist than ever when the opposite is the case. Is it perfect in that regard? No. But to compare today with the racism of apartheid and segregation brought to an end by the civil rights movement in the 1960s is to insult the memory of that movement and inspirations like Martin Luther King. Why is the 'anti-racism' industry (which it is) so dominated by privileged white people?

White supremacy: This is a label used by privileged white people to demonise poor and deprived white people pushing back on tyranny to marginalise and destroy them. White people are being especially targeted as the dominant race by number within Western society which the Cult seeks to transform in its image. If you want to change a society you must weaken and undermine its biggest group and once you have done that by using the other groups you next turn on them to do the same … 'Then they came for the Jews and I was not a Jew so I did nothing.'

Mass migration: The mass movement of people from the Middle East, Africa and Asia into Europe, from the south into the United States and from Asia into Australia are another way the Cult seeks to dilute the racial, cultural and political influence of white people on Western society. White people ask why their governments appear to be working against them while being politically and culturally biased towards incoming cultures. Well, here's your answer. In the same way sexually 'straight' people, men and women, ask why the authorities are biased against them in favour of other sexualities. The answer is the same – that's the way the Cult wants it to be for very sinister motives.

These are all central parts of the Cult agenda and central parts of the Woke agenda and Woke was created and continues to be funded to an immense degree by Cult billionaires and corporations. If anyone begins to say 'coincidence' the syllables should stick in their throat.

Billionaire 'social justice warriors'

Joe Biden is a 100 percent-owned asset of the Cult and the Wokers' man in the White House whenever he can remember his name and for however long he lasts with his rapidly diminishing cognitive function. Even walking up the steps of an aircraft without falling on his arse would appear to be a challenge. He's not an empty-shell puppet or anything. From the minute Biden took office (or the Cult did) he began his executive orders promoting the Woke wish-list. You will see the Woke agenda imposed ever more severely because it's really the *Cult* agenda. Woke organisations and activist networks spawned by the Cult are funded to the extreme so long as they promote what the Cult wants to happen. Woke is funded to promote 'social justice' by billionaires who become billionaires by destroying social justice. The social justice mantra is only a cover for dismantling social justice and funded by billionaires that couldn't give a damn about social justice. Everything makes sense when you see that. One of Woke's premier funders is Cult billionaire financier George Soros who said: 'I am basically there to make money, I cannot and do not look at the social consequences of what I do.' This is the same Soros who has given more than $32 billion to his Open Society Foundations global Woke network and funded Black Lives Matter, mass immigration into Europe and the United States, transgender activism, climate change activism, political correctness and groups targeting 'white supremacy' in the form of privileged white thugs that dominate Antifa. What a scam it all is and when you are dealing with the unquestioning fact-free zone of Woke scamming them is child's play. All you need to pull it off in all these organisations are a few in-the-know agents of the Cult and an army of naïve, reframed, uninformed, narcissistic, know-nothings convinced of their own self-righteousness, self-purity and virtue.

Soros and fellow billionaires and billionaire corporations have poured hundreds of millions into Black Lives Matter and connected groups and promoted them to a global audience. None of this is motivated by caring about black people. These are the billionaires that have controlled and exploited a system that leaves millions of black people in abject poverty and deprivation which they do absolutely nothing to address. The same Cult networks funding BLM were behind the *slave trade!* Black Lives Matter hijacked a phrase that few would challenge and they have turned this laudable concept into a political weapon to divide society. You know that BLM is a fraud when it claims that *All* Lives Matter, the most inclusive statement of all, is 'racist'. BLM and its Cult masters don't want to end racism. To them it's a means to an end to control all of humanity never mind the colour, creed, culture or background. What has destroying the nuclear family got to do with ending racism? Nothing – but that is one of the goals of BLM and also happens to be a goal of the Cult as I have been exposing in my books for decades. Stealing children from loving parents and giving schools ever more power to override parents is part of that same agenda. BLM is a Marxist organisation and why would that not be the case when the Cult created Marxism *and* BLM? Patrisse Cullors, a BLM co-founder, said in a 2015 video that she and her fellow organisers, including co-founder Alicia Garza, are 'trained Marxists'. The lady known after marriage as Patrisse Khan-Cullors bought a $1.4 million home in 2021 in one of the whitest areas of California with a black population of just 1.6 per cent and has so far bought *four* high-end homes for a total of $3.2 million. How very Marxist. There must be a bit of spare in the BLM coffers, however, when Cult corporations and billionaires have handed over the best part of $100 million. Many black people can see that Black Lives Matter is not working for them, but against them, and this is still more confirmation. Black journalist Jason Whitlock, who had his account suspended by Twitter for simply linking to the story about the 'Marxist's' home buying spree, said that BLM leaders are 'making millions of dollars off the backs of these dead black men who they wouldn't spit on if they were on fire and alive'.

Black Lies Matter

Cult assets and agencies came together to promote BLM in the wake of the death of career criminal George Floyd who had been jailed a number of times including for forcing his way into the home of a black woman with others in a raid in which a gun was pointed at her stomach. Floyd was filmed being held in a Minneapolis street in 2020 with the knee of a police officer on his neck and he subsequently died. It was an appalling thing for the officer to do, but the same technique has been used by police on peaceful protestors of lockdown without any outcry from the Woke brigade. As unquestioning supporters of the Cult agenda Wokers have supported lockdown and all the 'Covid' claptrap while attacking anyone standing up to the tyranny imposed in its name. Court documents would later include details of an autopsy on Floyd by County Medical Examiner Dr Andrew Baker who concluded that Floyd had taken a fatal level of the drug fentanyl. None of this mattered to fact-free, question-free, Woke. Floyd's death was followed by worldwide protests against police brutality amid calls to defund the police. Throwing babies out with the bathwater is a Woke speciality. In the wake of the murder of British woman Sarah Everard a Green Party member of the House of Lords, Baroness Jones of Moulescoomb (Nincompoopia would have been better), called for a 6pm curfew for all men. This would be in breach of the Geneva Conventions on war crimes which ban collective punishment, but that would never have crossed the black and white Woke mind of Baroness Nincompoopia who would have been far too convinced of her own self-righteousness to compute such details. Many American cities did defund the police in the face of Floyd riots and after $15 million was deleted from the police budget in Washington DC under useless Woke mayor Muriel Bowser car-jacking alone rose by 300 percent and within six months the US capital recorded its highest murder rate in 15 years. The same happened in Chicago and other cities in line with the Cult/Soros plan to bring fear to streets and neighbourhoods by reducing the police, releasing violent criminals and not prosecuting crime. This is the mob-rule agenda that I have warned in the books was coming for so long. Shootings in the area of Minneapolis where Floyd was arrested increased by 2,500 percent compared with the year before. Defunding the police over George Floyd has led to a big increase in dead people with many of them black. Police protection for politicians making these decisions stayed the same or increased as you would expect from professional hypocrites. The Cult doesn't actually want to abolish the police. It wants to abolish local control over the police and hand it to federal government as the psychopaths advance the Hunger Games Society. Many George Floyd protests turned into violent riots with black stores and businesses destroyed by fire and looting across America fuelled by Black Lives Matter. Woke doesn't do irony. If you want civil rights you must loot the liquor store and the supermarket and make off with a smart TV. It's the only way.

It's not a race war – it's a class war

Black people are patronised by privileged blacks and whites alike and told they are victims of white supremacy. I find it extraordinary to watch privileged blacks supporting the very system and bloodline networks behind the slave trade and parroting the same Cult-serving manipulative crap of their privileged white, often billionaire, associates. It is indeed not a race war but a class war and colour is just a diversion. Black Senator Cory Booker and black Congresswoman Maxine Waters, more residents of Nincompoopia, personify this. Once you tell people they are victims of someone else you devalue both their own responsibility for their plight and the power they have to impact on their reality and experience. Instead we have: ‘You are only in your situation because of whitey – turn on them and everything will change.’ It won’t change. Nothing changes in our lives unless *we* change it. Crucial to that is never seeing yourself as a victim and always as the creator of your reality. Life is a simple sequence of choice and consequence. Make different choices and you create different consequences. *You* have to make those choices – not Black Lives Matter, the Woke Mafia and anyone else that seeks to dictate your life. Who are they these Wokers, an emotional and psychological road traffic accident, to tell you what to do? Personal empowerment is the last thing the Cult and its Black Lives Matter want black people or anyone else to have. They claim to be defending the underdog while *creating* and perpetuating the underdog. The Cult’s worst nightmare is human unity and if they are going to keep blacks, whites and every other race under economic servitude and control then the focus must be diverted from what they have in common to what they can be manipulated to believe divides them. Blacks have to be told that their poverty and plight is the fault of the white bloke living on the street in the same poverty and with the same plight they are experiencing. The difference is that your plight black people is due to him, a white supremacist with ‘white privilege’ living on the street. Don’t unite as one human family against your mutual oppressors and suppressors – fight the oppressor with the white face who is as financially deprived as you are. The Cult knows that as its ‘Covid’ agenda moves into still new levels of extremism people are going to respond and it has been spreading the seeds of disunity everywhere to stop a united response to the evil that targets *all of us*.

Racist attacks on ‘whiteness’ are getting ever more outrageous and especially through the American Democratic Party which has an appalling history for anti-black racism. Barack Obama, Joe Biden, Hillary Clinton and Nancy Pelosi all eulogised about Senator Robert Byrd at his funeral in 2010 after a nearly 60-year career in Congress. Byrd was a brutal Ku Klux Klan racist and a violent abuser of Cathy O’Brien in MKUltra. He said he would never fight in the military ‘with a negro by my side’ and ‘rather I should die a thousand times, and see Old Glory trampled in the dirt never to rise again, than to see this beloved land of ours become degraded by race mongrels, a throwback to the blackest specimen from the wilds'. Biden called Byrd a 'very close friend and mentor'. These 'Woke' hypocrites are not anti-racist they are anti-poor and anti-people not of their perceived class. Here is an illustration of the scale of anti-white racism to which we have now descended. Seriously Woke and moronic *New York Times* contributor Damon Young described whiteness as a 'virus' that 'like other viruses will not die until there are no bodies left for it to infect'. He went on: '… the only way to stop it is to locate it, isolate it, extract it, and kill it.' Young can say that as a black man with no consequences when a white man saying the same in reverse would be facing a jail sentence. *That's* racism. We had super-Woke numbskull senators Tammy Duckworth and Mazie Hirono saying they would object to future Biden Cabinet appointments if he did not nominate more Asian Americans and Pacific Islanders. Never mind the ability of the candidate what do they look like? Duckworth said: 'I will vote for racial minorities and I will vote for LGBTQ, but anyone else I'm not voting for.' Appointing people on the grounds of race is illegal, but that was not a problem for this ludicrous pair. They were on-message and that's a free pass in any situation.

Critical race racism

White children are told at school they are intrinsically racist as they are taught the divisive 'critical race theory'. This claims that the law and legal institutions are inherently racist and that race is a socially constructed concept used by white people to further their economic and political interests at the expense of people of colour. White is a 'virus' as we've seen. Racial inequality results from 'social, economic, and legal differences that white people create between races to maintain white interests which leads to poverty and criminality in minority communities'. I must tell that to the white guy sleeping on the street. The principal of East Side Community School in New York sent white parents a manifesto that called on them to become ‘white traitors’ and advocate for full ‘white abolition’. These people are teaching your kids when they urgently need a psychiatrist. The ‘school’ included a chart with ‘eight white identities’ that ranged from ‘white supremacist’ to ‘white abolition’ and defined the behaviour white people must follow to end ‘the regime of whiteness’. Woke blacks and their privileged white associates are acting exactly like the slave owners of old and Ku Klux Klan racists like Robert Byrd. They are too full of their own self-purity to see that, but it’s true. Racism is not a body type; it’s a state of mind that can manifest through any colour, creed or culture.

Another racial fraud is *‘equity’*. Not equality of treatment and opportunity – equity. It’s a term spun as equality when it means something very different. Equality in its true sense is a raising up while ‘equity’ is a race to the bottom. Everyone in the same level of poverty is ‘equity’. Keep everyone down – that’s equity. The Cult doesn’t want anyone in the human family to be empowered and BLM leaders, like all these ‘anti-racist’ organisations, continue their privileged, pampered existence by perpetuating the perception of gathering racism. When is the last time you heard an ‘anti-racist’ or ‘anti-Semitism’ organisation say that acts of racism and discrimination have *fallen?* It’s not in the interests of their fund-raising and power to influence and the same goes for the professional soccer anti-racism operation, Kick It Out. Two things confirmed that the Black Lives Matter riots in the summer of 2020 were Cult creations. One was that while anti-lockdown protests were condemned in this same period for ‘transmitting ‘Covid’ the authorities supported mass gatherings of Black Lives Matter supporters. I even saw self-deluding people claiming to be doctors say the two types of protest were not the same. No – the non-existent ‘Covid’ was in favour of lockdowns and attacked those that protested against them while ‘Covid’ supported Black Lives Matter and kept well away from its protests. The whole thing was a joke and as lockdown protestors were arrested, often brutally, by reframed Face-Nappies we had the grotesque sight of police officers taking the knee to Black Lives Matter, a Cult-funded Marxist organisation that supports violent riots and wants to destroy the nuclear family and white people.

He's not white? Shucks!

Woke obsession with race was on display again when ten people were shot dead in Boulder, Colorado, in March, 2021. Cult-owned Woke TV channels like CNN said the shooter appeared to be a white man and Wokers were on Twitter condemning 'violent white men' with the usual mantras. Then the shooter's name was released as Ahmad Al Aliwi Alissa, an anti-Trump Arab-American, and the sigh of disappointment could be heard five miles away. Never mind that ten people were dead and what that meant for their families. Race baiting was all that mattered to these sick Cult-serving people like Barack Obama who exploited the deaths to further divide America on racial grounds which is his job for the Cult. This is the man that 'racist' white Americans made the first black president of the United States and then gave him a second term. Not-very-bright Obama has become filthy rich on the back of that and today appears to have a big influence on the Biden administration. Even so he's still a downtrodden black man and a victim of white supremacy. This disingenuous fraud reveals the contempt he has for black people when he puts on a Deep South Alabama accent whenever he talks to them, no, *at* them.

Another BLM red flag was how the now fully-Woke (fully-Cult) and fully-virtue-signalled professional soccer authorities had their teams taking the knee before every match in support of Marxist Black Lives Matter. Soccer authorities and clubs displayed 'Black Lives Matter' on the players' shirts and flashed the name on electronic billboards around the pitch. Any fans that condemned what is a Freemasonic taking-the-knee ritual were widely condemned as you would expect from the Woke virtue-signallers of professional sport and the now fully-Woke media. We have reverse racism in which you are banned from criticising any race or culture except for white people for whom anything goes – say what you like, no problem. What has this got to do with racial harmony and equality? We've had black supremacists from Black Lives Matter telling white people to fall to their knees in the street and apologise for their white supremacy. Black supremacists acting like white supremacist slave owners of the past couldn't breach their self-obsessed, race-obsessed sense of self-purity. Joe Biden appointed a race-obsessed black supremacist Kristen Clarke to head the Justice Department Civil Rights Division. Clarke claimed that blacks are endowed with 'greater mental, physical and spiritual abilities' than whites. If anyone reversed that statement they would be vilified. Clarke is on-message so no problem. She's never seen a black-white situation in which the black figure is anything but a virtuous victim and she heads the Civil Rights Division which should treat everyone the same or it isn't civil rights. Another perception of the Renegade Mind: If something or someone is part of the Cult agenda they will be supported by Woke governments and media no matter what. If they're not, they will be condemned and censored. It really is that simple and so racist Clarke prospers despite (make that because of) her racism.

The end of culture

Biden's administration is full of such racial, cultural and economic bias as the Cult requires the human family to be divided into warring factions. We are now seeing racially-segregated graduations and everything, but everything, is defined through the lens of perceived 'racism. We have 'racist' mathematics, 'racist' food and even 'racist' *plants*. World famous Kew Gardens in London said it was changing labels on plants and flowers to tell its pre-'Covid' more than two million visitors a year how racist they are. Kew director Richard Deverell said this was part of an effort to 'move quickly to decolonise collections' after they were approached by one Ajay Chhabra 'an actor with an insight into how sugar cane was linked to slavery'. They are *plants* you idiots. 'Decolonisation' in the Woke manual really means colonisation of society with its mentality and by extension colonisation by the Cult. We are witnessing a new Chinese-style 'Cultural Revolution' so essential to the success of all

Marxist takeovers. Our cultural past and traditions have to be swept away to allow a new culture to be built-back-better. Woke targeting of long-standing Western cultural pillars including historical monuments and cancelling of historical figures is what happened in the Mao revolution in China which 'purged remnants of capitalist and traditional elements from Chinese society' and installed Maoism as the dominant ideology'. For China see the Western world today and for 'dominant ideology' see Woke. Better still see Marxism or Maoism. The 'Covid' hoax has specifically sought to destroy the arts and all elements of Western culture from people meeting in a pub or restaurant to closing theatres, music venues, sports stadiums, places of worship and even banning *singing*. Destruction of Western society is also why criticism of any religion is banned except for Christianity which again is the dominant religion as white is the numerically-dominant race. Christianity may be fading rapidly, but its history and traditions are weaved through the fabric of Western society. Delete the pillars and other structures will follow until the whole thing collapses. I am not a Christian defending that religion when I say that. I have no religion. It's just a fact. To this end Christianity has itself been turned Woke to usher its own downfall and its ranks are awash with 'change agents' – knowing and unknowing – at every level including Pope Francis (*definitely* knowing) and the clueless Archbishop of Canterbury Justin Welby (possibly not, but who can be sure?). Woke seeks to coordinate attacks on Western culture, traditions, and ways of life through 'intersectionality' defined as 'the complex, cumulative way in which the effects of multiple forms of discrimination (such as racism, sexism, and classism) combine, overlap, or intersect especially in the experiences of marginalised individuals or groups'. Wade through the Orwellian Woke-speak and this means coordinating disparate groups in a common cause to overthrow freedom and liberal values.

The entire structure of public institutions has been infested with Woke – government at all levels, political parties, police, military, schools, universities, advertising, media and trade unions. This abomination has been achieved through the Cult web by appointing

Wokers to positions of power and battering non-Wokers into line through intimidation, isolation and threats to their job. Many have been fired in the wake of the empathy-deleted, vicious hostility of 'social justice' Wokers and the desire of gutless, spineless employers to virtue-signal their Wokeness. Corporations are filled with Wokers today, most notably those in Silicon Valley. Ironically at the top they are not Woke at all. They are only exploiting the mentality their Cult masters have created and funded to censor and enslave while the Wokers cheer them on until it's their turn. Thus the Woke 'liberal left' is an inversion of the traditional liberal left. Campaigning for justice on the grounds of power and wealth distribution has been replaced by campaigning for identity politics. The genuine traditional left would never have taken money from today's billionaire abusers of fairness and justice and nor would the billionaires have wanted to fund that genuine left. It would not have been in their interests to do so. The division of opinion in those days was between the haves and have nots. This all changed with Cult manipulated and funded identity politics. The division of opinion today is between Wokers and non-Wokers and not income brackets. Cult corporations and their billionaires may have taken wealth disparity to cataclysmic levels of injustice, but as long as they speak the language of Woke, hand out the dosh to the Woke network and censor the enemy they are 'one of us'. Billionaires who don't give a damn about injustice are laughing at them till their bellies hurt. Wokers are not even close to self-aware enough to see that. The transformed 'left' dynamic means that Wokers who drone on about 'social justice' are funded by billionaires that have destroyed social justice the world over. It's *why* they are billionaires.

The climate con

Nothing encapsulates what I have said more comprehensively than the hoax of human-caused global warming. I have detailed in my books over the years how Cult operatives and organisations were the pump-primers from the start of the climate con. A purpose-built vehicle for this is the Club of Rome established by the Cult in 1968 with the Rockefellers and Rothschilds centrally involved all along. Their gofer frontman Maurice Strong, a Canadian oil millionaire, hosted the Earth Summit in Rio de Janeiro, Brazil, in 1992 where the global 'green movement' really expanded in earnest under the guiding hand of the Cult. The Earth Summit established Agenda 21 through the Cult-created-and-owned United Nations to use the illusion of human-caused climate change to justify the transformation of global society to save the world from climate disaster. It is a No-Problem-Reaction-Solution sold through governments, media, schools and universities as whole generations have been terrified into believing that the world was going to end in their lifetimes unless what old people had inflicted upon them was stopped by a complete restructuring of how everything is done. Chill, kids, it's all a hoax. Such restructuring is precisely what the Cult agenda demands (purely by coincidence of course). Today this has been given the codename of the Great Reset which is only an updated term for Agenda 21 and its associated Agenda 2030. The latter, too, is administered through the UN and was voted into being by the General Assembly in 2015. Both 21 and 2030 seek centralised control of all resources and food right down to the raindrops falling on your own land. These are some of the demands of Agenda 21 established in 1992. See if you recognise this society emerging today:

- End national sovereignty
- State planning and management of all land resources, ecosystems, deserts, forests, mountains, oceans and fresh water; agriculture; rural development; biotechnology; and ensuring *'equity'*
- The state to 'define the role' of business and financial resources
- Abolition of private property
- 'Restructuring' the family unit (see BLM)
- Children raised by the state
- People told what their job will be
- Major restrictions on movement
- Creation of 'human settlement zones'

- Mass resettlement as people are forced to vacate land where they live
- Dumbing down education
- Mass global depopulation in pursuit of all the above

The United Nations was created as a Trojan horse for world government. With the climate con of critical importance to promoting that outcome you would expect the UN to be involved. Oh, it's involved all right. The UN is promoting Agenda 21 and Agenda 2030 justified by 'climate change' while also driving the climate hoax through its Intergovernmental Panel on Climate Change (IPCC), one of the world's most corrupt organisations. The IPCC has been lying ferociously and constantly since the day it opened its doors with the global media hanging unquestioningly on its every mendacious word. The Green movement is entirely Woke and has long lost its original environmental focus since it was co-opted by the Cult. An obsession with 'global warming' has deleted its values and scrambled its head. I experienced a small example of what I mean on a beautiful country walk that I have enjoyed several times a week for many years. The path merged into the fields and forests and you felt at one with the natural world. Then a 'Green' organisation, the Hampshire and Isle of Wight Wildlife Trust, took over part of the land and proceeded to cut down a large number of trees, including mature ones, to install a horrible big, bright steel 'this-is-ours-stay-out' fence that destroyed the whole atmosphere of this beautiful place. No one with a feel for nature would do that. Day after day I walked to the sound of chainsaws and a magnificent mature weeping willow tree that I so admired was cut down at the base of the trunk. When I challenged a Woke young girl in a green shirt (of course) about this vandalism she replied: 'It's a weeping willow – it will grow back.' This is what people are paying for when they donate to the Hampshire and Isle of Wight Wildlife Trust and many other 'green' organisations today. It is not the environmental movement that I knew and instead has become a support-system – as with Extinction Rebellion – for a very dark agenda.

Private jets for climate justice

The Cult-owned, Gates-funded, World Economic Forum and its founder Klaus Schwab were behind the emergence of Greta Thunberg to harness the young behind the climate agenda and she was invited to speak to the world at … the UN. Schwab published a book, *Covid-19: The Great Reset* in 2020 in which he used the 'Covid' hoax and the climate hoax to lay out a new society straight out of Agenda 21 and Agenda 2030. Bill Gates followed in early 2021 when he took time out from destroying the world to produce a book in his name about the way to save it. Gates flies across the world in private jets and admitted that 'I probably have one of the highest greenhouse gas footprints of anyone on the planet … my personal flying alone is gigantic.' He has also bid for the planet's biggest private jet operator. Other climate change saviours who fly in private jets include John Kerry, the US Special Presidential Envoy for Climate, and actor Leonardo DiCaprio, a 'UN Messenger of Peace with special focus on climate change'. These people are so full of bullshit they could corner the market in manure. We mustn't be sceptical, though, because the Gates book, *How to Avoid a Climate Disaster: The Solutions We Have and the Breakthroughs We Need*, is a genuine attempt to protect the world and not an obvious pile of excrement attributed to a mega-psychopath aimed at selling his masters' plans for humanity. The Gates book and the other shite-pile by Klaus Schwab could have been written by the same person and may well have been. Both use 'climate change' and 'Covid' as the excuses for their new society and by coincidence the Cult's World Economic Forum and Bill and Melinda Gates Foundation promote the climate hoax and hosted Event 201 which pre-empted with a 'simulation' the very 'coronavirus' hoax that would be simulated for real on humanity within weeks. The British 'royal' family is promoting the 'Reset' as you would expect through Prince 'climate change caused the war in Syria' Charles and his hapless son Prince William who said that we must 'reset our relationship with nature and our trajectory as a species' to avoid a climate disaster. Amazing how many promotors of the 'Covid' and 'climate change' control systems are connected to Gates and the World Economic Forum. A 'study' in early 2021 claimed that carbon dioxide emissions must fall by the equivalent of a global lockdown roughly every two years for the next decade to save the planet. The 'study' appeared in the same period that the Schwab mob claimed in a video that lockdowns destroying the lives of billions are good because they make the earth 'quieter' with less 'ambient noise'. They took down the video amid a public backlash for such arrogant, empathy-deleted stupidity You see, however, where they are going with this. Corinne Le Quéré, a professor at the Tyndall Centre for Climate Change Research, University of East Anglia, was lead author of the climate lockdown study, and she writes for … the World Economic Forum. Gates calls in 'his' book for changing 'every aspect of the economy' (long-time Cult agenda) and for humans to eat synthetic 'meat' (predicted in my books) while cows and other farm animals are eliminated. Australian TV host and commentator Alan Jones described what carbon emission targets would mean for farm animals in Australia alone if emissions were reduced as demanded by 35 percent by 2030 and zero by 2050:

> Well, let's take agriculture, the total emissions from agriculture are about 75 million tonnes of carbon dioxide, equivalent. Now reduce that by 35 percent and you have to come down to 50 million tonnes, I've done the maths. So if you take for example 1.5 million cows, you're going to have to reduce the herd by 525,000 [by] 2030, nine years, that's 58,000 cows a year. The beef herd's 30 million, reduce that by 35 percent, that's 10.5 million, which means 1.2 million cattle have to go every year between now and 2030. This is insanity!
>
> There are 75 million sheep. Reduce that by 35 percent, that's 26 million sheep, that's almost 3 million a year. So under the Paris Agreement over 30 million beasts. dairy cows, cattle, pigs and sheep would go. More than 8,000 every minute of every hour for the next decade, do these people know what they're talking about?

Clearly they don't at the level of campaigners, politicians and administrators. The Cult *does* know; that's the outcome it wants. We are faced with not just a war on humanity. Animals and the natural world are being targeted and I have been saying since the 'Covid' hoax began that the plan eventually was to claim that the 'deadly virus' is able to jump from animals, including farm animals and domestic pets, to humans. Just before this book went into production came this story: ‘Russia registers world’s first Covid-19 vaccine for cats & dogs as makers of Sputnik V warn pets & farm animals could spread virus’. The report said ‘top scientists warned that the deadly pathogen could soon begin spreading through homes and farms’ and ‘the next stage is the infection of farm and domestic animals’. Know the outcome and you’ll see the journey. Think what that would mean for animals and keep your eye on a term called zoonosis or zoonotic diseases which transmit between animals and humans. The Cult wants to break the connection between animals and people as it does between people and people. Farm animals fit with the Cult agenda to transform food from natural to synthetic.

The gas of life is killing us

There can be few greater examples of Cult inversion than the condemnation of carbon dioxide as a dangerous pollutant when it is the gas of life. Without it the natural world would be dead and so we would all be dead. We breathe in oxygen and breathe out carbon dioxide while plants produce oxygen and absorb carbon dioxide. It is a perfect symbiotic relationship that the Cult wants to dismantle for reasons I will come to in the final two chapters. Gates, Schwab, other Cult operatives and mindless repeaters, want the world to be ‘carbon neutral’ by at least 2050 and the earlier the better. ‘Zero carbon’ is the cry echoed by lunatics calling for ‘Zero Covid’ when we already have it. These carbon emission targets will deindustrialise the world in accordance with Cult plans – the post-industrial, post-democratic society – and with so-called renewables like solar and wind not coming even close to meeting human energy needs blackouts and cold are inevitable. Texans got the picture in the winter of 2021 when a snow storm stopped wind turbines and solar panels from working and the lights went down along with water which relies on electricity for its supply system. Gates wants everything to be powered by electricity to ensure that his masters have the kill switch to stop all human activity, movement, cooking, water and warmth any time they like. The climate lie is so stupendously inverted that it claims we must urgently reduce carbon dioxide when we *don't have enough*.

Co2 in the atmosphere is a little above 400 parts per million when the optimum for plant growth is 2,000 ppm and when it falls anywhere near 150 ppm the natural world starts to die and so do we. It fell to as low as 280 ppm in an 1880 measurement in Hawaii and rose to 413 ppm in 2019 with industrialisation which is why the planet has become *greener* in the industrial period. How insane then that psychopathic madman Gates is not satisfied only with blocking the rise of Co2. He's funding technology to suck it out of the atmosphere. The reason why will become clear. The industrial era is not destroying the world through Co2 and has instead turned around a potentially disastrous ongoing fall in Co2. Greenpeace co-founder and scientist Patrick Moore walked away from Greenpeace in 1986 and has exposed the green movement for fear-mongering and lies. He said that 500 million years ago there was *17 times* more Co2 in the atmosphere than we have today and levels have been falling for hundreds of millions of years. In the last 150 million years Co2 levels in Earth's atmosphere had reduced by *90 percent*. Moore said that by the time humanity began to unlock carbon dioxide from fossil fuels we were at '38 seconds to midnight' and in that sense: 'Humans are [the Earth's] salvation.' Moore made the point that only half the Co2 emitted by fossil fuels stays in the atmosphere and we should remember that all pollution pouring from chimneys that we are told is carbon dioxide is in fact nothing of the kind. It's pollution. Carbon dioxide is an invisible gas.

William Happer, Professor of Physics at Princeton University and long-time government adviser on climate, has emphasised the Co2 deficiency for maximum growth and food production. Greenhouse growers don't add carbon dioxide for a bit of fun. He said that most of the warming in the last 100 years, after the earth emerged from the super-cold period of the 'Little Ice Age' into a natural warming cycle, was over by 1940. Happer said that a peak year for warming in 1988 can be explained by a 'monster El Nino' which is a natural and cyclical warming of the Pacific that has nothing to do with 'climate change'. He said the effect of Co2 could be compared to painting a wall with red paint in that once two or three coats have been applied it didn't matter how much more you slapped on because the wall will not get much redder. Almost all the effect of the rise in Co2 has already happened, he said, and the volume in the atmosphere would now have to *double* to increase temperature by a single degree. Climate hoaxers know this and they have invented the most ridiculously complicated series of 'feedback' loops to try to overcome this rather devastating fact. You hear puppet Greta going on cluelessly about feedback loops and this is why.

The Sun affects temperature? No you *climate denier*

Some other nonsense to contemplate: Climate graphs show that rises in temperature do not follow rises in Co2 – *it's the other way round* with a lag between the two of some 800 years. If we go back 800 years from present time we hit the Medieval Warm Period when temperatures were higher than now without any industrialisation and this was followed by the Little Ice Age when temperatures plummeted. The world was still emerging from these centuries of serious cold when many climate records began which makes the ever-repeated line of the 'hottest year since records began' meaningless when you are not comparing like with like. The coldest period of the Little Ice Age corresponded with the lowest period of sunspot activity when the Sun was at its least active. Proper scientists will not be at all surprised by this when it confirms the obvious fact that earth temperature is affected by the scale of Sun activity and the energetic power that it subsequently emits; but when is the last time you heard a climate hoaxer talking about the Sun as a source of earth temperature?? Everything has to be focussed on Co2 which makes up just 0.117 percent of so-called greenhouse gases and only a fraction of even that is generated by human activity. The rest is natural. More than *90 percent* of those greenhouse gases are water vapour and clouds (Fig 9). Ban moisture I say. Have you noticed that the climate hoaxers no longer use the polar bear as their promotion image? That's because far from becoming extinct polar bear communities are stable or thriving. Joe Bastardi, American meteorologist, weather forecaster and outspoken critic of the climate lie, documents in his book *The Climate Chronicles* how weather patterns and events claimed to be evidence of climate change have been happening since long before industrialisation: 'What happened before naturally is happening again, as is to be expected given the cyclical nature of the climate due to the design of the planet.' If you read the detailed background to the climate hoax in my other books you will shake your head and wonder how anyone could believe the crap which has spawned a multi-trillion dollar industry based on absolute garbage (see HIV causes AIDs and Sars-Cov-2 causes 'Covid-19'). Climate and 'Covid' have much in common given they have the same source. They both have the contradictory *everything* factor in which everything is explained by reference to them. It's hot – 'it's climate change'. It's cold – 'it's climate change'. I got a sniffle – 'it's Covid'. I haven't got a sniffle – 'it's Covid'. Not having a sniffle has to be a symptom of 'Covid'. Everything is and not having a sniffle is especially dangerous if you are a slow walker. For sheer audacity I offer you a Cambridge University 'study' that actually linked 'Covid' to 'climate change'. It had to happen eventually. They concluded that climate change played a role in 'Covid-19' spreading from animals to humans because … wait for it … I kid you not … *the two groups were forced closer together as populations grow*. Er, that's it. The whole foundation on which this depended was that 'Bats are the likely zoonotic origin of SARS-CoV-1 and SARS-CoV-2'. Well, they are not. They are nothing to do with it. Apart from bats not being the origin and therefore 'climate change' effects on bats being irrelevant I am in awe of their academic insight. Where would we be without them? Not where we are that's for sure.

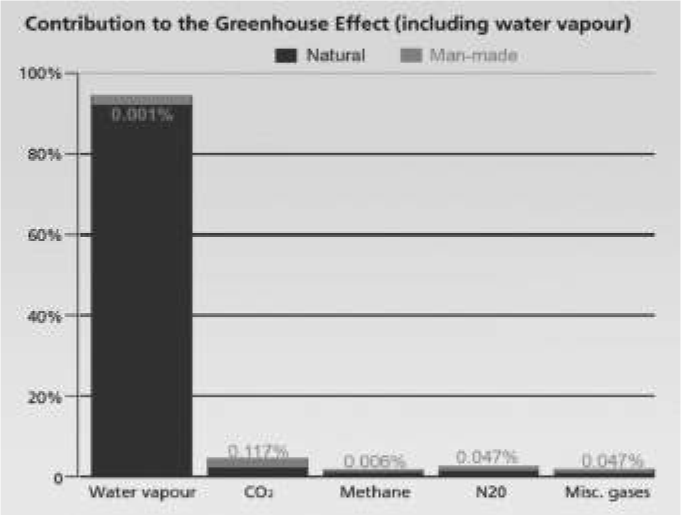


Figure 9: The idea that the gas of life is disastrously changing the climate is an insult to brain cell activity.

One other point about the weather is that climate modification is now well advanced and not every major weather event is natural – or earthquake come to that. I cover this subject at some length in other books. China is openly planning a rapid expansion of its weather modification programme which includes changing the climate in an area more than one and a half times the size of India. China used weather manipulation to ensure clear skies during the 2008 Olympics in Beijing. I have quoted from US military documents detailing how to employ weather manipulation as a weapon of war and they did that in the 1960s and 70s during the conflict in Vietnam with Operation Popeye manipulating monsoon rains for military purposes. Why would there be international treaties on weather modification if it wasn't possible? Of course it is. Weather is energetic information and it can be changed.

How was the climate hoax pulled off? See 'Covid'

If you can get billions to believe in a 'virus' that doesn't exist you can get them to believe in human-caused climate change that doesn't exist. Both are being used by the Cult to transform global society in the way it has long planned. Both hoaxes have been achieved in pretty much the same way. First you declare a lie is a fact. There's a ‘virus’ you call SARS-Cov-2 or humans are warming the planet with their behaviour. Next this becomes, via Cult networks, the foundation of government, academic and science policy and belief. Those who parrot the mantra are given big grants to produce research that confirms the narrative is true and ever more ‘symptoms’ are added to make the ‘virus’/’climate change’ sound even more scary. Scientists and researchers who challenge the narrative have their grants withdrawn and their careers destroyed. The media promote the lie as the unquestionable truth and censor those with an alternative view or evidence. A great percentage of the population believe what they are told as the lie becomes an everybody-knows-that and the believing-masses turn on those with a mind of their own. The technique has been used endlessly throughout human history. Wokers are the biggest promotors of the climate lie *and* ‘Covid’ fascism because their minds are owned by the Cult; their sense of self-righteous self-purity knows no bounds; and they exist in a bubble of reality in which facts are irrelevant and only get in the way of looking without seeing.

Running through all of this like veins in a blue cheese is control of information, which means control of perception, which means control of behaviour, which collectively means control of human society. The Cult owns the global media and Silicon Valley fascists for the simple reason that it *has* to. Without control of information it can’t control perception and through that human society. Examine every facet of the Cult agenda and you will see that anything supporting its introduction is never censored while anything pushing back is always censored. I say again: Psychopaths that know why they are doing this must go before Nuremberg trials and those that follow their orders must trot along behind them into the same dock. ‘I was just following orders’ didn’t work the first time and it must not work now. Nuremberg trials must be held all over the world before public juries for politicians, government officials, police, compliant doctors, scientists and virologists, and all Cult operatives such as Gates, Tedros, Fauci, Vallance, Whitty, Ferguson, Zuckerberg, Wojcicki, Brin, Page, Dorsey, the whole damn lot of them – including, no *especially*, the psychopath psychologists. Without them and the brainless, gutless excuses for journalists that have repeated their lies, none of this could be happening. Nobody can be allowed to escape justice for the psychological and economic Armageddon they are all responsible for visiting upon the human race.

As for the compliant, unquestioning, swathes of humanity, and the self-obsessed, all-knowing ignorance of the Wokers … don’t start me. God help their kids. God help their grandkids. God *help them*.

CHAPTER NINE

We must have it? So what is it?

Well I won't back down. No, I won't back down. You can stand me up at the Gates of Hell. But I won't back down
Tom Petty

I will now focus on the genetically-manipulating 'Covid vaccines' which do not meet this official definition of a vaccine by the US Centers for Disease Control (CDC): 'A product that stimulates a person's immune system to produce immunity to a specific disease, protecting the person from that disease.' On that basis 'Covid vaccines' are not a vaccine in that the makers don't even claim they stop infection or transmission.

They are instead part of a multi-levelled conspiracy to change the nature of the human body and what it means to be 'human' and to depopulate an enormous swathe of humanity. What I shall call Human 1.0 is on the cusp of becoming Human 2.0 and for very sinister reasons. Before I get to the 'Covid vaccine' in detail here's some background to vaccines in general. Government regulators do not test vaccines – the makers do – and the makers control which data is revealed and which isn't. Children in America are given 50 vaccine doses by age six and 69 by age 19 and the effect of the whole combined schedule has never been tested. Autoimmune diseases when the immune system attacks its own body have soared in the mass vaccine era and so has disease in general in children and the young. Why wouldn't this be the case when vaccines target the *immune system?* The US government gave Big Pharma drug companies immunity from prosecution for vaccine death and injury in the 1986 National Childhood Vaccine Injury Act (NCVIA) and since then the government (taxpayer) has been funding compensation for the consequences of Big Pharma vaccines. The criminal and satanic drug giants can't lose and the vaccine schedule has increased dramatically since 1986 for this reason. There is no incentive to make vaccines safe and a big incentive to make money by introducing ever more. Even against a ridiculously high bar to prove vaccine liability, and with the government controlling the hearing in which it is being challenged for compensation, the vaccine court has so far paid out more than $4 billion. These are the vaccines we are told are safe and psychopaths like Zuckerberg censor posts saying otherwise. The immunity law was even justified by a ruling that vaccines by their nature were 'unavoidably unsafe'.

Check out the ingredients of vaccines and you will be shocked if you are new to this. *They put that in children's bodies?? What??* Try aluminium, a brain toxin connected to dementia, aborted foetal tissue and formaldehyde which is used to embalm corpses. World-renowned aluminium expert Christopher Exley had his research into the health effect of aluminium in vaccines shut down by Keele University in the UK when it began taking funding from the Bill and Melinda Gates Foundation. Research when diseases 'eradicated' by vaccines began to decline and you will find the fall began long *before* the vaccine was introduced. Sometimes the fall even plateaued after the vaccine. Diseases like scarlet fever for which there was no vaccine declined in the same way because of environmental and other factors. A perfect case in point is the polio vaccine. Polio began when lead arsenate was first sprayed as an insecticide and residues remained in food products. Spraying started in 1892 and the first US polio epidemic came in Vermont in 1894. The simple answer was to stop spraying, but Rockefeller-created Big Pharma had a better idea. Polio was decreed to be caused by the *poliovirus* which 'spreads from person to person and can infect a person's spinal cord'. Lead arsenate was replaced by the lethal DDT which had the same effect of causing paralysis by damaging the brain and central nervous system. Polio plummeted when DDT was reduced and then banned, but the vaccine is still given the credit for something it didn't do. Today by far the biggest cause of polio is the vaccines promoted by Bill Gates. Vaccine justice campaigner Robert Kennedy Jr, son of assassinated (by the Cult) US Attorney General Robert Kennedy, wrote:

> In 2017, the World Health Organization (WHO) reluctantly admitted that the global explosion in polio is predominantly vaccine strain. The most frightening epidemics in Congo, Afghanistan, and the Philippines, are all linked to vaccines. In fact, by 2018, 70% of global polio cases were vaccine strain.

Vaccines make fortunes for Cult-owned Gates and Big Pharma while undermining the health and immune systems of the population. We had a glimpse of the mentality behind the Big Pharma cartel with a report on WION (World is One News), an international English language TV station based in India, which exposed the extraordinary behaviour of US drug company Pfizer over its 'Covid vaccine'. The WION report told how Pfizer had made fantastic demands of Argentina, Brazil and other countries in return for its 'vaccine'. These included immunity from prosecution, even for Pfizer negligence, government insurance to protect Pfizer from law suits and handing over as collateral sovereign assets of the country to include Argentina's bank reserves, military bases and embassy buildings. Pfizer demanded the same of Brazil in the form of waiving sovereignty of its assets abroad; exempting Pfizer from Brazilian laws; and giving Pfizer immunity from all civil liability. This is a 'vaccine' developed with government funding. Big Pharma is evil incarnate as a creation of the Cult and all must be handed tickets to Nuremberg.

Phantom 'vaccine' for a phantom 'disease'

I'll expose the 'Covid vaccine' fraud and then go on to the wider background of why the Cult has set out to 'vaccinate' every man, woman and child on the planet for an alleged 'new disease' with a survival rate of 99.77 percent (or more) even by the grotesquelymanipulated figures of the World Health Organization and Johns Hopkins University. The ‘infection’ to ‘death’ ratio is 0.23 to 0.15 percent according to Stanford epidemiologist Dr John Ioannidis and while estimates vary the danger remains tiny. I say that if the truth be told the fake infection to fake death ratio is zero. Never mind all the evidence I have presented here and in *The Answer* that there is no ‘virus’ let us just focus for a moment on that death-rate figure of say 0.23 percent. The figure includes all those worldwide who have tested positive with a test not testing for the ‘virus’ and then died within 28 days or even longer of any other cause – *any other cause*. Now subtract all those illusory ‘Covid’ deaths on the global data sheets from the 0.23 percent. What do you think you would be left with? *Zero*. A vaccination has never been successfully developed for a so-called coronavirus. They have all failed at the animal testing stage when they caused hypersensitivity to what they were claiming to protect against and made the impact of a disease far worse. Cult-owned vaccine corporations got around that problem this time by bypassing animal trials, going straight to humans and making the length of the ‘trials’ before the public rollout as short as they could get away with. Normally it takes five to ten years or more to develop vaccines that still cause demonstrable harm to many people and that’s without including the long-term effects that are never officially connected to the vaccination. ‘Covid’ non-vaccines have been officially produced and approved in a matter of months from a standing start and part of the reason is that (a) they were developed before the ‘Covid’ hoax began and (b) they are based on computer programs and not natural sources. Official non-trials were so short that government agencies gave *emergency*, not full, approval. ‘Trials’ were not even completed and full approval cannot be secured until they are. Public ‘Covid vaccination’ is actually a *continuation of the trial*. Drug company ‘trials’ are not scheduled to end until 2023 by which time a lot of people are going to be dead. Data on which government agencies gave this emergency approval was supplied by the Big Pharma corporations themselves in the form of Pfizer/BioNTech, AstraZeneca, Moderna, Johnson & Johnson, and others, and this is the case with all vaccines. By its very nature *emergency* approval means drug companies do not have to prove that the 'vaccine' is 'safe and effective'. How could they with trials way short of complete? Government regulators only have to *believe* that they *could* be safe and effective. It is criminal manipulation to get products in circulation with no testing worth the name. Agencies giving that approval are infested with Big Pharma-connected place-people and they act in the interests of Big Pharma (the Cult) and not the public about whom they do not give a damn.

More human lab rats

'Covid vaccines' produced in record time by Pfizer/BioNTech and Moderna employ a technique *never approved before for use on humans*. They are known as mRNA 'vaccines' and inject a synthetic version of 'viral' mRNA or 'messenger RNA'. The key is in the term 'messenger'. The body works, or doesn't, on the basis of information messaging. Communications are constantly passing between and within the genetic system and the brain. Change those messages and you change the state of the body and even its very nature and you can change psychology and behaviour by the way the brain processes information. I think you are going to see significant changes in personality and perception of many people who have had the 'Covid vaccine' synthetic potions. Insider Aldous Huxley predicted the following in 1961 and mRNA 'vaccines' can be included in the term 'pharmacological methods':

> There will be, in the next generation or so, a pharmacological method of making people love their servitude, and producing dictatorship without tears, so to speak, producing a kind of painless concentration camp for entire societies, so that people will in fact have their own liberties taken away from them, but rather enjoy it, because they will be distracted from any desire to rebel by propaganda or brainwashing, or brainwashing enhanced by pharmacological methods. And this seems to be the final revolution.

Apologists claim that mRNA synthetic 'vaccines' don't change the DNA genetic blueprint because RNA does not affect DNA only the other way round. This is so disingenuous. A process called 'reverse transcription' can convert RNA into DNA and be integrated into DNA in the cell nucleus. This was highlighted in December, 2020, by scientists at Harvard and Massachusetts Institute of Technology (MIT). Geneticists report that more than 40 percent of mammalian genomes results from reverse transcription. On the most basic level if messaging changes then that sequence must lead to changes in DNA which is receiving and transmitting those communications. How can introducing synthetic material into cells not change the cells where DNA is located? The process is known as transfection which is defined as 'a technique to insert foreign nucleic acid (DNA or RNA) into a cell, typically with the intention of altering the properties of the cell'. Researchers at the Sloan Kettering Institute in New York found that changes in messenger RNA can deactivate tumour-suppressing proteins and thereby promote cancer. This is what happens when you mess with messaging. 'Covid vaccine' maker Moderna was founded in 2010 by Canadian stem cell biologist Derrick J. Rossi after his breakthrough discovery in the field of transforming and reprogramming stem cells. These are neutral cells that can be programmed to become any cell including sperm cells. Moderna was therefore founded on the principle of genetic manipulation and has never produced any vaccine or drug before its genetically-manipulating synthetic 'Covid' shite. Look at the name – Mode-RNA or Modify-RNA. Another important point is that the US Supreme Court has ruled that genetically-modified DNA, or complementary DNA (cDNA) synthesized in the laboratory from messenger RNA, can be patented and owned. These psychopaths are doing this to the human body.

Cells replicate synthetic mRNA in the 'Covid vaccines' and in theory the body is tricked into making antigens which trigger antibodies to target the 'virus spike proteins' which as Dr Tom Cowan said have *never been seen*. Cut the crap and these 'vaccines' deliver *self-replicating* synthetic material to the cells with the effect of changing human DNA. The more of them you have the more that process is compounded while synthetic material is all the time self-replicating. 'Vaccine'-maker Moderna describes mRNA as 'like software for the cell' and so they are messing with the body's software. What happens when you change the software in a computer? Everything changes. For this reason the Cult is preparing a production line of mRNA 'Covid vaccines' and a long list of excuses to use them as with all the 'variants' of a 'virus' never shown to exist. The plan is further to transfer the mRNA technique to other vaccines mostly given to children and young people. The cumulative consequences will be a transformation of human DNA through a constant infusion of synthetic genetic material which will kill many and change the rest. Now consider that governments that have given emergency approval for a vaccine that's not a vaccine; never been approved for humans before; had no testing worth the name; and the makers have been given immunity from prosecution for any deaths or adverse effects suffered by the public. The UK government awarded *permanent legal indemnity* to itself and its employees for harm done when a patient is being treated for 'Covid-19' or 'suspected Covid-19'. That is quite a thought when these are possible 'side-effects' from the 'vaccine' (they are not 'side', they are effects) listed by the US Food and Drug Administration:

Guillain-Barre syndrome; acute disseminated encephalomyelitis; transverse myelitis; encephalitis; myelitis; encephalomyelitis; meningoencephalitis; meningitis; encephalopathy; convulsions; seizures; stroke; narcolepsy; cataplexy; anaphylaxis; acute myocardial infarction (heart attack); myocarditis; pericarditis; autoimmune disease; death; implications for pregnancy, and birth outcomes; other acute demyelinating diseases; non anaphylactic allergy reactions; thrombocytopenia ; disseminated intravascular coagulation; venous thromboembolism; arthritis; arthralgia; joint pain; Kawasaki disease; multisystem inflammatory syndrome in children; vaccine enhanced disease. The latter is the way the 'vaccine' has the potential to make diseases far worse than they would otherwise be.

UK doctor and freedom campaigner Vernon Coleman described the conditions in this list as 'all unpleasant, most of them very serious, and you can't get more serious than death'. The thought that anyone at all has had the 'vaccine' in these circumstances is testament to the potential that humanity has for clueless, unquestioning, stupidity and for many that programmed stupidity has already been terminal.

An insider speaks

Dr Michael Yeadon is a former Vice President, head of research and Chief Scientific Adviser at vaccine giant Pfizer. Yeadon worked on the inside of Big Pharma, but that did not stop him becoming a vocal critic of 'Covid vaccines' and their potential for multiple harms, including infertility in women. By the spring of 2021 he went much further and even used the no, no, term 'conspiracy'. When you begin to see what is going on it is impossible not to do so. Yeadon spoke out in an interview with freedom campaigner James Delingpole and I mentioned earlier how he said that no one had samples of 'the virus'. He explained that the mRNA technique originated in the anti-cancer field and ways to turn on and off certain genes which could be advantageous if you wanted to stop cancer growing out of control. 'That's the origin of them. They are a very unusual application, really.' Yeadon said that treating a cancer patient with an aggressive procedure might be understandable if the alternative was dying, but it was quite another thing to use the same technique as a public health measure. Most people involved wouldn't catch the infectious agent you were vaccinating against and if they did they probably wouldn't die:

> If you are really using it as a public health measure you really want to as close as you can get to zero sides-effects … I find it odd that they chose techniques that were really cutting their teeth in the field of oncology and I'm worried that in using gene-based vaccines that have to be injected in the body and spread around the body, get taken up into some cells, and the regulators haven't quite told us which cells they get taken up into … you are going to be generating a wide range of responses … with multiple steps each of which could go well or badly.

I doubt the Cult intends it to go well. Yeadon said that you can put any gene you like into the body through the 'vaccine'. 'You can certainly give them a gene that would do them some harm if you wanted.' I was intrigued when he said that when used in the cancer field the technique could turn genes on and off. I explore this process in *The Answer* and with different genes having different functions you could create mayhem – physically and psychologically – if you turned the wrong ones on and the right ones off. I read reports of an experiment by researchers at the University of Washington's school of computer science and engineering in which they encoded DNA to infect computers. The body is itself a biological computer and if human DNA can inflict damage on a computer why can't the computer via synthetic material mess with the human body? It can. The Washington research team said it was possible to insert malicious malware into 'physical DNA strands' and corrupt the computer system of a gene sequencing machine as it 'reads gene letters and stores them as binary digits 0 and 1'. They concluded that hackers could one day use blood or spit samples to access computer systems and obtain sensitive data from police forensics labs or infect genome files. It is at this level of digital interaction that synthetic 'vaccines' need to be seen to get the full picture and that will become very clear later on. Michael Yeadon said it made no sense to give the 'vaccine' to younger people who were in no danger from the 'virus'. What was the benefit? It was all downside with potential effects:

> The fact that my government in what I thought was a civilised, rational country, is raining [the 'vaccine'] on people in their 30s and 40s, even my children in their 20s, they're getting letters and phone calls, I know this is not right and any of you doctors who are vaccinating you know it's not right, too. They are not at risk. They are not at risk from the disease, so you are now hoping that the side-effects are so rare that you get away with it. You don't give new technology … that you don't understand to 100 percent of the population.

Blood clot problems with the AstraZeneca 'vaccine' have been affecting younger people to emphasise the downside risks with no benefit. AstraZeneca's version, produced with Oxford University, does not use mRNA, but still gets its toxic cocktail inside cells where it targets DNA. The Johnson & Johnson 'vaccine' which uses a similar technique has also produced blood clot effects to such an extent that the United States paused its use at one point. They are all 'gene therapy' (cell modification) procedures and not 'vaccines'. The truth is that once the content of these injections enter cells we have no idea what the effect will be. People can speculate and some can give very educated opinions and that's good. In the end, though, only the makers know what their potions are designed to do and even they won't know every last consequence. Michael Yeadon was scathing about doctors doing what they knew to be wrong. 'Everyone's mute', he said. Doctors in the NHS must know this was not right, coming into work and injecting people. 'I don't know how they sleep at night. I know I couldn't do it. I know that if I were in that position I'd have to quit.' He said he knew enough about toxicology to know this was not a good risk-benefit. Yeadon had spoken to seven or eight university professors and all except two would not speak out publicly. Their universities had a policy that no one said anything that countered the government and its medical advisors. They were afraid of losing their government grants. This is how intimidation has been used to silence the truth at every level of the system. I say silence, but these people could still speak out if they made that choice. Yeadon called them 'moral cowards' – 'This is about your children and grandchildren's lives and you have just buggered off and left it.'

'Variant' nonsense

Some of his most powerful comments related to the alleged 'variants' being used to instil more fear, justify more lockdowns, and introduce more 'vaccines'. He said government claims about 'variants' were nonsense. He had checked the alleged variant 'codes' and they were 99.7 percent identical to the 'original'. This was the human identity difference equivalent to putting a baseball cap on and off or wearing it the other way round. A 0.3 percent difference would make it impossible for that 'variant' to escape immunity from the 'original'. This made no sense of having new 'vaccines' for ‘variants’. He said there would have to be at least a *30 percent* difference for that to be justified and even then he believed the immune system would still recognise what it was. Gates-funded ‘variant modeller’ and ‘vaccine’-pusher John Edmunds might care to comment. Yeadon said drug companies were making new versions of the ‘vaccine’ as a ‘top up’ for ‘variants’. Worse than that, he said, the ‘regulators’ around the world like the MHRA in the UK had got together and agreed that because ‘vaccines’ for ‘variants’ were so similar to the first ‘vaccines’ *they did not have to do safety studies*. How transparently sinister that is. This is when Yeadon said: ‘There is a conspiracy here.’ There was no need for another vaccine for ‘variants’ and yet we were told that there was and the country had shut its borders because of them. ‘They are going into hundreds of millions of arms without passing ‘go’ or any regulator. Why did they do that? Why did they pick this method of making the vaccine?’

The reason had to be something bigger than that it seemed and ‘it’s not protection against the virus’. It’s was a far bigger project that meant politicians and advisers were willing to do things and not do things that knowingly resulted in avoidable deaths – ‘that’s already happened when you think about lockdown and deprivation of health care for a year.’ He spoke of people prepared to do something that results in the avoidable death of their fellow human beings and it not bother them. This is the penny-drop I have been working to get across for more than 30 years – the level of pure evil we are dealing with. Yeadon said his friends and associates could not believe there could be that much evil, but he reminded them of Stalin, Pol Pot and Hitler and of what Stalin had said: ‘One death is a tragedy. A million? A statistic.’ He could not think of a benign explanation for why you need top-up vaccines ‘which I’m sure you don’t’ and for the regulators ‘to just get out of the way and wave them through’. Why would the regulators do that when they were still wrestling with the dangers of the ‘parent’ vaccine? He was clearly shocked by what he had seen since the ‘Covid’ hoax began and now he was thinking the previously unthinkable:

> If you wanted to depopulate a significant proportion of the world and to do it in a way that doesn't involve destruction of the environment with nuclear weapons, poisoning everyone with anthrax or something like that, and you wanted plausible deniability while you had a multi-year infectious disease crisis, I actually don't think you could come up with a better plan of work than seems to be in front of me. I can't say that's what they are going to do, but I can't think of a benign explanation why they are doing it.

He said he never thought that they would get rid of 99 percent of humans, but now he wondered. 'If you wanted to that this would be a hell of a way to do it – it would be unstoppable folks.' Yeadon had concluded that those who submitted to the 'vaccine' would be allowed to have some kind of normal life (but for how long?) while screws were tightened to coerce and mandate the last few percent. 'I think they'll put the rest of them in a prison camp. I wish I was wrong, but I don't think I am.' Other points he made included: There were no coronavirus vaccines then suddenly they all come along at the same time; we have no idea of the long term affect with trials so short; coercing or forcing people to have medical procedures is against the Nuremberg Code instigated when the Nazis did just that; people should at least delay having the 'vaccine'; a quick Internet search confirms that masks don't reduce respiratory viral transmission and 'the government knows that'; they have smashed civil society and they know that, too; two dozen peer-reviewed studies show no connection between lockdown and reducing deaths; he knew from personal friends the elite were still flying around and going on holiday while the public were locked down; the elite were not having the 'vaccines'. He was also asked if 'vaccines' could be made to target difference races. He said he didn't know, but the document by the Project for the New American Century in September, 2000, said developing 'advanced forms of biological warfare that can target *specific genotypes* may transform biological warfare from the realm of terror to a politically useful tool.' Oh, they're evil all right. Of that we can be *absolutely* sure.

Another cull of old people

We have seen from the CDC definition that the mRNA 'Covid vaccine' is not a vaccine and nor are the others that *claim* to reduce 'severity of symptoms' in *some* people, but not protect from infection or transmission. What about all the lies about returning to 'normal' if people were 'vaccinated'? If they are not claimed to stop infection and transmission of the alleged 'virus', how does anything change? This was all lies to manipulate people to take the jabs and we are seeing that now with masks and distancing still required for the 'vaccinated'. How did they think that elderly people with fragile health and immune responses were going to be affected by infusing their cells with synthetic material and other toxic substances? They *knew* that in the short and long term it would be devastating and fatal as the culling of the old that began with the first lockdowns was continued with the 'vaccine'. Death rates in care homes soared immediately residents began to be 'vaccinated' – infused with synthetic material. Brave and committed whistleblower nurses put their careers at risk by exposing this truth while the rest kept their heads down and their mouths shut to put their careers before those they are supposed to care for. A long-time American Certified Nursing Assistant who gave his name as James posted a video in which he described emotionally what happened in his care home when vaccination began. He said that during 2020 very few residents were sick with 'Covid' and no one died during the entire year; but shortly after the Pfizer mRNA injections 14 people died within two weeks and many others were near death. 'They're dropping like flies', he said. Residents who walked on their own before the shot could no longer and they had lost their ability to conduct an intelligent conversation. The home's management said the sudden deaths were caused by a 'super-spreader' of 'Covid-19'. Then how come, James asked, that residents who refused to take the injections were not sick? It was a case of inject the elderly with mRNA synthetic potions and blame their illness and death that followed on the 'virus'. James described what was happening in care homes as 'the greatest crime of genocide this country has ever seen'. Remember the NHS staff nurse from earlier who used the same word ‘genocide’ for what was happening with the ‘vaccines’ and that it was an ‘act of human annihilation’. A UK care home whistleblower told a similar story to James about the effect of the ‘vaccine’ in deaths and ‘outbreaks’ of illness dubbed ‘Covid’ after getting the jab. She told how her care home management and staff had zealously imposed government regulations and no one was allowed to even question the official narrative let alone speak out against it. She said the NHS was even worse. Again we see the results of reframing. A worker at a local care home where I live said they had not had a single case of ‘Covid’ there for almost a year and when the residents were ‘vaccinated’ they had 19 positive cases in two weeks with eight dying.

It’s not the ‘vaccine’ – honest

The obvious cause and effect was being ignored by the media and most of the public. Australia’s health minister Greg Hunt (a former head of strategy at the World Economic Forum) was admitted to hospital after he had the ‘vaccine’. He was suffering according to reports from the skin infection ‘cellulitis’ and it must have been a severe case to have warranted days in hospital. Immediately the authorities said this was nothing to do with the ‘vaccine’ when an effect of some vaccines is a ‘cellulitis-like reaction’. We had families of perfectly healthy old people who died after the ‘vaccine’ saying that if only they had been given the ‘vaccine’ earlier they would still be alive. As a numbskull rating that is off the chart. A father of four ‘died of Covid’ at aged 48 when he was taken ill two days after having the ‘vaccine’. The man, a health administrator, had been ‘shielding during the pandemic’ and had ‘not really left the house’ until he went for the ‘vaccine’. Having the ‘vaccine’ and then falling ill and dying does not seem to have qualified as a possible cause and effect and ‘Covid-19’ went on his death certificate. His family said they had no idea how he ‘caught the virus’. A family member said: ‘Tragically, it could be that going for a vaccination ultimately led to him catching Covid …The sad truth is that they are never going to know where it came from.’ The family warned people to remember that the virus still existed and was ‘very real’. So was their stupidity. Nurses and doctors who had the first round of the ‘vaccine’ were collapsing, dying and ending up in a hospital bed while they or their grieving relatives were saying they’d still have the ‘vaccine’ again despite what happened. I kid you not. You mean if your husband returned from the dead he’d have the same ‘vaccine’ again that killed him??

Doctors at the VCU Medical Center in Richmond, Virginia, said the Johnson & Johnson ‘vaccine’ was to blame for a man’s skin peeling off. Patient Richard Terrell said: ‘It all just happened so fast. My skin peeled off. It’s still coming off on my hands now.’ He said it was stinging, burning and itching and when he bent his arms and legs it was very painful with ‘the skin swollen and rubbing against itself’. Pfizer/BioNTech and Moderna vaccines use mRNA to change the cell while the Johnson & Johnson version uses DNA in a process similar to AstraZeneca’s technique. Johnson & Johnson and AstraZeneca have both had their ‘vaccines’ paused by many countries after causing serious blood problems. Terrell’s doctor Fnu Nutan said he could have died if he hadn’t got medical attention. It sounds terrible so what did Nutan and Terrell say about the ‘vaccine’ now? Oh, they still recommend that people have it. A nurse in a hospital bed 40 minutes after the vaccination and unable to swallow due to throat swelling was told by a doctor that he lost mobility in his arm for 36 hours following the vaccination. What did he say to the ailing nurse? ‘Good for you for getting the vaccination.’ We are dealing with a serious form of cognitive dissonance madness in both public and medical staff. There is a remarkable correlation between those having the ‘vaccine’ and trumpeting the fact and suffering bad happenings shortly afterwards. Witold Rogiewicz, a Polish doctor, made a video of his ‘vaccination’ and ridiculed those who were questioning its safety and the intentions of Bill Gates: ‘Vaccinate yourself to protect yourself, your loved ones, friends and also patients. And to mention quickly I have info for anti-vaxxers and anti-Coviders if you want to contact Bill Gates you can do this through me.’ He further ridiculed the dangers of 5G. Days later he was dead, but naturally the vaccination wasn't mentioned in the verdict of 'heart attack'.

Lies, lies and more lies

So many members of the human race have slipped into extreme states of insanity and unfortunately they include reframed doctors and nursing staff. Having a 'vaccine' and dying within minutes or hours is not considered a valid connection while death from any cause within 28 days or longer of a positive test with a test not testing for the 'virus' means 'Covid-19' goes on the death certificate. How could that 'vaccine'-death connection not have been made except by calculated deceit? US figures in the initial rollout period to February 12th, 2020, revealed that a third of the deaths reported to the CDC after 'Covid vaccines' happened within 48 hours. Five men in the UK suffered an 'extremely rare' blood clot problem after having the AstraZeneca 'vaccine', but no causal link was established said the Gates-funded Medicines and Healthcare products Regulatory Agency (MHRA) which had given the 'vaccine' emergency approval to be used. Former Pfizer executive Dr Michael Yeadon explained in his interview how the procedures could cause blood coagulation and clots. People who should have been at no risk were dying from blood clots in the brain and he said he had heard from medical doctor friends that people were suffering from skin bleeding and massive headaches. The AstraZeneca 'shot' was stopped by some 20 countries over the blood clotting issue and still the corrupt MHRA, the European Medicines Agency (EMA) and the World Health Organization said that it should continue to be given even though the EMA admitted that it 'still cannot rule out definitively' a link between blood clotting and the 'vaccine'. Later Marco Cavaleri, head of EMA vaccine strategy, said there was indeed a clear link between the 'vaccine' and thrombosis, but they didn't know why. So much for the trials showing the 'vaccine' is safe. Blood clots were affecting younger people who would be under virtually no danger from 'Covid' even if it existed which makes it all the more stupid and sinister.

The British government responded to public alarm by wheeling out June Raine, the terrifyingly weak infant school headmistress sound-alike who heads the UK MHRA drug 'regulator'. The idea that she would stand up to Big Pharma and government pressure is laughable and she told us that all was well in the same way that she did when allowing untested, never-used-on-humans-before, genetically-manipulating 'vaccines' to be exposed to the public in the first place. Mass lying is the new normal of the 'Covid' era. The MHRA later said 30 cases of rare blood clots had by then been connected with the AstraZeneca 'vaccine' (that means a lot more in reality) while stressing that the benefits of the jab in preventing 'Covid-19' outweighed any risks. A more ridiculous and disingenuous statement with callous disregard for human health it is hard to contemplate. Immediately after the mendacious 'all-clears' two hospital workers in Denmark experienced blood clots and cerebral haemorrhaging following the AstraZeneca jab and one died. Top Norwegian health official Pål Andre Holme said the 'vaccine' was the only common factor: 'There is nothing in the patient history of these individuals that can give such a powerful immune response … I am confident that the antibodies that we have found are the cause, and I see no other explanation than it being the vaccine which triggers it.' Strokes, a clot or bleed in the brain, were clearly associated with the 'vaccine' from word of mouth and whistleblower reports. Similar consequences followed with all these 'vaccines' that we were told were so safe and as the numbers grew by the day it was clear we were witnessing human carnage.

Learning the hard way

A woman interviewed by UKColumn told how her husband suffered dramatic health effects after the vaccine when he'd been in good health all his life. He went from being a little unwell to losing all feeling in his legs and experiencing 'excruciating pain'. Misdiagnosis followed twice at Accident and Emergency (an 'allergy' and 'sciatica') before he was admitted to a neurology ward where doctors said his serious condition had been caused by the 'vaccine'. Another seven 'vaccinated' people were apparently being treated on the same ward for similar symptoms. The woman said he had the 'vaccine' because they believed media claims that it was safe. 'I didn't think the government would give out a vaccine that does this to somebody; I believed they would be bringing out a vaccination that would be safe.' What a tragic way to learn that lesson. Another woman posted that her husband was transporting stroke patients to hospital on almost every shift and when he asked them if they had been 'vaccinated' for 'Covid' they all replied 'yes'. One had a 'massive brain bleed' the day after his second dose. She said her husband reported the 'just been vaccinated' information every time to doctors in A and E only for them to ignore it, make no notes and appear annoyed that it was even mentioned. This particular report cannot be verified, but it expresses a common theme that confirms the monumental underreporting of 'vaccine' consequences. Interestingly as the 'vaccines' and their brain blood clot/stroke consequences began to emerge the UK National Health Service began a publicity campaign telling the public what to do in the event of a stroke. A Scottish NHS staff nurse who quit in disgust in March, 2021, said:

> I have seen traumatic injuries from the vaccine, they're not getting reported to the yellow card [adverse reaction] scheme, they're treating the symptoms, not asking why, why it's happening. It's just treating the symptoms and when you speak about it you're dismissed like you're crazy, I'm not crazy, I'm not crazy because every other colleague I've spoken to is terrified to speak out, they've had enough.

Videos appeared on the Internet of people uncontrollably shaking after the 'vaccine' with no control over muscles, limbs and even their face. A Scottish mother broke out in a severe rash all over her body almost immediately after she was given the AstraZeneca 'vaccine'. The pictures were horrific. Leigh King, a 41-year-old hairdresser from Lanarkshire said: 'Never in my life was I prepared for what I was about to experience … My skin was so sore and constantly hot … I have never felt pain like this …' But don't you worry, the 'vaccine' is perfectly safe. Then there has been the effect on medical staff who have been pressured to have the 'vaccine' by psychopathic 'health' authorities and government. A London hospital consultant who gave the name K. Polyakova wrote this to the *British Medical Journal* or *BMJ*:

> I am currently struggling with … the failure to report the reality of the morbidity caused by our current vaccination program within the health service and staff population. The levels of sickness after vaccination is unprecedented and staff are getting very sick and some with neurological symptoms which is having a huge impact on the health service function. Even the young and healthy are off for days, some for weeks, and some requiring medical treatment. Whole teams are being taken out as they went to get vaccinated together.
>
> Mandatory vaccination in this instance is stupid, unethical and irresponsible when it comes to protecting our staff and public health. We are in the voluntary phase of vaccination, and encouraging staff to take an unlicensed product that is impacting on their immediate health … it is clearly stated that these vaccine products do not offer immunity or stop transmission. In which case why are we doing it?

Not to protect health that's for sure. Medical workers are lauded by governments for agenda reasons when they couldn't give a toss about them any more than they can for the population in general. Schools across America faced the same situation as they closed due to the high number of teachers and other staff with bad reactions to the Pfizer/BioNTech, Moderna, and Johnson & Johnson 'Covid vaccines' all of which were linked to death and serious adverse effects. The *BMJ* took down the consultant's comments pretty quickly on the grounds that they were being used to spread 'disinformation'. They were exposing the truth about the 'vaccine' was the real reason. The cover-up is breathtaking.

Hiding the evidence

The scale of the 'vaccine' death cover-up worldwide can be confirmed by comparing official figures with the personal experience of the public. I heard of many people in my community who died immediately or soon after the vaccine that would never appear in the media or even likely on the official totals of 'vaccine' fatalities and adverse reactions when only about ten percent are estimated to be reported and I have seen some estimates as low as one percent in a Harvard study. In the UK alone by April 29th, 2021, some 757,654 adverse reactions had been officially reported from the Pfizer/BioNTech, Oxford/AstraZeneca and Moderna 'vaccines' with more than a thousand deaths linked to jabs and that means an estimated ten times this number in reality from a ten percent reporting rate percentage. That's seven million adverse reactions and 10,000 potential deaths and a one percent reporting rate would be ten times *those* figures. In 1976 the US government pulled the swine flu vaccine after *53* deaths. The UK data included a combined 10,000 eye disorders from the 'Covid vaccines' with more than 750 suffering visual impairment or blindness and again multiply by the estimated reporting percentages. As 'Covid cases' officially fell hospitals virtually empty during the 'Covid crisis' began to fill up with a range of other problems in the wake of the 'vaccine' rollout. The numbers across America have also been catastrophic. Deaths linked to *all* types of vaccine increased by *6,000 percent* in the first quarter of 2021 compared with 2020. A 39-year-old woman from Ogden, Utah, died four days after receiving a second dose of Moderna's 'Covid vaccine' when her liver, heart and kidneys all failed despite the fact that she had no known medical issues or conditions. Her family sought an autopsy, but Dr Erik Christensen, Utah's chief medical examiner, said proving vaccine injury as a cause of death almost never happened. He could think of only one instance where an autopsy would name a vaccine as the official cause of death and that would be anaphylaxis where someone received a vaccine and died almost instantaneously. 'Short of that, it would be difficult for us to definitively say this is the vaccine,' Christensen said. If that is true this must be added to the estimated ten percent (or far less) reporting rate of vaccine deaths and serious reactions and the conclusion can only be that vaccine deaths and serious reactions – including these 'Covid' potions' – are phenomenally understated in official figures. The same story can be found everywhere. Endless accounts of deaths and serious reactions among the public, medical and care home staff while official figures did not even begin to reflect this.

Professional script-reader Dr David Williams, a 'top public-health official' in Ontario, Canada, insulted our intelligence by claiming only four serious adverse reactions and no deaths from the more than 380,000 vaccine doses then given. This bore no resemblance to what people knew had happened in their owns circles and we had Dirk Huyer in charge of getting millions vaccinated in Ontario while at the same time he was Chief Coroner for the province investigating causes of death including possible death from the vaccine. An aide said he had stepped back from investigating deaths, but evidence indicated otherwise. Rosemary Frei, who secured a Master of Science degree in molecular biology at the Faculty of Medicine at Canada's University of Calgary before turning to investigative journalism, was one who could see that official figures for 'vaccine' deaths and reactions made no sense. She said that doctors seldom reported adverse events and when people got really sick or died after getting a vaccination they would attribute that to anything except the vaccines. It had been that way for years and anyone who wondered aloud whether the 'Covid vaccines' or other shots cause harm is immediately branded as 'anti-vax' and 'anti-science'. This was 'career-threatening' for health professionals. Then there was the huge pressure to support the push to 'vaccinate' billions in the quickest time possible. Frei said:

> So that's where we're at today. More than half a million vaccine doses have been given to people in Ontario alone. The rush is on to vaccinate all 15 million of us in the province by September. And the mainstream media are screaming for this to be sped up even more. That all adds up to only a very slim likelihood that we're going to be told the truth by officials about how many people are getting sick or dying from the vaccines.

What is true of Ontario is true of everywhere.

They KNEW – and still did it

The authorities knew what was going to happen with multiple deaths and adverse reactions. The UK government's Gates-funded and Big Pharma-dominated Medicines and Healthcare products Regulatory Agency (MHRA) hired a company to employ AI in compiling the projected reactions to the 'vaccine' that would otherwise be uncountable. The request for applications said: 'The MHRA urgently seeks an Artificial Intelligence (AI) software tool to process the expected high volume of Covid-19 vaccine Adverse Drug Reaction …' This was from the agency, headed by the disingenuous June Raine, that gave the 'vaccines' emergency approval and the company was hired before the first shot was given. 'We are going to kill and maim you – is that okay?' 'Oh, yes, perfectly fine – I'm very grateful, thank you, doctor.' The range of 'Covid vaccine' adverse reactions goes on for page after page in the MHRA criminally underreported 'Yellow Card' system and includes affects to eyes, ears, skin, digestion, blood and so on. Raine's MHRA amazingly claimed that the 'overall safety experience … is so far as expected from the clinical trials'. The death, serious adverse effects, deafness and blindness were *expected?* When did they ever mention that? If these human tragedies were expected then those that gave approval for the use of these 'vaccines' must be guilty of crimes against humanity including murder – a definition of which is 'killing a person with malice aforethought or with recklessness manifesting extreme indifference to the value of human life.' People involved at the MHRA, the CDC in America and their equivalent around the world must go before Nuremberg trials to answer for their callous inhumanity. We are only talking here about the immediate effects of the 'vaccine'. The longer-term impact of the DNA synthetic manipulation is the main reason they are so hysterically desperate to inoculate the entire global population in the shortest possible time.

Africa and the developing world are a major focus for the 'vaccine' depopulation agenda and a mass vaccination sales-pitch is underway thanks to caring people like the Rockefellers and other Cult assets. The Rockefeller Foundation, which pre-empted the 'Covid pandemic' in a document published in 2010 that 'predicted' what happened a decade later, announced an initial $34.95 million grant in February, 2021, 'to ensure more equitable access to Covid-19 testing and vaccines' among other things in Africa in collaboration with '24 organizations, businesses, and government agencies'. The pan-Africa initiative would focus on 10 countries: Burkina Faso, Ethiopia, Ghana, Kenya, Nigeria, Rwanda, South Africa, Tanzania, Uganda, and Zambia'. Rajiv Shah, President of the Rockefeller Foundation and former administrator of CIA-controlled USAID, said that if Africa was not mass-vaccinated (to change the DNA of its people) it was a 'threat to all of humanity' and not fair on Africans. When someone from the Rockefeller Foundation says they want to do something to help poor and deprived people and countries it is time for a belly-laugh. They are doing this out of the goodness of their 'heart' because 'vaccinating' the entire global population is what the 'Covid' hoax set out to achieve. Official 'decolonisation' of Africa by the Cult was merely a prelude to financial colonisation on the road to a return to physical colonisation. The 'vaccine' is vital to that and the sudden and convenient death of the 'Covid' sceptic president of Tanzania can be seen in its true light. A lot of people in Africa are aware that this is another form of colonisation and exploitation and they need to stand their ground.

The 'vaccine is working' scam

A potential problem for the Cult was that the 'vaccine' is meant to change human DNA and body messaging and not to protect anyone from a 'virus' never shown to exist. The vaccine couldn't work because it was not designed to work and how could they make it *appear* to be working so that more people would have it? This was overcome by lowering the amplification rate of the PCR test to produce fewer 'cases' and therefore fewer 'deaths'. Some of us had been pointing out since March, 2020, that the amplification rate of the test not testing for the 'virus' had been made artificially high to generate positive tests which they could call 'cases' to justify lockdowns. The World Health Organization recommended an absurdly high 45 amplification cycles to ensure the high positives required by the Cult and then remained silent on the issue until January 20th, 2021 – Biden's Inauguration Day. This was when the 'vaccinations' were seriously underway and on that day the WHO recommended after discussions with America's CDC that laboratories *lowered their testing amplification*. Dr David Samadi, a certified urologist and health writer, said the WHO was encouraging all labs to reduce their cycle count for PCR tests. He said the current cycle was much too high and was 'resulting in any particle being declared a positive case'. Even one mainstream news report I saw said this meant the number of 'Covid' infections may have been 'dramatically inflated'. Oh, just a little bit. The CDC in America issued new guidance to laboratories in April, 2021, to use 28 cycles *but only for 'vaccinated' people*. The timing of the CDC/WHO interventions were cynically designed to make it appear the 'vaccines' were responsible for falling cases and deaths when the real reason can be seen in the following examples. New York's state lab, the Wadsworth Center, identified 872 positive tests in July, 2020, based on a threshold of 40 cycles. When the figure was lowered to 35 cycles *43 percent* of the 872 were no longer 'positives'. At 30 cycles the figure was 63 percent. A Massachusetts lab found that between *85 to 90 percent* of people who tested positive in July with a cycle threshold of 40 would be negative at 30 cycles, Ashish Jha, MD, director of the Harvard Global Health Institute, said: 'I'm really shocked that it could be that high … Boy, does it really change the way we need to be thinking about testing.' I'm shocked that I could see the obvious in the spring of 2020, with no medical background, and most medical professionals still haven't worked it out. No, that's not shocking – it's terrifying.

Three weeks after the WHO directive to lower PCR cycles the London *Daily Mail* ran this headline: 'Why ARE Covid cases plummeting? New infections have fallen 45% in the US and 30% globally in the past 3 weeks but experts say vaccine is NOT the main driver because only 8% of Americans and 13% of people worldwide have received their first dose.' They acknowledged that the drop could not be attributed to the 'vaccine', but soon this morphed throughout the media into the 'vaccine' has caused cases and deaths to fall when it was the PCR threshold. In December, 2020, there was chaos at English Channel ports with truck drivers needing negative ‘Covid’ tests before they could board a ferry home for Christmas. The government wanted to remove the backlog as fast as possible and they brought in troops to do the ‘testing’. Out of 1,600 drivers just *36* tested positive and the rest were given the all clear to cross the Channel. I guess the authorities thought that 36 was the least they could get away with without the unquestioning catching on. The amplification trick which most people believed in the absence of information in the mainstream applied more pressure on those refusing the ‘vaccine’ to succumb when it ‘obviously worked’. The truth was the exact opposite with deaths in care homes soaring with the ‘vaccine’ and in Israel the term used was ‘skyrocket’. A re-analysis of published data from the Israeli Health Ministry led by Dr Hervé Seligmann at the Medicine Emerging Infectious and Tropical Diseases at Aix-Marseille University found that Pfizer’s ‘Covid vaccine’ killed ‘about 40 times more [elderly] people than the disease itself would have killed’ during a five-week vaccination period and *260 times* more younger people than would have died from the ‘virus’ even according to the manipulated ‘virus’ figures. Dr Seligmann and his co-study author, Haim Yativ, declared after reviewing the Israeli ‘vaccine’ death data: ‘This is a new Holocaust.’

Then, in mid-April, 2021, after vast numbers of people worldwide had been ‘vaccinated’, the story changed with clear coordination. The UK government began to prepare the ground for more future lockdowns when Nuremberg-destined Boris Johnson told yet another whopper. He said that cases had fallen because of *lockdowns* not ‘vaccines’. Lockdowns are irrelevant when *there is no ‘virus’* and the test and fraudulent death certificates are deciding the number of ‘cases’ and ‘deaths’. Study after study has shown that lockdowns don’t work and instead kill and psychologically destroy people. Meanwhile in the United States Anthony Fauci and Rochelle Walensky, the ultra-Zionist head of the CDC, peddled the same line. More lockdown was the answer and not the ‘vaccine’, a line repeated on cue by the moron that is Canadian Prime Minister Justin Trudeau. Why all the hysteria to get everyone ‘vaccinated’ if lockdowns and not ‘vaccines’ made the difference? None of it makes sense on the face of it. Oh, but it does. The Cult wants lockdowns *and* the ‘vaccine’ and if the ‘vaccine’ is allowed to be seen as the total answer lockdowns would no longer be justified when there are still livelihoods to destroy. ‘Variants’ and renewed upward manipulation of PCR amplification are planned to instigate never-ending lockdown *and* more ‘vaccines’.

You *must* have it – we’re desperate

Israel, where the Jewish and Arab population are ruled by the Sabbatian Cult, was the front-runner in imposing the DNA-manipulating ‘vaccine’ on its people to such an extent that Jewish refusers began to liken what was happening to the early years of Nazi Germany. This would seem to be a fantastic claim. Why would a government of Jewish people be acting like the Nazis did? If you realise that the Sabbatian Cult was behind the Nazis and that Sabbatians hate Jews the pieces start to fit and the question of why a ‘Jewish’ government would treat Jews with such callous disregard for their lives and freedom finds an answer. Those controlling the government of Israel *aren’t Jewish* – they’re Sabbatian. Israeli lawyer Tamir Turgal was one who made the Nazi comparison in comments to German lawyer Reiner Fuellmich who is leading a class action lawsuit against the psychopaths for crimes against humanity. Turgal described how the Israeli government was vaccinating children and pregnant women on the basis that there was no evidence that this was dangerous when they had no evidence that it *wasn’t* dangerous either. They just had no evidence. This was medical experimentation and Turgal said this breached the Nuremberg Code about medical experimentation and procedures requiring informed consent and choice. Think about that. A Nuremberg Code developed because of Nazi experimentation on Jews and others in concentration camps by people like the evil-beyond-belief Josef Mengele is being breached by the *Israeli* government; but when you know that it’s a *Sabbatian* government along with its intelligence and military agencies like Mossad, Shin Bet and the Israeli Defense Forces, and that Sabbatians were the force behind the Nazis, the kaleidoscope comes into focus. What have we come to when Israeli Jews are suing their government for violating the Nuremberg Code by essentially making Israelis subject to a medical experiment using the controversial 'vaccines'? It's a shocker that this has to be done in the light of what happened in Nazi Germany. The Anshe Ha-Emet, or 'People of the Truth', made up of Israeli doctors, lawyers, campaigners and public, have launched a lawsuit with the International Criminal Court. It says:

> When the heads of the Ministry of Health as well as the prime minister presented the vaccine in Israel and began the vaccination of Israeli residents, the vaccinated were not advised, that, in practice, they are taking part in a medical experiment and that their consent is required for this under the Nuremberg Code.

The irony is unbelievable, but easily explained in one word: Sabbatians. The foundation of Israeli 'Covid' apartheid is the 'green pass' or 'green passport' which allows Jews and Arabs who have had the DNA-manipulating 'vaccine' to go about their lives – to work, fly, travel in general, go to shopping malls, bars, restaurants, hotels, concerts, gyms, swimming pools, theatres and sports venues, while non-'vaccinated' are banned from all those places and activities. Israelis have likened the 'green pass' to the yellow stars that Jews in Nazi Germany were forced to wear – the same as the yellow stickers that a branch of UK supermarket chain Morrisons told exempt mask-wears they had to display when shopping. How very sensitive. The Israeli system is blatant South African-style apartheid on the basis of compliance or non-compliance to fascism rather than colour of the skin. How appropriate that the Sabbatian Israeli government was so close to the pre-Mandela apartheid regime in Pretoria. The Sabbatian-instigated 'vaccine passport' in Israel is planned for everywhere. Sabbatians struck a deal with Pfizer that allowed them to lead the way in the percentage of a national population infused with synthetic material and the result was catastrophic. Israeli freedom activist Shai Dannon told me how chairs were appearing on beaches that said 'vaccinated only'. Health Minister Yuli Edelstein said that anyone unwilling or unable to get the jabs that 'confer immunity' will be 'left behind'. The man's a liar. Not even the makers claim the 'vaccines' confer immunity. When you see those figures of 'vaccine' deaths these psychopaths were saying that you must take the chance the 'vaccine' will kill you or maim you while knowing it will change your DNA or lockdown for you will be permanent. That's fascism. The Israeli parliament passed a law to allow personal information of the non-vaccinated to be shared with local and national authorities for three months. This was claimed by its supporters to be a way to 'encourage' people to be vaccinated. Hadas Ziv from Physicians for Human Rights described this as a 'draconian law which crushed medical ethics and the patient rights'. But that's the idea, the Sabbatians would reply.

Your papers, please

Sabbatian Israel was leading what has been planned all along to be a global 'vaccine pass' called a 'green passport' without which you would remain in permanent lockdown restriction and unable to do anything. This is how badly – *desperately* – the Cult is to get everyone 'vaccinated'. The term and colour 'green' was not by chance and related to the psychology of fusing the perception of the green climate hoax with the 'Covid' hoax and how the 'solution' to both is the same Great Reset. Lying politicians, health officials and psychologists denied there were any plans for mandatory vaccinations or restrictions based on vaccinations, but they knew that was exactly what was meant to happen with governments of all countries reaching agreements to enforce a global system. 'Free' Denmark and 'free' Sweden unveiled digital vaccine certification. Cyprus, Czech Republic, Estonia, Greece, Hungary, Iceland, Italy, Poland, Portugal, Slovakia, and Spain have all committed to a vaccine passport system and the rest including the whole of the EU would follow. The satanic UK government will certainly go this way despite mendacious denials and at the time of writing it is trying to manipulate the public into having the 'vaccine' so they could go abroad on a summer holiday. How would that work without something to prove you had the synthetic toxicity injected into you?

Documents show that the EU's European Commission was moving towards 'vaccine certificates' in 2018 and 2019 before the 'Covid' hoax began. They knew what was coming. Abracadabra – Ursula von der Leyen, the German President of the Commission, announced in March, 2021, an EU 'Digital Green Certificate' – green again – to track the public's 'Covid status'. The passport sting is worldwide and the Far East followed the same pattern with South Korea ruling that only those with 'vaccination' passports – again the *green* pass – would be able to 'return to their daily lives'.

Bill Gates has been preparing for this 'passport' with other Cult operatives for years and beyond the paper version is a Gates-funded 'digital tattoo' to identify who has been vaccinated and who hasn't. The 'tattoo' is reported to include a substance which is externally readable to confirm who has been vaccinated. This is a bio-luminous light-generating enzyme (think fireflies) called … *Luciferase*. Yes, named after the Cult 'god' Lucifer the 'light bringer' of whom more to come. Gates said he funded the readable tattoo to ensure children in the developing world were vaccinated and no one was missed out. He cares so much about poor kids as we know. This was just the cover story to develop a vaccine tagging system for everyone on the planet. Gates has been funding the ID2020 'alliance' to do just that in league with other lovely people at Microsoft, GAVI, the Rockefeller Foundation, Accenture and IDEO.org. He said in interviews in March, 2020, before any 'vaccine' publicly existed, that the world must have a globalised digital certificate to track the 'virus' and who had been vaccinated. Gates knew from the start that the mRNA vaccines were coming and when they would come and that the plan was to tag the 'vaccinated' to marginalise the intelligent and stop them doing anything including travel. Evil just doesn't suffice. Gates was exposed for offering a $10 million bribe to the Nigerian House of Representatives to invoke compulsory 'Covid' vaccination of all Nigerians. Sara Cunial, a member of the Italian Parliament, called Gates a 'vaccine criminal'. She urged the Italian President to hand him over to the International Criminal Court for crimes against humanity and condemned his plans to 'chip the human race' through ID2020.

You know it's a long-planned agenda when war criminal and Cult gofer Tony Blair is on the case. With the scale of arrogance only someone as dark as Blair can muster he said: 'Vaccination in the end is going to be your route to liberty.' Blair is a disgusting piece of work and he confirms that again. The media has given a lot of coverage to a bloke called Charlie Mullins, founder of London's biggest independent plumbing company, Pimlico Plumbers, who has said he won't employ anyone who has not been vaccinated or have them go to any home where people are not vaccinated. He said that if he had his way no one would be allowed to walk the streets if they have not been vaccinated. Gates was cheering at the time while I was alerting the white coats. The plan is that people will qualify for 'passports' for having the first two doses and then to keep it they will have to have all the follow ups and new ones for invented 'variants' until human genetics is transformed and many are dead who can't adjust to the changes. Hollywood celebrities – the usual propaganda stunt – are promoting something called the WELL Health-Safety Rating to verify that a building or space has 'taken the necessary steps to prioritize the health and safety of their staff, visitors and other stakeholders'. They included Lady Gaga, Jennifer Lopez, Michael B. Jordan, Robert DeNiro, Venus Williams, Wolfgang Puck, Deepak Chopra and 17th Surgeon General Richard Carmona. Yawn. WELL Health-Safety has big connections with China. Parent company Delos is headed by former Goldman Sachs partner Paul Scialla. This is another example – and we will see so many others – of using the excuse of 'health' to dictate the lives and activities of the population. I guess one confirmation of the 'safety' of buildings is that only 'vaccinated' people can go in, right?

Electronic concentration camps

I wrote decades ago about the plans to restrict travel and here we are for those who refuse to bow to tyranny. This can be achieved in one go with air travel if the aviation industry makes a blanket decree.

The ‘vaccine’ and guaranteed income are designed to be part of a global version of China’s social credit system which tracks behaviour 24/7 and awards or deletes ‘credits’ based on whether your behaviour is supported by the state or not. I mean your entire lifestyle – what you do, eat, say, everything. Once your credit score falls below a certain level consequences kick in. In China tens of millions have been denied travel by air and train because of this. All the locations and activities denied to refusers by the ‘vaccine’ passports will be included in one big mass ban on doing almost anything for those that don’t bow their head to government. It’s beyond fascist and a new term is required to describe its extremes – I guess fascist technocracy will have to do. The way the Chinese system of technological – technocratic – control is sweeping the West can be seen in the Los Angeles school system and is planned to be expanded worldwide. Every child is required to have a ‘Covid’-tracking app scanned daily before they can enter the classroom. The so-called Daily Pass tracking system is produced by Gates’ Microsoft which I’m sure will shock you rigid. The pass will be scanned using a barcode (one step from an inside-the-body barcode) and the information will include health checks, ‘Covid’ tests and vaccinations. Entry codes are for one specific building only and access will only be allowed if a student or teacher has a negative test with a test not testing for the ‘virus’, has no symptoms of anything alleged to be related to ‘Covid’ (symptoms from a range of other illness), and has a temperature under 100 degrees. No barcode, no entry, is planned to be the case for everywhere and not only schools.

Kids are being psychologically prepared to accept this as ‘normal’ their whole life which is why what they can impose in schools is so important to the Cult and its gofers. Long-time American freedom campaigner John Whitehead of the Rutherford Institute was not exaggerating when he said: ‘Databit by databit, we are building our own electronic concentration camps.’ Canada under its Cult gofer prime minister Justin Trudeau has taken a major step towards the real thing with people interned against their will if they test positive with a test not testing for the ‘virus’ when they arrive at a Canadian airport. They are jailed in internment hotels often without food or water for long periods and with many doors failing to lock there have been sexual assaults. The interned are being charged sometimes $2,000 for the privilege of being abused in this way. Trudeau is fully on board with the Cult and says the 'Covid pandemic' has provided an opportunity for a global 'reset' to permanently change Western civilisation. His number two, Deputy Prime Minister Chrystia Freeland, is a trustee of the World Economic Forum and a Rhodes Scholar. The Trudeau family have long been servants of the Cult. See *The Biggest Secret* and Cathy O'Brien's book *Trance-Formation of America* for the horrific background to Trudeau's father Pierre Trudeau another Canadian prime minister. Hide your fascism behind the façade of a heart-on-the-sleeve liberal. It's a well-honed Cult technique.

What can the 'vaccine' *really* do?

We have a 'virus' never shown to exist and 'variants' of the 'virus' that have also never been shown to exist except, like the 'original', as computer-generated fictions. Even if you believe there's a 'virus' the 'case' to 'death' rate is in the region of 0.23 to 0.15 percent and those 'deaths' are concentrated among the very old around the same average age that people die anyway. In response to this lack of threat (in truth none) psychopaths and idiots, knowingly and unknowingly answering to Gates and the Cult, are seeking to 'vaccinate' every man, woman and child on Planet Earth. Clearly the 'vaccine' is not about 'Covid' – none of this ever has been. So what is it all about *really*? Why the desperation to infuse genetically-manipulating synthetic material into everyone through mRNA fraudulent 'vaccines' with the intent of doing this over and over with the excuses of 'variants' and other 'virus' inventions? Dr Sherri Tenpenny, an osteopathic medical doctor in the United States, has made herself an expert on vaccines and their effects as a vehement campaigner against their use. Tenpenny was board certified in emergency medicine, the director of a level two trauma centre for 12 years, and moved to Cleveland in 1996 to start an integrative medicine practice which has treated patients from all 50 states and some 17 other countries. Weaning people off pharmaceutical drugs is a speciality.

She became interested in the consequences of vaccines after attending a meeting at the National Vaccine Information Center in Washington DC in 2000 where she 'sat through four days of listening to medical doctors and scientists and lawyers and parents of vaccine injured kids' and asked: 'What's going on?' She had never been vaccinated and never got ill while her father was given a list of vaccines to be in the military and was 'sick his entire life'. The experience added to her questions and she began to examine vaccine documents from the Centers for Disease Control (CDC). After reading the first one, the 1998 version of *The General Recommendations of Vaccination*, she thought: 'This is it?' The document was poorly written and bad science and Tenpenny began 20 years of research into vaccines that continues to this day. She began her research into 'Covid vaccines' in March, 2020, and she describes them as 'deadly'. For many, as we have seen, they already have been. Tenpenny said that in the first 30 days of the 'vaccine' rollout in the United States there had been more than 40,000 adverse events reported to the vaccine adverse event database. A document had been delivered to her the day before that was 172 pages long. 'We have over 40,000 adverse events; we have over 3,100 cases of [potentially deadly] anaphylactic shock; we have over 5,000 neurological reactions.' Effects ranged from headaches to numbness, dizziness and vertigo, to losing feeling in hands or feet and paraesthesia which is when limbs 'fall asleep' and people have the sensation of insects crawling underneath their skin. All this happened in the first 30 days and remember that only about *ten percent* (or far less) of adverse reactions and vaccine-related deaths are estimated to be officially reported. Tenpenny said:

> So can you think of one single product in any industry, any industry, for as long as products have been made on the planet that within 30 days we have 40,000 people complaining of side effects that not only is still on the market but … we've got paid actors telling us how great

> they are for getting their vaccine. We're offering people $500 if they will just get their vaccine and we've got nurses and doctors going; 'I got the vaccine, I got the vaccine'.

Tenpenny said they were not going to be 'happy dancing folks' when they began to suffer Bell's palsy (facial paralysis), neuropathies, cardiac arrhythmias and autoimmune reactions that kill through a blood disorder. 'They're not going to be so happy, happy then, but we're never going to see pictures of those people' she said. Tenpenny described the 'vaccine' as 'a well-designed killing tool'.

No off-switch

Bad as the initial consequences had been Tenpenny said it would be maybe 14 months before we began to see the 'full ravage' of what is going to happen to the 'Covid vaccinated' with full-out consequences taking anything between two years and 20 years to show. You can understand why when you consider that variations of the 'Covid vaccine' use mRNA (messenger RNA) to in theory activate the immune system to produce protective antibodies without using the actual 'virus'. How can they when it's a computer program and they've never isolated what they claim is the 'real thing'? Instead they use *synthetic* mRNA. They are inoculating synthetic material into the body which through a technique known as the Trojan horse is absorbed into cells to change the nature of DNA. Human DNA is changed by an infusion of messenger RNA and with each new 'vaccine' of this type it is changed even more. Say so and you are banned by Cult Internet platforms. The contempt the contemptuous Mark Zuckerberg has for the truth and human health can be seen in an internal Facebook video leaked to the Project Veritas investigative team in which he said of the 'Covid vaccines': '… I share some caution on this because we just don't know the long term side-effects of basically modifying people's DNA and RNA.' At the same time this disgusting man's Facebook was censoring and banning anyone saying exactly the same. He must go before a Nuremberg trial for crimes against humanity when he *knows* that he is censoring legitimate concerns and denying the right of informed consent on behalf of the Cult that owns him. People have been killed and damaged by the very 'vaccination' technique he cast doubt on himself when they may not have had the 'vaccine' with access to information that he denied them. The plan is to have at least annual 'Covid vaccinations', add others to deal with invented 'variants', and change all other vaccines into the mRNA system. Pfizer executives told shareholders at a virtual Barclays Global Healthcare Conference in March, 2021, that the public may need a third dose of 'Covid vaccine', plus regular yearly boosters and the company planned to hike prices to milk the profits in a 'significant opportunity for our vaccine'. These are the professional liars, cheats and opportunists who are telling you their 'vaccine' is safe. Given this volume of mRNA planned to be infused into the human body and its ability to then replicate we will have a transformation of human genetics from biological to synthetic biological – exactly the long-time Cult plan for reasons we'll see – and many will die. Sherri Tenpenny said of this replication:

> It's like having an on-button but no off-button and that whole mechanism … they actually give it a name and they call it the Trojan horse mechanism, because it allows that [synthetic] virus and that piece of that [synthetic] virus to get inside of your cells, start to replicate and even get inserted into other parts of your DNA as a Trojan-horse.

Ask the overwhelming majority of people who have the 'vaccine' what they know about the contents and what they do and they would reply: 'The government says it will stop me getting the virus.' Governments give that false impression on purpose to increase take-up. You can read Sherri Tenpenny's detailed analysis of the health consequences in her blog at Vaxxter.com, but in summary these are some of them. She highlights the statement by Bill Gates about how human beings can become their own 'vaccine manufacturing machine'. The man is insane. ['Vaccine'-generated] 'antibodies' carry synthetic messenger RNA into the cells and the damage starts, Tenpenny contends, and she says that lungs can be adversely affected through varying degrees of pus and bleeding which obviously affects breathing and would be dubbed ‘Covid-19’. Even more sinister was the impact of ‘antibodies’ on macrophages, a white blood cell of the immune system. They consist of Type 1 and Type 2 which have very different functions. She said Type 1 are ‘hyper-vigilant’ white blood cells which ‘gobble up’ bacteria etc. However, in doing so, this could cause inflammation and in extreme circumstances be fatal. She says these affects are mitigated by Type 2 macrophages which kick in to calm down the system and stop it going rogue. They clear up dead tissue debris and reduce inflammation that the Type 1 ‘fire crews’ have caused. Type 1 kills the infection and Type 2 heals the damage, she says. This is her punchline with regard to ‘Covid vaccinations’: She says that mRNA ‘antibodies’ block Type 2 macrophages by attaching to them and deactivating them. This meant that when the Type 1 response was triggered by infection there was nothing to stop that getting out of hand by calming everything down. There’s an on-switch, but no off-switch, she says. What follows can be ‘over and out, see you when I see you’.

Genetic suicide

Tenpenny also highlights the potential for autoimmune disease – the body attacking itself – which has been associated with vaccines since they first appeared. Infusing a synthetic foreign substance into cells could cause the immune system to react in a panic believing that the body is being overwhelmed by an invader (it is) and the consequences can again be fatal. There is an autoimmune response known as a ‘cytokine storm’ which I have likened to a homeowner panicked by an intruder and picking up a gun to shoot randomly in all directions before turning the fire on himself. The immune system unleashes a storm of inflammatory response called cytokines to a threat and the body commits hara-kiri. The lesson is that you mess with the body’s immune response at your peril and these ‘vaccines’ seriously – fundamentally – mess with immune response. Tenpenny refers to a consequence called anaphylactic shock which is a severe and highly dangerous allergic reaction when the immune system floods the body with chemicals. She gives the example of having a bee sting which primes the immune system and makes it sensitive to those chemicals. When people are stung again maybe years later the immune response can be so powerful that it leads to anaphylactic shock. Tenpenny relates this 'shock' with regard to the 'Covid vaccine' to something called polyethylene glycol or PEG. Enormous numbers of people have become sensitive to this over decades of use in a whole range of products and processes including food, drink, skin creams and 'medicine'. Studies have claimed that some 72 percent of people have antibodies triggered by PEG compared with two percent in the 1960s and allergic hypersensitive reactions to this become a gathering cause for concern. Tenpenny points out that the 'mRNA vaccine' is coated in a 'bubble' of polyethylene glycol which has the potential to cause anaphylactic shock through immune sensitivity. Many reports have appeared of people reacting this way after having the 'Covid vaccine'. What do we think is going to happen as humanity has more and more of these 'vaccines'? Tenpenny said: 'All these pictures we have seen with people with these rashes … these weepy rashes, big reactions on their arms and things like that – it's an acute allergic reaction most likely to the polyethylene glycol that you've been previously primed and sensitised to.'

Those who have not studied the conspiracy and its perpetrators at length might think that making the population sensitive to PEG and then putting it in these 'vaccines' is just a coincidence. It is not. It is instead testament to how carefully and coldly-planned current events have been and the scale of the conspiracy we are dealing with. Tenpenny further explains that the 'vaccine' mRNA procedure can breach the blood-brain barrier which protects the brain from toxins and other crap that will cause malfunction. In this case they could make two proteins corrupt brain function to cause Amyotrophic lateral sclerosis (ALS) , a progressive nervous system disease leading to loss of muscle control, and frontal lobe degeneration – Alzheimer's and dementia. Immunologist J. Bart Classon published a paper connecting mRNA 'vaccines' to prion disease which can lead to Alzheimer's and other forms of neurogenerative disease while others have pointed out the potential to affect the placenta in ways that make women infertile. This will become highly significant in the next chapter when I will discuss other aspects of this non-vaccine that relate to its nanotechnology and transmission from the injected to the uninjected.

Qualified in idiocy

Tenpenny describes how research has confirmed that these 'vaccine'-generated antibodies can interact with a range of other tissues in the body and attack many other organs including the lungs. 'This means that if you have a hundred people standing in front of you that all got this shot they could have a hundred different symptoms.' Anyone really think that Cult gofers like the Queen, Tony Blair, Christopher Whitty, Anthony Fauci, and all the other psychopaths have really had this 'vaccine' in the pictures we've seen? Not a bloody chance. Why don't doctors all tell us about all these dangers and consequences of the 'Covid vaccine'? Why instead do they encourage and pressure patients to have the shot? Don't let's think for a moment that doctors and medical staff can't be stupid, lazy, and psychopathic and that's without the financial incentives to give the jab. Tenpenny again:

> Some people are going to die from the vaccine directly but a large number of people are going to start to get horribly sick and get all kinds of autoimmune diseases 42 days to maybe a year out. What are they going to do, these stupid doctors who say; 'Good for you for getting that vaccine.' What are they going to say; 'Oh, it must be a mutant, we need to give an extra dose of that vaccine.'
>
> Because now the vaccine, instead of one dose or two doses we need three or four because the stupid physicians aren't taking the time to learn anything about it. If I can learn this sitting in my living room reading a 19 page paper and several others so can they. There's nothing special about me, I just take the time to do it.

Remember how Sara Kayat, the NHS and TV doctor, said that the 'Covid vaccine' would '100 percent prevent hospitalisation and death'. Doctors can be idiots like every other profession and they should not be worshipped as infallible. They are not and far from it. Behind many medical and scientific 'experts' lies an uninformed prat trying to hide themselves from you although in the 'Covid' era many have failed to do so as with UK narrative-repeating 'TV doctor' Hilary Jones. Pushing back against the minority of proper doctors and scientists speaking out against the 'vaccine' has been the entire edifice of the Cult global state in the form of governments, medical systems, corporations, mainstream media, Silicon Valley, and an army of compliant doctors, medical staff and scientists willing to say anything for money and to enhance their careers by promoting the party line. If you do that you are an 'expert' and if you won't you are an 'anti-vaxxer' and 'Covidiot'. The pressure to be 'vaccinated' is incessant. We have even had reports claiming that the 'vaccine' can help cure cancer and Alzheimer's and make the lame walk. I am waiting for the announcement that it can bring you coffee in the morning and cook your tea. Just as the symptoms of 'Covid' seem to increase by the week so have the miracles of the 'vaccine'. American supermarket giant Kroger Co. offered nearly 500,000 employees in 35 states a $100 bonus for having the 'vaccine' while donut chain Krispy Kreme promised 'vaccinated' customers a free glazed donut every day for the rest of 2021. Have your DNA changed and you will get a doughnut although we might not have to give you them for long. Such offers and incentives confirm the desperation.

Perhaps the worse vaccine-stunt of them all was UK 'Health' Secretary Matt-the-prat Hancock on live TV after watching a clip of someone being 'vaccinated' when the roll-out began. Hancock faked tears so badly it was embarrassing. Brain-of-Britain Piers Morgan, the lockdown-supporting, 'vaccine' supporting, 'vaccine' passport-supporting, TV host played along with Hancock – 'You're quite emotional about that' he said in response to acting so atrocious it would have been called out at a school nativity which will presumably today include Mary and Jesus in masks, wise men keeping their camels six feet apart, and shepherds under tent arrest. System-serving Morgan tweeted this: 'Love the idea of covid vaccine passports for everywhere: flights, restaurants, clubs, football, gyms, shops etc. It's time covid-denying, anti-vaxxer loonies had their bullsh*t bluff called & bar themselves from going anywhere that responsible citizens go.' If only I could aspire to his genius. To think that Morgan, who specialises in shouting over anyone he disagrees with, was lauded as a free speech hero when he lost his job after storming off the set of his live show like a child throwing his dolly out of the pram. If he is a free speech hero we are in real trouble. I have no idea what 'bullsh*t' means, by the way, the * throws me completely.

The Cult is desperate to infuse its synthetic DNA-changing concoction into everyone and has been using every lie, trick and intimidation to do so. The question of '*Why*?' we shall now address.

CHAPTER TEN

Human 2.0

I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted –
Alan Turing (1912-1954), the 'Father of artificial intelligence'

I have been exposing for decades the plan to transform the human body from a biological to a synthetic-biological state. The new human that I will call Human 2.0 is planned to be connected to artificial intelligence and a global AI 'Smart Grid' that would operate as one global system in which AI would control everything from your fridge to your heating system to your car to your mind. Humans would no longer be 'human', but post-human and sub-human, with their thinking and emotional processes replaced by AI.

What I said sounded crazy and beyond science fiction and I could understand that. To any balanced, rational, mind it *is* crazy. Today, however, that world is becoming reality and it puts the 'Covid vaccine' into its true context. Ray Kurzweil is the ultra-Zionist 'computer scientist, inventor and futurist' and co-founder of the Singularity University. Singularity refers to the merging of humans with machines or 'transhumanism'. Kurzweil has said humanity would be connected to the cyber 'cloud' in the period of the ever-recurring year of 2030:

> Our thinking … will be a hybrid of biological and non-biological thinking … humans will be able to extend their limitations and 'think in the cloud' … We're going to put gateways to the

> cloud in our brains … We're going to gradually merge and enhance ourselves … In my view, that's the nature of being human – we transcend our limitations. As the technology becomes vastly superior to what we are then the small proportion that is still human gets smaller and smaller and smaller until it's just utterly negligible.

They are trying to sell this end-of-humanity-as-we-know-it as the next stage of 'evolution' when we become super-human and 'like the gods'. They are lying to you. Shocked, eh? The population, and again especially the young, have been manipulated into addiction to technologies designed to enslave them for life. First they induced an addiction to smartphones (holdables); next they moved to technology on the body (wearables); and then began the invasion of the body (implantables). I warned way back about the plan for microchipped people and we are now entering that era. We should not be diverted into thinking that this refers only to chips we can see. Most important are the nanochips known as smart dust, neural dust and nanobots which are far too small to be seen by the human eye. Nanotechnology is everywhere, increasingly in food products, and released into the atmosphere by the geoengineering of the skies funded by Bill Gates to 'shut out the Sun' and 'save the planet from global warming'. Gates has been funding a project to spray millions of tonnes of chalk (calcium carbonate) into the stratosphere over Sweden to 'dim the Sun' and cool the Earth. Scientists warned the move could be disastrous for weather systems in ways no one can predict and opposition led to the Swedish space agency announcing that the 'experiment' would not be happening as planned in the summer of 2021; but it shows where the Cult is going with dimming the impact of the Sun and there's an associated plan to change the planet's atmosphere. Who gives psychopath Gates the right to dictate to the entire human race and dismantle planetary systems? The world will not be safe while this man is at large.

The global warming hoax has made the Sun, like the gas of life, something to fear when both are essential to good health and human survival (more inversion). The body transforms sunlight into vital vitamin D through a process involving … *cholesterol*. This is the cholesterol we are also told to fear. We are urged to take Big Pharma statin drugs to reduce cholesterol and it's all systematic. Reducing cholesterol means reducing vitamin D uptake with all the multiple health problems that will cause. At least if you take statins long term it saves the government from having to pay you a pension. The delivery system to block sunlight is widely referred to as chemtrails although these have a much deeper agenda, too. They appear at first to be contrails or condensation trails streaming from aircraft into cold air at high altitudes. Contrails disperse very quickly while chemtrails do not and spread out across the sky before eventually their content falls to earth. Many times I have watched aircraft cross-cross a clear blue sky releasing chemtrails until it looks like a cloudy day. Chemtrails contain many things harmful to humans and the natural world including toxic heavy metals, aluminium (see Alzheimer's) and nanotechnology. Ray Kurzweil reveals the reason without actually saying so: 'Nanobots will infuse all the matter around us with information. Rocks, trees, everything will become these intelligent creatures.' How do you deliver that? *From the sky*. Self-replicating nanobots would connect everything to the Smart Grid. The phenomenon of Morgellons disease began in the chemtrail era and the correlation has led to it being dubbed the 'chemtrail disease'. Self-replicating fibres appear in the body that can be pulled out through the skin. Morgellons fibres continue to grow outside the body and have a form of artificial intelligence. I cover this at greater length in *Phantom Self*.

'Vaccine' operating system

'Covid vaccines' with their self-replicating synthetic material are also designed to make the connection between humanity and Kurzweil's 'cloud'. American doctor and dedicated campaigner for truth, Carrie Madej, an Internal Medicine Specialist in Georgia with more than 20 years medical experience, has highlighted the nanotechnology aspect of the fake 'vaccines'. She explains how one of the components in at least the Moderna and Pfizer synthetic potions are 'lipid nanoparticles' which are 'like little tiny computer bits' – a 'sci-fi substance' known as nanobots and hydrogel which can be 'triggered at any moment to deliver its payload' and act as 'biosensors'. The synthetic substance had 'the ability to accumulate data from your body like your breathing, your respiration, thoughts and emotions, all kind of things' and each syringe could carry a *million* nanobots:

> This substance because it's like little bits of computers in your body, crazy, but it's true, it can do that, [and] obviously has the ability to act through Wi-Fi. It can receive and transmit energy, messages, frequencies or impulses. That issue has never been addressed by these companies. What does that do to the human?
>
> Just imagine getting this substance in you and it can react to things all around you, the 5G, your smart device, your phones, what is happening with that? What if something is triggering it, too, like an impulse, a frequency? We have something completely foreign in the human body.

Madej said her research revealed that electromagnetic (EMF) frequencies emitted by phones and other devices had increased dramatically in the same period of the 'vaccine' rollout and she was seeing more people with radiation problems as 5G and other electromagnetic technology was expanded and introduced to schools and hospitals. She said she was 'floored with the EMF coming off' the devices she checked. All this makes total sense and syncs with my own work of decades when you think that Moderna refers in documents to its mRNA 'vaccine' as an 'operating system':

> Recognizing the broad potential of mRNA science, we set out to create an mRNA technology platform that functions very much like an operating system on a computer. It is designed so that it can plug and play interchangeably with different programs. In our case, the 'program' or 'app' is our mRNA drug – the unique mRNA sequence that codes for a protein …
>
> … Our MRNA Medicines – 'The 'Software Of Life': When we have a concept for a new mRNA medicine and begin research, fundamental components are already in place. Generally, the only thing that changes from one potential mRNA medicine to another is the coding region – the actual genetic code that instructs ribosomes to make protein. Utilizing these instruction sets gives our investigational mRNA medicines a software-like quality. We also have the ability to combine different mRNA sequences encoding for different proteins in a single mRNA investigational medicine.

Who needs a real 'virus' when you can create a computer version to justify infusing your operating system into the entire human race on the road to making living, breathing people into cyborgs? What is missed with the 'vaccines' is the *digital* connection between synthetic material and the body that I highlighted earlier with the study that hacked a computer with human DNA. On one level the body is digital, based on mathematical codes, and I'll have more about that in the next chapter. Those who ridiculously claim that mRNA 'vaccines' are not designed to change human genetics should explain the words of Dr Tal Zaks, chief medical officer at Moderna, in a 2017 TED talk. He said that over the last 30 years 'we've been living this phenomenal digital scientific revolution, and I'm here today to tell you, that we are actually *hacking the software of life*, and that it's changing the way we think about prevention and treatment of disease':

> In every cell there's this thing called messenger RNA, or mRNA for short, that transmits the critical information from the DNA in our genes to the protein, which is really the stuff we're all made out of. This is the critical information that determines what the cell will do. So we think about it as an operating system. So if you could change that, if you could introduce a line of code, or change a line of code, it turns out, that has profound implications for everything, from the flu to cancer.

Zaks should more accurately have said that this has profound implications for the human genetic code and the nature of DNA. Communications within the body go both ways and not only one. But, hey, no, the 'Covid vaccine' will not affect your genetics. Cult fact-checkers say so even though the man who helped to develop the mRNA technique says that it does. Zaks said in 2017:

> If you think about what it is we're trying to do. We've taken information and our understanding of that information and how that information is transmitted in a cell, and we've taken our understanding of medicine and how to make drugs, and we're fusing the two. We think of it as information therapy.

I have been writing for decades that the body is an information field communicating with itself and the wider world. This is why radiation which is information can change the information field of body and mind through phenomena like 5G and change their nature and function. ‘Information therapy’ means to change the body’s information field and change the way it operates. DNA is a receiver-transmitter of information and can be mutated by information like mRNA synthetic messaging. Technology to do this has been ready and waiting in the underground bases and other secret projects to be rolled out when the ‘Covid’ hoax was played. ‘Trials’ of such short and irrelevant duration were only for public consumption. When they say the ‘vaccine’ is ‘experimental’ that is not true. It may appear to be ‘experimental’ to those who don’t know what’s going on, but the trials have already been done to ensure the Cult gets the result it desires. Zaks said that it took decades to sequence the human genome, completed in 2003, but now they could do it in a week. By ‘they’ he means scientists operating in the public domain. In the secret projects they were sequencing the genome in a week long before even 2003.

Deluge of mRNA

Highly significantly the Moderna document says the guiding premise is that if using mRNA as a medicine works for one disease then it should work for many diseases. They were leveraging the flexibility afforded by their platform and the fundamental role mRNA plays in protein synthesis to pursue mRNA medicines for a broad spectrum of diseases. Moderna is confirming what I was saying through 2020 that multiple ‘vaccines’ were planned for ‘Covid’ (and later invented ‘variants’) and that previous vaccines would be converted to the mRNA system to infuse the body with massive amounts of genetically-manipulating synthetic material to secure a transformation to a synthetic-biological state. The ‘vaccines’ are designed to kill stunning numbers as part of the long-exposed Cult depopulation agenda and transform the rest. Given this is the goal you can appreciate why there is such hysterical demand for every human to be ‘vaccinated’ for an alleged ‘disease’ that has an estimated ‘infection’ to ‘death’ ratio of 0.23-0.15 percent. As I write children are being given the ‘vaccine’ in trials (their parents are a disgrace) and ever-younger people are being offered the vaccine for a ‘virus’ that even if you believe it exists has virtually zero chance of harming them. Horrific effects of the ‘trials’ on a 12-year-old girl were revealed by a family member to be serious brain and gastric problems that included a bowel obstruction and the inability to swallow liquids or solids. She was unable to eat or drink without throwing up, had extreme pain in her back, neck and abdomen, and was paralysed from the waist down which stopped her urinating unaided. When the girl was first taken to hospital doctors said it was all in her mind. She was signed up for the ‘trial’ by her parents for whom no words suffice. None of this ‘Covid vaccine’ insanity makes any sense unless you see what the ‘vaccine’ really is – a body-changer. Synthetic biology or ‘SynBio’ is a fast-emerging and expanding scientific discipline which includes everything from genetic and molecular engineering to electrical and computer engineering. Synthetic biology is defined in these ways:

- A multidisciplinary area of research that seeks to create new biological parts, devices, and systems, or to redesign systems that are already found in nature.
- The use of a mixture of physical engineering and genetic engineering to create new (and therefore synthetic) life forms.
- An emerging field of research that aims to combine the knowledge and methods of biology, engineering and related disciplines in the design of chemically-synthesized DNA to create organisms with novel or enhanced characteristics and traits (synthetic organisms including humans).

We now have synthetic blood, skin, organs and limbs being developed along with synthetic body parts produced by 3D printers. These are all elements of the synthetic human programme and this comment by Kurzweil’s co-founder of the Singularity University, Peter Diamandis, can be seen in a whole new light with the 'Covid' hoax and the sanctions against those that refuse the 'vaccine':

> Anybody who is going to be resisting the progress forward [to transhumanism] is going to be resisting evolution and, fundamentally, they will die out. It's not a matter of whether it's good or bad. It's going to happen.

'Resisting evolution'? What absolute bollocks. The arrogance of these people is without limit. His 'it's going to happen' mantra is another way of saying 'resistance is futile' to break the spirit of those pushing back and we must not fall for it. Getting this genetically-transforming 'vaccine' into everyone is crucial to the Cult plan for total control and the desperation to achieve that is clear for anyone to see. Vaccine passports are a major factor in this and they, too, are a form of resistance is futile. It's NOT. The paper funded by the Rockefeller Foundation for the 2013 'health conference' in China said:

> We will interact more with artificial intelligence. The use of robotics, bio-engineering to augment human functioning is already well underway and will advance. Re-engineering of humans into potentially separate and unequal forms through genetic engineering or mixed human-robots raises debates on ethics and equality.
>
> A new demography is projected to emerge after 2030 [that year again] of technologies (robotics, genetic engineering, nanotechnology) producing robots, engineered organisms, 'nanobots' and artificial intelligence (AI) that can self-replicate. Debates will grow on the implications of an impending reality of human designed life.

What is happening today is so long planned. The world army enforcing the will of the world government is intended to be a robot army, not a human one. Today's military and its technologically 'enhanced' troops, pilotless planes and driverless vehicles are just stepping stones to that end. Human soldiers are used as Cult fodder and its time they woke up to that and worked for the freedom of the population instead of their own destruction and their family's destruction – the same with the police. Join us and let's sort this out. The phenomenon of enforce my own destruction is widespread in the 'Covid' era with Woker 'luvvies' in the acting and entertainment industries supporting ‘Covid’ rules which have destroyed their profession and the same with those among the public who put signs on the doors of their businesses ‘closed due to Covid – stay safe’ when many will never reopen. It’s a form of masochism and most certainly insanity.

Transgender = transhumanism

When something explodes out of nowhere and is suddenly everywhere it is always the Cult agenda and so it is with the tidal wave of claims and demands that have infiltrated every aspect of society under the heading of ‘transgenderism’. The term ‘trans’ is so ‘in’ and this is the dictionary definition:

> A prefix meaning ‘across’, ’through’, occurring … in loanwords from Latin, used in particular for denoting movement or conveyance from place to place (transfer; transmit; transplant) or complete change (transform; transmute), or to form adjectives meaning ’crossing’, ‘on the other side of’, or ‘going beyond’ the place named (transmontane; transnational; trans-Siberian).

Transgender means to go beyond gender and transhuman means to go beyond human. Both are aspects of the Cult plan to transform the human body to a synthetic state with *no gender*. Human 2.0 is not designed to procreate and would be produced technologically with no need for parents. The new human would mean the end of parents and so men, and increasingly women, are being targeted for the deletion of their rights and status. Parental rights are disappearing at an ever-quickening speed for the same reason. The new human would have no need for men or women when there is no procreation and no gender. Perhaps the transgender movement that appears to be in a permanent state of frenzy might now contemplate on how it is being used. This was never about transgender rights which are only the interim excuse for confusing gender, particularly in the young, on the road to *fusing* gender. Transgender activism is not an end; it is a *means* to an end. We see again the technique of creative destruction in which you destroy the status quo to ‘build back better’ in the form that you want. The gender status quo had to be destroyed by persuading the Cult-created Woke mentality to believe that you can have 100 genders or more. A programme for 9 to 12 year olds produced by the Cult-owned BBC promoted the 100 genders narrative. The very idea may be the most monumental nonsense, but it is not what is true that counts, only what you can make people *believe* is true. Once the gender of 2 + 2 = 4 has been dismantled through indoctrination, intimidation and 2 + 2 = 5 then the new no-gender normal can take its place with Human 2.0. Aldous Huxley revealed the plan in his prophetic *Brave New World* in 1932:

> Natural reproduction has been done away with and children are created, decanted', and raised in 'hatcheries and conditioning centres'. From birth, people are genetically designed to fit into one of five castes, which are further split into 'Plus' and 'Minus' members and designed to fulfil predetermined positions within the social and economic strata of the World State.

How could Huxley know this in 1932? For the same reason George Orwell knew about the Big Brother state in 1948, Cult insiders I have quoted knew about it in 1969, and I have known about it since the early 1990s. If you are connected to the Cult or you work your balls off to uncover the plan you can predict the future. The process is simple. If there is a plan for the world and nothing intervenes to stop it then it will happen. Thus if you communicate the plan ahead of time you are perceived to have predicted the future, but you haven't. You have revealed the plan which without intervention will become the human future. The whole reason I have done what I have is to alert enough people to inspire an intervention and maybe at last that time has come with the Cult and its intentions now so obvious to anyone with a brain in working order.

The future is here

Technological wombs that Huxley described to replace parent procreation are already being developed and they are only the projects we know about in the public arena. Israeli scientists told *The Times of Israel* in March, 2021, that they have grown 250-cell embryos into mouse foetuses with fully formed organs using artificial wombs in a development they say could pave the way for gestating humans outside the womb. Professor Jacob Hanna of the Weizmann Institute of Science said:

> We took mouse embryos from the mother at day five of development, when they are just of 250 cells, and had them in the incubator from day five until day 11, by which point they had grown all their organs.
>
> By day 11 they make their own blood and have a beating heart, a fully developed brain. Anybody would look at them and say, 'this is clearly a mouse foetus with all the characteristics of a mouse.' It's gone from being a ball of cells to being an advanced foetus.

A special liquid is used to nourish embryo cells in a laboratory dish and they float on the liquid to duplicate the first stage of embryonic development. The incubator creates all the right conditions for its development, Hanna said. The liquid gives the embryo 'all the nutrients, hormones and sugars they need' along with a custom-made electronic incubator which controls gas concentration, pressure and temperature. The cutting-edge in the underground bases and other secret locations will be light years ahead of that, however, and this was reported by the London *Guardian* in 2017:

> We are approaching a biotechnological breakthrough. Ectogenesis, the invention of a complete external womb, could completely change the nature of human reproduction. In April this year, researchers at the Children's Hospital of Philadelphia announced their development of an artificial womb.

The article was headed 'Artificial wombs could soon be a reality. What will this mean for women?' What would it mean for children is an even bigger question. No mother to bond with only a machine in preparation for a life of soulless interaction and control in a world governed by machines (see the *Matrix* movies). Now observe the calculated manipulations of the 'Covid' hoax as human interaction and warmth has been curtailed by distancing, isolation and fear with people communicating via machines on a scale never seen before.

These are all dots in the same picture as are all the personal assistants, gadgets and children's toys through which kids and adults communicate with AI as if it is human. The AI 'voice' on Sat-Nav should be included. All these things are psychological preparation for the Cult endgame. Before you can make a physical connection with AI you have to make a psychological connection and that is what people are being conditioned to do with this ever gathering human-AI interaction. Movies and TV programmes depicting the transhuman, robot dystopia relate to a phenomenon known as 'pre-emptive programming' in which the world that is planned is portrayed everywhere in movies, TV and advertising. This is conditioning the conscious and subconscious mind to become familiar with the planned reality to dilute resistance when it happens for real. What would have been a shock such is the change is made less so. We have young children put on the road to transgender transition surgery with puberty blocking drugs at an age when they could never be able to make those life-changing decisions.

Rachel Levine, a professor of paediatrics and psychiatry who believes in treating children this way, became America's highest-ranked openly-transgender official when she was confirmed as US Assistant Secretary at the Department of Health and Human Services after being nominated by Joe Biden (the Cult). Activists and governments press for laws to deny parents a say in their children's transition process so the kids can be isolated and manipulated into agreeing to irreversible medical procedures. A Canadian father Robert Hoogland was denied bail by the Vancouver Supreme Court in 2021 and remained in jail for breaching a court order that he stay silent over his young teenage daughter, a minor, who was being offered life-changing hormone therapy without parental consent. At the age of 12 the girl's 'school counsellor' said she may be transgender, referred her to a doctor and told the school to treat her like a boy. This is another example of state-serving schools imposing ever more control over children's lives while parents have ever less.

Contemptible and extreme child abuse is happening all over the world as the Cult gender-fusion operation goes into warp-speed.

Why the war on men – and now women?

The question about what artificial wombs mean for women should rightly be asked. The answer can be seen in the deletion of women's rights involving sport, changing rooms, toilets and status in favour of people in male bodies claiming to identify as women. I can identify as a mountain climber, but it doesn't mean I can climb a mountain any more than a biological man can be a biological woman. To believe so is a triumph of belief over factual reality which is the very perceptual basis of everything Woke. Women's sport is being destroyed by allowing those with male bodies who say they identify as female to 'compete' with girls and women. Male body 'women' dominate 'women's' competition with their greater muscle mass, bone density, strength and speed. With that disadvantage sport for women loses all meaning. To put this in perspective nearly 300 American high school boys can run faster than the quickest woman sprinter in the world. Women are seeing their previously protected spaces invaded by male bodies simply because they claim to identify as women. That's all they need to do to access all women's spaces and activities under the Biden 'Equality Act' that destroys equality for women with the usual Orwellian Woke inversion. Male sex offenders have already committed rapes in women's prisons after claiming to identify as women to get them transferred. Does this not matter to the Woke 'equality' hypocrites? Not in the least. What matters to Cult manipulators and funders behind transgender activists is to advance gender fusion on the way to the no-gender 'human'. When you are seeking to impose transparent nonsense like this, or the 'Covid' hoax, the only way the nonsense can prevail is through censorship and intimidation of dissenters, deletion of factual information, and programming of the unquestioning, bewildered and naive. You don't have to scan the world for long to see that all these things are happening.

Many women's rights organisations have realised that rights and status which took such a long time to secure are being eroded and that it is systematic. Kara Dansky of the global Women's Human Rights Campaign said that Biden's transgender executive order immediately he took office, subsequent orders, and Equality Act legislation that followed 'seek to erase women and girls in the law as a category'. *Exactly*. I said during the long ago-started war on men (in which many women play a crucial part) that this was going to turn into a war on them. The Cult is phasing out *both* male and female genders. To get away with that they are brought into conflict so they are busy fighting each other while the Cult completes the job with no unity of response. Unity, people, *unity*. We need unity everywhere. Transgender is the only show in town as the big step towards the no-gender human. It's not about rights for transgender people and never has been. Woke political correctness is deleting words relating to genders to the same end. Wokers believe this is to be 'inclusive' when the opposite is true. They are deleting words describing gender because gender *itself* is being deleted by Human 2.0. Terms like 'man', 'woman', 'mother' and 'father' are being deleted in the universities and other institutions to be replaced by the *no*-gender, not trans-gender, 'individuals' and 'guardians'. Women's rights campaigner Maria Keffler of Partners for Ethical Care said: 'Children are being taught from kindergarten upward that some boys have a vagina, some girls have a penis, and that kids can be any gender they want to be.' Do we really believe that suddenly countries all over the world at the same time had the idea of having drag queens go into schools or read transgender stories to very young children in the local library? It's coldly-calculated confusion of gender on the way to the fusion of gender. Suzanne Vierling, a psychologist from Southern California, made another important point:

> Yesterday's slave woman who endured gynecological medical experiments is today's girl-child being butchered in a booming gender-transitioning sector. Ovaries removed, pushing her into menopause and osteoporosis, uncharted territory, and parents' rights and authority decimated.

The erosion of parental rights is a common theme in line with the Cult plans to erase the very concept of parents and 'ovaries removed, pushing her into menopause' means what? Those born female lose the ability to have children – another way to discontinue humanity as we know it.

Eliminating Human 1.0 (before our very eyes)

To pave the way for Human 2.0 you must phase out Human 1.0. This is happening through plummeting sperm counts and making women infertile through an onslaught of chemicals, radiation (including smartphones in pockets of men) and mRNA 'vaccines'. Common agriculture pesticides are also having a devastating impact on human fertility. I have been tracking collapsing sperm counts in the books for a long time and in 2021 came a book by fertility scientist and reproductive epidemiologist Shanna Swan, *Count Down: How Our Modern World Is Threatening Sperm Counts, Altering Male and Female Reproductive Development and Imperiling the Future of the Human Race*. She reports how the global fertility rate dropped by *half* between 1960 and 2016 with America's birth rate 16 percent below where it needs to be to sustain the population. Women are experiencing declining egg quality, more miscarriages, and more couples suffer from infertility. Other findings were an increase in erectile dysfunction, infant boys developing more genital abnormalities, male problems with conception, and plunging levels of the male hormone testosterone which would explain why so many men have lost their backbone and masculinity. This has been very evident during the 'Covid' hoax when women have been prominent among the Pushbackers and big strapping blokes have bowed their heads, covered their faces with a nappy and quietly submitted. Mind control expert Cathy O'Brien also points to how global education introduced the concept of 'we're all winners' in sport and classrooms: 'Competition was defused, and it in turn defused a sense of fighting back.' This is another version of the 'equity' doctrine in which you drive down rather than raise up. What a contrast in Cult-controlled China with its global ambitions where the government published plans in January, 2021, to ‘cultivate masculinity’ in boys from kindergarten through to high school in the face of a ‘masculinity crisis’. A government adviser said boys would be soon become ‘delicate, timid and effeminate’ unless action was taken. Don’t expect any similar policy in the targeted West. A 2006 study showed that a 65-year-old man in 2002 had testosterone levels *15 percent* lower than a 65-year-old man in 1987 while a 2020 study found a similar story with young adults and adolescents. Men are getting prescriptions for testosterone replacement therapy which causes an even greater drop in sperm count with up to 99 percent seeing sperm counts drop to zero during the treatment. More sperm is defective and malfunctioning with some having two heads or not pursuing an egg.

A class of *synthetic* chemicals known as phthalates are being blamed for the decline. These are found everywhere in plastics, shampoos, cosmetics, furniture, flame retardants, personal care products, pesticides, canned foods and even receipts. Why till receipts? Everyone touches them. Let no one delude themselves that all this is not systematic to advance the long-time agenda for human body transformation. Phthalates mimic hormones and disrupt the hormone balance causing testosterone to fall and genital birth defects in male infants. Animals and fish have been affected in the same way due to phthalates and other toxins in rivers. When fish turn gay or change sex through chemicals in rivers and streams it is a pointer to why there has been such an increase in gay people and the sexually confused. It doesn’t matter to me what sexuality people choose to be, but if it’s being affected by chemical pollution and consumption then we need to know. Does anyone really think that this is not connected to the transgender agenda, the war on men and the condemnation of male ‘toxic masculinity’? You watch this being followed by ‘toxic femininity’. It’s already happening. When breastfeeding becomes ‘chest-feeding’, pregnant women become pregnant people along with all the other Woke claptrap you know that the world is going insane and there’s a Cult scam in progress. Transgender activists are promoting the Cult agenda while Cult billionaires support and fund the insanity as they laugh themselves to sleep at the sheer stupidity for which humans must be infamous in galaxies far, far away.

‘Covid vaccines’ and female infertility

We can now see why the ‘vaccine’ has been connected to potential infertility in women. Dr Michael Yeadon, former Vice President and Chief Scientific Advisor at Pfizer, and Dr Wolfgang Wodarg in Germany, filed a petition with the European Medicines Agency in December, 2020, urging them to stop trials for the Pfizer/BioNTech shot and all other mRNA trials until further studies had been done. They were particularly concerned about possible effects on fertility with ‘vaccine’-produced antibodies attacking the protein Syncytin-1 which is responsible for developing the placenta. The result would be infertility ‘of indefinite duration’ in women who have the ‘vaccine’ with the placenta failing to form. Section 10.4.2 of the Pfizer/BioNTech trial protocol says that pregnant women or those who might become so should not have mRNA shots. Section 10.4 warns men taking mRNA shots to ‘be abstinent from heterosexual intercourse’ and not to donate sperm. The UK government said that it *did not know* if the mRNA procedure had an effect on fertility. *Did not know?* These people have to go to jail. UK government advice did not recommend at the start that pregnant women had the shot and said they should avoid pregnancy for at least two months after ‘vaccination’. The ‘advice’ was later updated to pregnant women should only have the ‘vaccine’ if the benefits outweighed the risks to mother and foetus. What the hell is that supposed to mean? Then ‘spontaneous abortions’ began to appear and rapidly increase on the adverse reaction reporting schemes which include only a fraction of adverse reactions. Thousands and ever-growing numbers of ‘vaccinated’ women are describing changes to their menstrual cycle with heavier blood flow, irregular periods and menstruating again after going through the menopause – all links to reproduction effects. Women are passing blood clots and the lining of their uterus while men report erectile dysfunction and blood effects. Most significantly of all *un*vaccinated women began to report similar menstrual changes after interaction with *'vaccinated'* people and men and children were also affected with bleeding noses, blood clots and other conditions. 'Shedding' is when vaccinated people can emit the content of a vaccine to affect the unvaccinated, but this is different. 'Vaccinated' people were not shedding a 'live virus' allegedly in 'vaccines' as before because the fake 'Covid vaccines' involve synthetic material and other toxicity. Doctors exposing what is happening prefer the term 'transmission' to shedding. Somehow those that have had the shots are transmitting effects to those that haven't. Dr Carrie Madej said the nano-content of the 'vaccines' can 'act like an antenna' to others around them which fits perfectly with my own conclusions. This 'vaccine' transmission phenomenon was becoming known as the book went into production and I deal with this further in the Postscript.

Vaccine effects on sterility are well known. The World Health Organization was accused in 2014 of sterilising millions of women in Kenya with the evidence confirmed by the content of the vaccines involved. The same WHO behind the 'Covid' hoax admitted its involvement for more than ten years with the vaccine programme. Other countries made similar claims. Charges were lodged by Tanzania, Nicaragua, Mexico, and the Philippines. The Gardasil vaccine claimed to protect against a genital 'virus' known as HPV has also been linked to infertility. Big Pharma and the WHO (same thing) are criminal and satanic entities. Then there's the Bill Gates Foundation which is connected through funding and shared interests with 20 pharmaceutical giants and laboratories. He stands accused of directing the policy of United Nations Children's Fund (UNICEF), vaccine alliance GAVI, and other groupings, to advance the vaccine agenda and silence opposition at great cost to women and children. At the same time Gates wants to reduce the global population. Coincidence?

Great Reset = Smart Grid = new human

The Cult agenda I have been exposing for 30 years is now being openly promoted by Cult assets like Gates and Klaus Schwab of the World Economic Forum under code-terms like the 'Great Reset', 'Build Back Better' and 'a rare but narrow window of opportunity to reflect, reimagine, and reset our world'. What provided this 'rare but narrow window of opportunity'? The 'Covid' hoax did. Who created that? *They* did. My books from not that long ago warned about the planned 'Internet of Things' (IoT) and its implications for human freedom. This was the plan to connect all technology to the Internet and artificial intelligence and today we are way down that road with an estimated 36 billion devices connected to the World Wide Web and that figure is projected to be 76 billion by 2025. I further warned that the Cult planned to go beyond that to the Internet of *Everything* when the human brain was connected via AI to the Internet and Kurzweil's 'cloud'. Now we have Cult operatives like Schwab calling for precisely that under the term 'Internet of Bodies', a fusion of the physical, digital and biological into one centrally-controlled Smart Grid system which the Cult refers to as the 'Fourth Industrial Revolution'. They talk about the 'biological', but they really mean the synthetic-biological which is required to fully integrate the human body and brain into the Smart Grid and artificial intelligence planned to replace the human mind. We have everything being synthetically manipulated including the natural world through GMO and smart dust, the food we eat and the human body itself with synthetic 'vaccines'. I said in *The Answer* that we would see the Cult push for synthetic meat to replace animals and in February, 2021, the so predictable psychopath Bill Gates called for the introduction of synthetic meat to save us all from 'climate change'. The climate hoax just keeps on giving like the 'Covid' hoax. The war on meat by vegan activists is a carbon (oops, sorry) copy of the manipulation of transgender activists. They have no idea (except their inner core) that they are being used to promote and impose the agenda of the Cult or that they are only the *vehicle* and not the *reason*. This is not to say those who choose not to eat meat shouldn't be respected and supported in that right, but there are ulterior motives for those in power. A *Forbes* article in December, 2019, highlighted the plan so beloved of Schwab and the Cult under the heading: 'What Is The Internet of Bodies? And How Is It Changing Our World?' The article said the human body is the latest data platform (remember 'our vaccine is an operating system'). *Forbes* described the plan very accurately and the words could have come straight out of my books from long before:

> The Internet of Bodies (IoB) is an extension of the IoT and basically connects the human body to a network through devices that are ingested, implanted, or connected to the body in some way. Once connected, data can be exchanged, and the body and device can be remotely monitored and controlled.

They were really describing a human hive mind with human perception centrally-dictated via an AI connection as well as allowing people to be 'remotely monitored and controlled'. Everything from a fridge to a human mind could be directed from a central point by these insane psychopaths and 'Covid vaccines' are crucial to this. Forbes explained the process I mentioned earlier of holdable and wearable technology followed by implantable. The article said there were three generations of the Internet of Bodies that include:

- Body external: These are wearable devices such as Apple Watches or Fitbits that can monitor our health.

- Body internal: These include pacemakers, cochlear implants, and digital pills that go inside our bodies to monitor or control various aspects of health.

- Body embedded: The third generation of the Internet of Bodies is embedded technology where technology and the human body are melded together and have a real-time connection to a remote machine.

Forbes noted the development of the Brain Computer Interface (BCI) which merges the brain with an external device for monitoring and controlling in real-time. 'The ultimate goal is to help restore function to individuals with disabilities by using brain signals rather than conventional neuromuscular pathways.' Oh, do fuck off. The goal of brain interface technology is controlling human thought and emotion from the central point in a hive mind serving its masters wishes. Many people are now agreeing to be chipped to open doors without a key. You can recognise them because they'll be wearing a mask, social distancing and lining up for the 'vaccine'. The Cult plans a Great Reset money system after they have completed the demolition of the global economy in which 'money' will be exchanged through communication with body operating systems. Rand Corporation, a Cult-owned think tank, said of the Internet of Bodies or IoB:

> Internet of Bodies technologies fall under the broader IoT umbrella. But as the name suggests, IoB devices introduce an even more intimate interplay between humans and gadgets. IoB devices monitor the human body, collect health metrics and other personal information, and transmit those data over the Internet. Many devices, such as fitness trackers, are already in use … IoB devices … and those in development can track, record, and store users' whereabouts, bodily functions, and what they see, hear, and even think.

Schwab's World Economic Forum, a long-winded way of saying 'fascism' or 'the Cult', has gone full-on with the Internet of Bodies in the 'Covid' era. 'We're entering the era of the Internet of Bodies', it declared, 'collecting our physical data via a range of devices that can be implanted, swallowed or worn'. The result would be a huge amount of health-related data that could improve human wellbeing around the world, and prove crucial in fighting the 'Covid-19 pandemic'. Does anyone think these clowns care about 'human wellbeing' after the death and devastation their pandemic hoax has purposely caused? Schwab and co say we should move forward with the Internet of Bodies because 'Keeping track of symptoms could help us stop the spread of infection, and quickly detect new cases'. How wonderful, but keeping track' is all they are really bothered about. Researchers were investigating if data gathered from smartwatches and similar devices could be used as viral infection alerts by tracking the user's heart rate and breathing. Schwab said in his 2018 book *Shaping the Future of the Fourth Industrial Revolution*:

> The lines between technologies and beings are becoming blurred and not just by the ability to create lifelike robots or synthetics. Instead it is about the ability of new technologies to literally become part of us. Technologies already influence how we understand ourselves, how we think about each other, and how we determine our realities. As the technologies … give us deeper access to parts of ourselves, we may begin to integrate digital technologies into our bodies.

You can see what the game is. Twenty-four hour control and people – if you could still call them that – would never know when something would go ping and take them out of circulation. It's the most obvious rush to a global fascist dictatorship and the complete submission of humanity and yet still so many are locked away in their Cult-induced perceptual coma and can't see it.

Smart Grid control centres

The human body is being transformed by the 'vaccines' and in other ways into a synthetic cyborg that can be attached to the global Smart Grid which would be controlled from a central point and other sub-locations of Grid manipulation. Where are these planned to be? Well, China for a start which is one of the Cult's biggest centres of operation. The technological control system and technocratic rule was incubated here to be unleashed across the world after the 'Covid' hoax came out of China in 2020. Another Smart Grid location that will surprise people new to this is Israel. I have exposed in *The Trigger* how Sabbatian technocrats, intelligence and military operatives were behind the horrors of 9/11 and not '19 Arab hijackers' who somehow manifested the ability to pilot big passenger airliners when instructors at puddle-jumping flying schools described some of them as a joke. The 9/11 attacks were made possible through control of civilian and military air computer systems and those of the White House, Pentagon and connected agencies. See *The Trigger* – it will blow your mind. The controlling and coordinating force were the Sabbatian networks in Israel and the United States which by then had infiltrated the entire US government, military and intelligence system. The real name of the American Deep State is 'Sabbatian State'. Israel is a tiny country of only nine million people, but it is one of the global centres of cyber operations and fast catching Silicon Valley in importance to the Cult. Israel is known as the 'start-up nation' for all the cyber companies spawned there with the Sabbatian specialisation of 'cyber security' that I mentioned earlier which gives those companies access to computer systems of their clients in real time through 'backdoors' written into the coding when security software is downloaded. The Sabbatian centre of cyber operations outside Silicon Valley is the Israeli military Cyber Intelligence Unit, the biggest infrastructure project in Israel's history, headquartered in the desert-city of Beersheba and involving some 20,000 'cyber soldiers'. Here are located a literal army of Internet trolls scanning social media, forums and comment lists for anyone challenging the Cult agenda. The UK military has something similar with its 77th Brigade and associated operations. The Beersheba complex includes research and development centres for other Cult operations such as Intel, Microsoft, IBM, Google, Apple, Hewlett-Packard, Cisco Systems, Facebook and Motorola. Techcrunch.com ran an article about the Beersheba global Internet technology centre headlined 'Israel's desert city of Beersheba is turning into a cybertech oasis':

> The military's massive relocation of its prestigious technology units, the presence of multinational and local companies, a close proximity to Ben Gurion University and generous government subsidies are turning Beersheba into a major global cybertech hub. Beersheba has all of the ingredients of a vibrant security technology ecosystem, including Ben Gurion University with its graduate program in cybersecurity and Cyber Security Research Center, and the presence of companies such as EMC, Deutsche Telekom, PayPal, Oracle, IBM, and Lockheed Martin. It's also the future home of the INCB (Israeli National Cyber Bureau); offers a special income tax incentive for cyber security companies, and was the site for the relocation of the army's intelligence corps units.

Sabbatians have taken over the cyber world through the following process: They scan the schools for likely cyber talent and develop them at Ben Gurion University and their period of conscription in the Israeli Defense Forces when they are stationed at the Beersheba complex. When the cyber talented officially leave the army they are funded to start cyber companies with technology developed by themselves or given to them by the state. Much of this is stolen through backdoors of computer systems around the world with America top of the list. Others are sent off to Silicon Valley to start companies or join the major ones and so we have many major positions filled by apparently 'Jewish' but really Sabbatian operatives. Google, YouTube and Facebook are all run by 'Jewish' CEOs while Twitter is all but run by ultra-Zionist hedge-fund shark Paul Singer. At the centre of the Sabbatian global cyber web is the Israeli army's Unit 8200 which specialises in hacking into computer systems of other countries, inserting viruses, gathering information, instigating malfunction, and even taking control of them from a distance. A long list of Sabbatians involved with 9/11, Silicon Valley and Israeli cyber security companies are operatives of Unit 8200. This is not about Israel. It's about the Cult. Israel is planned to be a Smart Grid hub as with China and what is happening at Beersheba is not for the benefit of Jewish people who are treated disgustingly by the Sabbatian elite that control the country. A glance at the Nuremberg Codes will tell you that.

The story is much bigger than 'Covid', important as that is to where we are being taken. Now, though, it's time to really strap in. There's more … much more …

CHAPTER ELEVEN

Who controls the Cult?

Awake, arise or be forever fall'n
John Milton, *Paradise Lost*

I have exposed this far the level of the Cult conspiracy that operates in the world of the seen and within the global secret society and satanic network which operates in the shadows one step back from the seen. The story, however, goes much deeper than that.

The 'Covid' hoax is major part of the Cult agenda, but only part, and to grasp the biggest picture we have to expand our attention beyond the realm of human sight and into the infinity of possibility that we cannot see. It is from here, ultimately, that humanity is being manipulated into a state of total control by the force which dictates the actions of the Cult. How much of reality can we see? Next to damn all is the answer. We may appear to see all there is to see in the 'space' our eyes survey and observe, but little could be further from the truth. The human 'world' is only a tiny band of frequency that the body's visual and perceptual systems can decode into *perception* of a 'world'. According to mainstream science the electromagnetic spectrum is 0.005 percent of what exists in the Universe (Fig 10). The maximum estimate I have seen is 0.5 percent and either way it's miniscule. I say it is far, far, smaller even than 0.005 percent when you compare reality we see with the totality of reality that we don't. Now get this if you are new to such information: Visible light, the only band of frequency that we can see, is a *fraction* of the 0.005

percent (Fig 11 overleaf). Take this further and realise that our universe is one of infinite universes and that universes are only a fragment of overall reality – *infinite* reality. Then compare that with the almost infinitesimal frequency band of visible light or human sight. You see that humans are as near blind as it is possible to be without actually being so. Artist and filmmaker, Sergio Toporek, said:

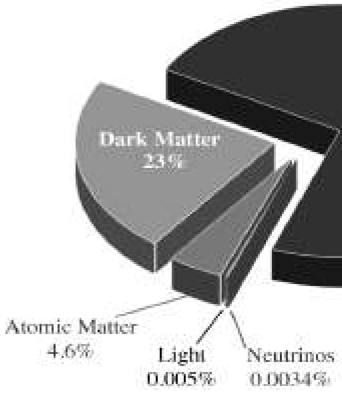


Figure 10: Humans can perceive such a tiny band of visual reality it's laughable.

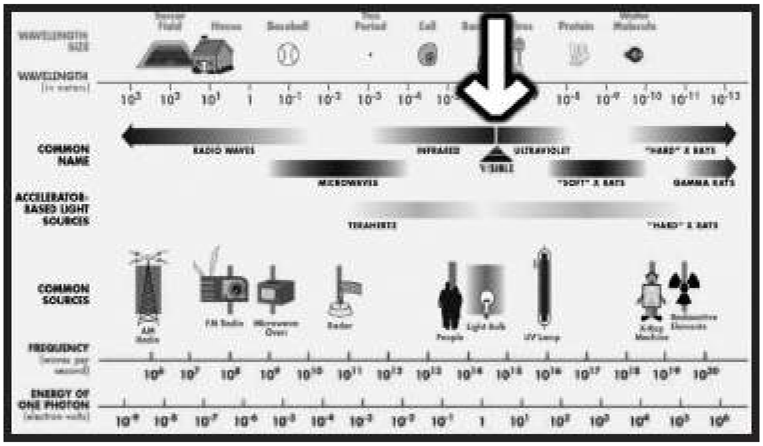

WAVELENGTH SIZE
Baseball
Cell
Protein
WAVELENGTH
COMMON NAME
RADIO WAVES
INFRARED
VISIBLE
ULTRAVIOLET
"HARD" X RAYS
MICROWAVES
"SOFT" X RAYS
GAMMA RAYS
ACCELERATOR-BASED LIGHT SOURCES
TERAHERTZ
"HARD" X RAYS
COMMON SOURCES
AM Radio
FM Radio
Microwave Oven
Radar
People
Light Bulb
UV Lamp
X-Ray Machine
Radioactive Elements
FREQUENCY
ENERGY OF ONE PHOTON

Figure 11: We can see a smear of the 0.005 percent electromagnetic spectrum, but we still know it all. Yep, makes sense.

Consider that you can see less than 1% of the electromagnetic spectrum and hear less than 1% of the acoustic spectrum. 90% of the cells in your body carry their own microbial DNA and are not 'you'. The atoms in your body are 99.9999999999999999% empty space and none of them are the ones you were born with ... Human beings have 46 chromosomes, two less than a potato.

The existence of the rainbow depends on the conical photoreceptors in your eyes; to animals without cones, the rainbow does not exist. So you don't just look at a rainbow, you create it. This is pretty amazing, especially considering that all the beautiful colours you see represent less than 1% of the electromagnetic spectrum.

Suddenly the 'world' of humans looks a very different place. Take into account, too, that Planet Earth when compared with the projected size of this single universe is the equivalent of a billionth of a pinhead. Imagine the ratio that would be when compared to infinite reality. To think that Christianity once insisted that Earth and humanity were the centre of everything. This background is vital if we are going to appreciate the nature of 'human' and how we can be manipulated by an unseen force. To human visual reality virtually *everything* is unseen and yet the prevailing perception within the institutions and so much of the public is that if we can't see it, touch it, hear it, taste it and smell it then it cannot exist. Such perception is indoctrinated and encouraged by the Cult and its agents because it isolates believers in the strictly limited, village-idiot, realm of the five senses where perceptions can be firewalled and information controlled. Most of those perpetuating the 'this-world-is-all-there-is' insanity are themselves indoctrinated into believing the same delusion. While major players and influencers know that official reality is laughable most of those in science, academia and medicine really believe the nonsense they peddle and teach succeeding generations. Those who challenge the orthodoxy are dismissed as nutters and freaks to protect the manufactured illusion from exposure. Observe the dynamic of the 'Covid' hoax and you will see how that takes the same form. The inner-circle psychopaths knows it's a gigantic scam, but almost the entirety of those imposing their fascist rules believe that 'Covid' is all that they're told it is.

Stolen identity

Ask people who they are and they will give you their name, place of birth, location, job, family background and life story. Yet that is not who they are – it is what they are *experiencing*. The difference is *absolutely crucial*. The true 'I', the eternal, infinite 'I', is consciousness, a state of being aware. Forget 'form'. That is a vehicle for a brief experience. Consciousness does not come *from* the brain, but *through* the brain and even that is more symbolic than literal. We are awareness, pure awareness, and this is what withdraws from the body at what we call 'death' to continue our eternal beingness, *isness*, in other realms of reality within the limitlessness of infinity or the Biblical 'many mansions in my father's house'. Labels of a human life, man, woman, transgender, black, white, brown, nationality, circumstances and income are not who we are. They are what we are – awareness – is *experiencing* in a brief connection with a band of frequency we call 'human'. The labels are not the self; they are, to use the title of one of my books, a *Phantom* Self. I am not David Icke born in Leicester, England, on April 29th, 1952. I am the consciousness *having that experience*. The Cult and its non-human masters seek to convince us through the institutions of 'education', science, medicine, media and government that what we are *experiencing* is who we *are*. It's so easy to control and direct perception locked away in the bewildered illusions of the five senses with no expanded radar. Try, by contrast, doing the same with a humanity aware of its true self and its true power to consciously create its reality and experience. How is it possible to do this? We do it all day every day. If you perceive yourself as 'little me' with no power to impact upon your life and the world then your life experience will reflect that. You will hand the power you don't think you have to authority in all its forms which will use it to control your experience. This, in turn, will appear to confirm your perception of 'little me' in a self-fulfilling feedback loop. But that is what 'little me' really is – a *perception*. We are all 'big-me', infinite me, and the Cult has to make us forget that if its will is to prevail. We are therefore manipulated and pressured into self-identifying with human labels and not the consciousness/awareness *experiencing* those human labels.

The phenomenon of identity politics is a Cult-instigated manipulation technique to sub-divide previous labels into even smaller ones. A United States university employs this list of letters to describe student identity: LGBTTQQFAGPBDSM or lesbian, gay, bisexual, transgender, transsexual, queer, questioning, flexual, asexual, gender-fuck, polyamorous, bondage/discipline, dominance/submission and sadism/masochism. I'm sure other lists are even longer by now as people feel the need to self-identity the 'I' with the minutiae of race and sexual preference. Wokers programmed by the Cult for generations believe this is about 'inclusivity' when it's really the Cult locking them away into smaller and smaller versions of Phantom Self while firewalling them from the influence of their true self, the infinite, eternal 'I'. You may notice that my philosophy which contends that we are all unique points of attention/awareness within the same infinite whole or Oneness is the ultimate non-racism. The very sense of Oneness makes the judgement of people by their body-type, colour or sexuality utterly ridiculous and confirms that racism has no understanding of reality (including anti-white racism). Yet despite my perception of life Cult agents and fast-asleep Wokers label me racist to discredit my information while they are themselves phenomenally racist and sexist. All they see is race and sexuality and they judge people as good or bad, demons or untouchables, by their race and sexuality. All they see is *Phantom Self* and perceive themselves in terms of Phantom Self. They are pawns and puppets of the Cult agenda to focus attention and self-identity in the five senses and play those identities against each other to divide and rule. Columbia University has introduced segregated graduations in another version of social distancing designed to drive people apart and teach them that different racial and cultural groups have nothing in common with each other. The last thing the Cult wants is unity. Again the pump-primers of this will be Cult operatives in the knowledge of what they are doing, but the rest are just the Phantom Self blind leading the Phantom Self blind. We *do* have something in common – we are all *the same consciousness* having different temporary experiences.

What is this 'human'?

Yes, what *is* 'human'? That is what we are supposed to be, right? I mean 'human'? True, but 'human' is the experience not the 'I'. Break it down to basics and 'human' is the way that information is processed. If we are to experience and interact with this band of frequency we call the 'world' we must have a vehicle that operates within that band of frequency. Our consciousness in its prime form cannot do that; it is way beyond the frequency of the human realm. My consciousness or awareness could not tap these keys and pick up the cup in front of me in the same way that radio station A cannot interact with radio station B when they are on different frequencies. The human body is the means through which we have that interaction. I have long described the body as a biological computer which processes information in a way that allows consciousness to experience this reality. The body is a receiver, transmitter and processor of information in a particular way that we call human. We visually perceive only the world of the five senses in a wakened state – that is the limit of the body's visual decoding system. In truth it's not even visual in the way we experience 'visual reality' as I will come to in a moment. We are 'human' because the body processes the information sources of human into a reality and behaviour system that we *perceive* as human. Why does an elephant act like an elephant and not like a human or a duck? The elephant's biological computer is a different information field and processes information according to that program into a visual and behaviour type we call an elephant. The same applies to everything in our reality. These body information fields are perpetuated through procreation (like making a copy of a software program). The Cult wants to break that cycle and intervene technologically to transform the human information field into one that will change what we call humanity. If it can change the human information field it will change the way that field processes information and change humanity both 'physically' and psychologically. Hence the *messenger* (information) RNA 'vaccines' and so much more that is targeting human genetics by changing the body's information – *messaging* – construct through food, drink, radiation, toxicity and other means.

Reality that we experience is nothing like reality as it really is in the same way that the reality people experience in virtual reality games is not the reality they are really living in. The game is only a decoded source of information that appears to be a reality. Our world is also an information construct – a *simulation* (more later). In its base form our reality is a wavefield of information much the same in theme as Wi-Fi. The five senses decode wavefield information into electrical information which they communicate to the brain to decode into holographic (illusory 'physical') information. Different parts of the brain specialise in decoding different senses and the information is fused into a reality that appears to be outside of us but is really inside the brain and the genetic structure in general (Fig 12 overleaf). DNA is a receiver-transmitter of information and a vital part of this decoding process and the body's connection to other realities. Change DNA and you change the way we decode and connect with reality – see 'Covid vaccines'. Think of computers decoding Wi-Fi. You have information encoded in a radiation field and the computer decodes that information into a very different form on the screen. You can't see the Wi-Fi until its information is made manifest on the screen and the information on the screen is inside the computer and not outside. I have just described how we decode the 'human world'. All five senses decode the waveform 'Wi-Fi' field into electrical signals and the brain (computer) constructs reality inside the brain and not outside – 'You don't just look at a rainbow, you create it'. Sound is a simple example. We don't hear sound until the brain decodes it. Waveform sound waves are picked up by the hearing sense and communicated to the brain in an electrical form to be decoded into the sounds that we hear. Everything we hear is inside the brain along with everything we see, feel, smell and taste. Words and language are waveform fields generated by our vocal chords which pass through this process until they are decoded by the brain into words that we hear. Different languages are different frequency fields or sound waves generated by vocal chords. Late British philosopher Alan Watts said:

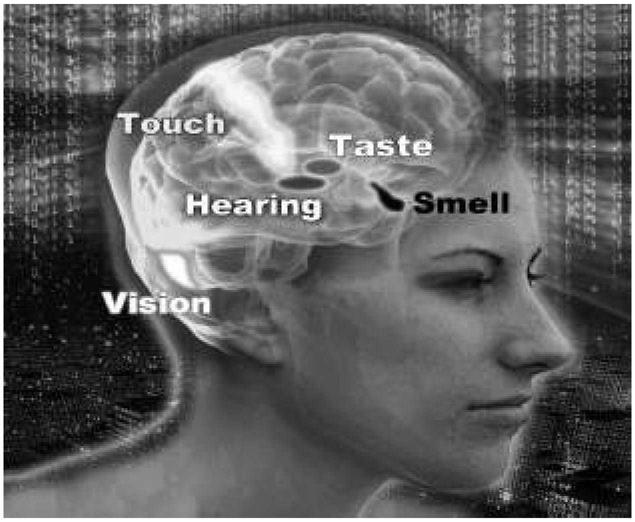


Figure 12: The brain receives information from the five senses and constructs from that our perceived reality.

> [Without the brain] the world is devoid of light, heat, weight, solidity, motion, space, time or any other imaginable feature. All these phenomena are interactions, or transactions, of vibrations with a certain arrangement of neurons.

That's exactly what they are and scientist Robert Lanza describes in his book, *Biocentrism*, how we decode electromagnetic waves and energy into visual and 'physical' experience. He uses the example of a flame emitting photons, electromagnetic energy, each pulsing electrically and magnetically:

> … these … invisible electromagnetic waves strike a human retina, and if (and only if) the waves happen to measure between 400 and 700 nano meters in length from crest to crest, then their energy is just right to deliver a stimulus to the 8 million cone-shaped cells in the retina.
>
> Each in turn send an electrical pulse to a neighbour neuron, and on up the line this goes, at 250 mph, until it reaches the … occipital lobe of the brain, in the back of the head. There, a cascading complex of neurons fire from the incoming stimuli, and we subjectively perceive this experience as a yellow brightness occurring in a place we have been conditioned to call the 'external world'.

You hear what you decode

If a tree falls or a building collapses they make no noise unless someone is there to decode the energetic waves generated by the disturbance into what we call sound. Does a falling tree make a noise? Only if you hear it – *decode* it. Everything in our reality is a frequency field of information operating within the overall 'Wi-Fi' field that I call The Field. A vibrational disturbance is generated in The Field by the fields of the falling tree or building. These disturbance waves are what we decode into the sound of them falling. If no one is there to do that then neither will make any noise. Reality is created by the observer – *decoder* – and the *perceptions* of the observer affect the decoding process. For this reason different people – different *perceptions* – will perceive the same reality or situation in a different way. What one may perceive as a nightmare another will see as an opportunity. The question of why the Cult is so focused on controlling human perception now answers itself. All experienced reality is the act of decoding and we don't experience Wi-Fi until it is decoded on the computer screen. The sight and sound of an Internet video is encoded in the Wi-Fi all around us, but we don't see or hear it until the computer decodes that information. Taste, smell and touch are all phenomena of the brain as a result of the same process. We don't taste, smell or feel anything except in the brain and there are pain relief techniques that seek to block the signal from the site of discomfort to the brain because if the brain doesn't decode that signal we don't feel pain. Pain is in the brain and only appears to be at the point of impact thanks to the feedback loop between them. We don't see anything until electrical information from the sight senses is decoded in an area at the back of the brain. If that area is damaged we can go blind when our eyes are perfectly okay. So why do we go blind if we damage an eye? We damage the information processing between the waveform visual information and the visual decoding area of the brain. If information doesn't reach the brain in a form it can decode then we can't see the visual reality that it represents. What's more the brain is decoding only a fraction of the information it receives and the rest is absorbed by the sub-conscious mind. This explanation is from the science magazine, *Wonderpedia*:

> Every second, 11 million sensations crackle along these [brain] pathways … The brain is confronted with an alarming array of images, sounds and smells which it rigorously filters down until it is left with a manageable list of around 40. Thus 40 sensations per second make up what we perceive as reality.

The 'world' is not what people are told to believe that is it and the inner circles of the Cult *know that*.

Illusory 'physical' reality

We can only see a smear of 0.005 percent of the Universe which is only one of a vast array of universes – 'mansions' – within infinite reality. Even then the brain decodes only 40 pieces of information ('sensations') from a potential *11 million* that we receive every second. Two points strike you from this immediately: The sheer breathtaking stupidity of believing we know anything so rigidly that there's nothing more to know; and the potential for these processes to be manipulated by a malevolent force to control the reality of the population. One thing I can say for sure with no risk of contradiction is that when you can perceive an almost indescribable fraction of infinite reality there is always more to know as in tidal waves of it. Ancient Greek philosopher Socrates was so right when he said that wisdom is to know how little we know. How obviously true that is when you think that we are experiencing a physical world of solidity that is neither physical nor solid and a world of apartness when everything is connected. Cult-controlled 'science' dismisses the so-called 'paranormal' and all phenomena related to that when the 'para'-normal is perfectly normal and explains the alleged 'great mysteries' which dumbfound scientific minds. There is a reason for this. A 'scientific mind' in terms of the mainstream is a material mind, a five-sense mind imprisoned in see it, touch it, hear it, smell it and taste it. Phenomena and happenings that can't be explained that way leave the 'scientific mind' bewildered and the rule is that if they can't account for why something is happening then it can't, by definition, be happening. I beg to differ. Telepathy is thought waves passing through The Field (think wave disturbance again) to be decoded by someone able to connect with that wavelength (information). For example: You can pick up the thought waves of a friend at any distance and at the very least that will bring them to mind. A few minutes later the friend calls you. 'My god', you say, 'that's incredible – I was just thinking of you.' Ah, but *they* were thinking of *you* before they made the call and that's what you decoded. Native peoples not entrapped in five-sense reality do this so well it became known as the 'bush telegraph'. Those known as psychics and mediums (genuine ones) are doing the same only across dimensions of reality. 'Mind over matter' comes from the fact that matter and mind are the *same*. The state of one influences the state of the other. Indeed one *and* the other are illusions. They are aspects of the same field. Paranormal phenomena are all explainable so why are they still considered 'mysteries' or not happening? Once you go down this road of understanding you begin to expand awareness beyond the five senses and that's the nightmare for the Cult.

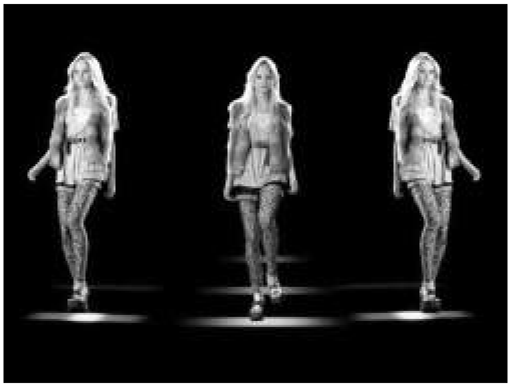


Figure 13: Holograms are not solid, but the best ones appear to be.

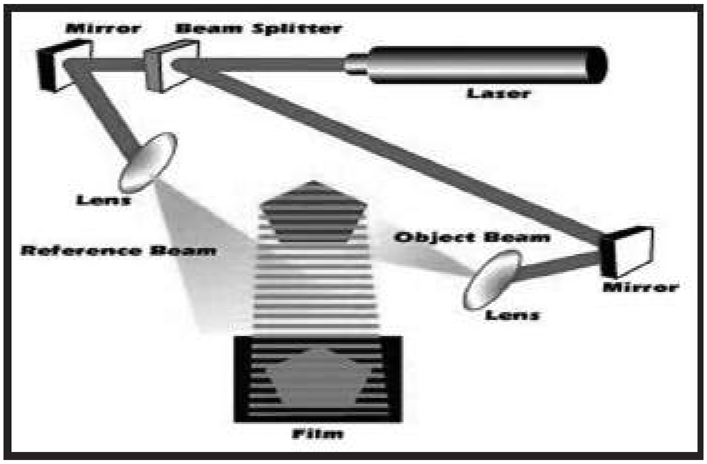


Figure 14: How holograms are created by capturing a waveform version of the subject image.

Holographic 'solidity'

Our reality is not solid, it is holographic. We are now well aware of holograms which are widely used today. Two-dimensional information is decoded into a three-dimensional reality that is not solid although can very much appear to be (Fig 13). Holograms are created with a laser divided into two parts. One goes directly onto a holographic photographic print ('reference beam') and the other takes a waveform image of the subject ('working beam') before being directed onto the print where it 'collides' with the other half of the laser (Fig 14). This creates a *waveform* interference pattern which contains the wavefield information of whatever is being photographed (Fig 15 overleaf). The process can be likened to dropping pebbles in a pond. Waves generated by each one spread out across the water to collide with the others and create a wave representation of where the stones fell and at what speed, weight and distance. A waveform interference pattern of a hologram is akin to the waveform information in The Field which the five senses decode into electrical signals to be decoded by the brain into a holographic illusory 'physical' reality. In the same way when a laser (think human attention) is directed at the waveform interference pattern a three-dimensional version of the subject is projected into apparently 'solid' reality (Fig 16). An amazing trait of holograms reveals more 'paranormal mysteries'. Information of the *whole* hologram is encoded in waveform in every part of the interference pattern by the way they are created. This means that every *part* of a hologram is a smaller version of the whole. Cut the interference wave-pattern into four and you won't get four parts of the image. You get quarter-sized versions of the *whole* image. The body is a hologram and the same applies. Here we have the basis of acupuncture, reflexology and other forms of healing which identify representations of the whole body in all of the parts, hands, feet, ears, everywhere. Skilled palm readers can do what they do because the information of whole body is encoded in the hand. The concept of as above, so below, comes from this.

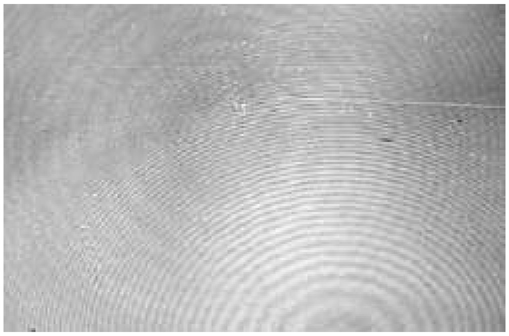


Figure 15: A waveform interference pattern that holds the information that transforms into a hologram.

She isn't really standing there
WILLIAM VIA HOLOGRAM
PERFORMER & OBAMA SUPPORTER
LIVE

Figure 16: Holographic people including 'Elvis' holographically inserted to sing a duet with Celine Dion.

The question will be asked of why, if solidity is illusory, we can't just walk through walls and each other. The resistance is not solid against solid; it is electromagnetic field against electromagnetic field and we decode this into the *experience* of solid against solid. We should also not underestimate the power of belief to dictate reality. What you believe is impossible *will be*. Your belief impacts on your decoding processes and they won't decode what you think is impossible. What we believe we perceive and what we perceive we experience. 'Can't dos' and 'impossibles' are like a firewall in a computer system that won't put on the screen what the firewall blocks. How vital that is to understanding how human experience has been hijacked. I explain in *The Answer*, *Everything You Need To Know But Have Never Been Told* and other books a long list of 'mysteries' and 'paranormal' phenomena that are not mysterious and perfectly normal once you realise what reality is and how it works. 'Ghosts' can be seen to pass through 'solid' walls because the walls are not solid and the ghost is a discarnate entity operating on a frequency so different to that of the wall that it's like two radio stations sharing the same space while never interfering with each other. I have seen ghosts do this myself. The apartness of people and objects is also an illusion. Everything is connected by the Field like all sea life is connected by the sea. It's just that within the limits of our visual reality we only 'see' holographic information and not the field of information that connects everything and from which the holographic world is made manifest. If you can only see holographic 'objects' and not the field that connects them they will appear to you as unconnected to each other in the same way that we see the computer while not seeing the Wi-Fi.

What you don't know *can* hurt you

Okay, we return to those 'two worlds' of human society and the Cult with its global network of interconnecting secret societies and satanic groups which manipulate through governments, corporations, media, religions, etc. The fundamental difference between them is *knowledge*. The idea has been to keep humanity ignorant of the plan for its total enslavement underpinned by a crucial ignorance of reality – who we are and where we are – and how we interact with it. 'Human' should be the interaction between our expanded eternal consciousness and the five-sense body experience. We are meant to be *in* this world in terms of the five senses but not *of* this world in relation to our greater consciousness and perspective. In that state we experience the small picture of the five senses within the wider context of the big picture of awareness beyond the five senses. Put another way the five senses see the dots and expanded awareness connects them into pictures and patterns that give context to the apparently random and unconnected. Without the context of expanded awareness the five senses see only apartness and randomness with apparently no meaning. The Cult and its other-dimensional controllers seek to intervene in the frequency realm where five-sense reality is supposed to connect with expanded reality and to keep the two apart (more on this in the final chapter). When that happens five-sense mental and emotional processes are no longer influenced by expanded awareness, or the True 'I', and instead are driven by the isolated perceptions of the body's decoding systems. They are in the world *and* of it. Here we have the human plight and why humanity with its potential for infinite awareness can be so easily manipulatable and descend into such extremes of stupidity.

Once the Cult isolates five-sense mind from expanded awareness it can then program the mind with perceptions and beliefs by controlling information that the mind receives through the 'education' system of the formative years and the media perceptual bombardment and censorship of an entire lifetime. Limit perception and a sense of the possible through limiting knowledge by limiting and skewing information while censoring and discrediting that which could set people free. As the title of another of my books says *… And The Truth Shall Set You Free*. For this reason the last thing the Cult wants in circulation is the truth about anything – especially the reality of the eternal 'I' – and that's why it is desperate to control information. The Cult knows that information becomes perception which becomes behaviour which, collectively, becomes human society. Cult-controlled and funded mainstream 'science' denies the existence of an eternal 'I' and seeks to dismiss and trash all evidence to the contrary. Cult-controlled mainstream religion has a version of 'God' that is little more than a system of control and dictatorship that employs threats of damnation in an afterlife to control perceptions and behaviour in the here and now through fear and guilt. Neither is true and it's the 'neither' that the Cult wishes to suppress. This 'neither' is that everything is an expression, a point of attention, within an infinite state of consciousness which is the real meaning of the term 'God'.

Perceptual obsession with the 'physical body' and five-senses means that 'God' becomes personified as a bearded bloke sitting among the clouds or a raging bully who loves us if we do what 'he' wants and condemns us to the fires of hell if we don't. These are no more than a 'spiritual' fairy tales to control and dictate events and behaviour through fear of this 'God' which has bizarrely made 'God-fearing' in religious circles a state to be desired. I would suggest that fearing *anything* is not to be encouraged and celebrated, but rather deleted. You can see why 'God fearing' is so beneficial to the Cult and its religions when *they* decide what 'God' wants and what 'God' demands (the Cult demands) that everyone do. As the great American comedian Bill Hicks said satirising a Christian zealot: 'I think what God meant to say.' How much of this infinite awareness ('God') that we access is decided by how far we choose to expand our perceptions, self-identity and sense of the possible. The scale of self-identity reflects itself in the scale of awareness that we can connect with and are influenced by – how much knowing and insight we have instead of programmed perception. You cannot expand your awareness into the infinity of possibility when you believe that you are little me Peter the postman or Mary in marketing and nothing more. I'll deal with this in the concluding chapter because it's crucial to how we turnaround current events.

Where the Cult came from

When I realised in the early 1990s there was a Cult network behind global events I asked the obvious question: When did it start? I took it back to ancient Rome and Egypt and on to Babylon and Sumer in Mesopotamia, the 'Land Between Two Rivers', in what we now call Iraq. The two rivers are the Tigris and Euphrates and this region is of immense historical and other importance to the Cult, as is the land called Israel only 550 miles away by air. There is much more going with deep esoteric meaning across this whole region. It's not only about 'wars for oil'. Priceless artefacts from Mesopotamia were stolen or destroyed after the American and British invasion of Iraq in 2003 justified by the lies of Boy Bush and Tony Blair (their Cult masters) about non-existent 'weapons of mass destruction'. Mesopotamia was the location of Sumer (about 5,400BC to 1,750BC), and Babylon (about 2,350BC to 539BC). Sabbatians may have become immensely influential in the Cult in modern times but they are part of a network that goes back into the mists of history. Sumer is said by historians to be the 'cradle of civilisation'. I disagree. I say it was the re-start of what we call human civilisation after cataclysmic events symbolised in part as the 'Great Flood' destroyed the world that existed before. These fantastic upheavals that I have been describing in detail in the books since the early1990s appear in accounts and legends of ancient cultures across the world and they are supported by geological and biological evidence. Stone tablets found in Iraq detailing the Sumer period say the cataclysms were caused by non-human 'gods' they call the Anunnaki. These are described in terms of extraterrestrial visitations in which knowledge supplied by the Anunnaki is said to have been the source of at least one of the world's oldest writing systems and developments in astronomy, mathematics and architecture that were way ahead of their time. I have covered this subject at length in *The Biggest Secret* and *Children of the Matrix* and the same basic 'Anunnaki' story can be found in Zulu accounts in South Africa where the late and very great Zulu high shaman Credo Mutwa told me that the Sumerian Anunnaki were known by Zulus as the Chitauri or 'children of the serpent'. See my six-hour video interview with Credo on this subject entitled *The*

Reptilian Agenda recorded at his then home near Johannesburg in 1999 which you can watch on the Ickonic media platform.

The Cult emerged out of Sumer, Babylon and Egypt (and elsewhere) and established the Roman Empire before expanding with the Romans into northern Europe from where many empires were savagely imposed in the form of Cult-controlled societies all over the world. Mass death and destruction was their calling card. The Cult established its centre of operations in Europe and European Empires were Cult empires which allowed it to expand into a global force. Spanish and Portuguese colonialists headed for Central and South America while the British and French targeted North America. Africa was colonised by Britain, France, Belgium, the Netherlands, Portugal, Spain, Italy, and Germany. Some like Britain and France moved in on the Middle East. The British Empire was by far the biggest for a simple reason. By now Britain was the headquarters of the Cult from which it expanded to form Canada, the United States, Australia and New Zealand. The Sun never set on the British Empire such was the scale of its occupation. London remains a global centre for the Cult along with Rome and the Vatican although others have emerged in Israel and China. It is no accident that the 'virus' is alleged to have come out of China while Italy was chosen as the means to terrify the Western population into compliance with 'Covid' fascism. Nor that Israel has led the world in 'Covid' fascism and mass 'vaccination'.

You would think that I would mention the United States here, but while it has been an important means of imposing the Cult's will it is less significant than would appear and is currently in the process of having what power it does have deleted. The Cult in Europe has mostly loaded the guns for the US to fire. America has been controlled from Europe from the start through Cult operatives in Britain and Europe. The American Revolution was an illusion to make it appear that America was governing itself while very different forces were pulling the strings in the form of Cult families such as the Rothschilds through the Rockefellers and other subordinates. The Rockefellers are extremely close to Bill Gates and established both scalpel and drug 'medicine' and the World Health Organization. They play a major role in the development and circulation of vaccines through the Rockefeller Foundation on which Bill Gates said his Foundation is based. Why wouldn't this be the case when the Rockefellers and Gates are on the same team? Cult infiltration of human society goes way back into what we call history and has been constantly expanding and centralising power with the goal of establishing a global structure to dictate everything. Look how this has been advanced in great leaps with the 'Covid' hoax.

The non-human dimension

I researched and observed the comings and goings of Cult operatives through the centuries and even thousands of years as they were born, worked to promote the agenda within the secret society and satanic networks, and then died for others to replace them. Clearly there had to be a coordinating force that spanned this entire period while operatives who would not have seen the end goal in their lifetimes came and went advancing the plan over millennia. I went in search of that coordinating force with the usual support from the extraordinary synchronicity of my life which has been an almost daily experience since 1990. I saw common themes in religious texts and ancient cultures about a non-human force manipulating human society from the hidden. Christianity calls this force Satan, the Devil and demons; Islam refers to the Jinn or Djinn; Zulus have their Chitauri (spelt in other ways in different parts of Africa); and the Gnostic people in Egypt in the period around and before 400AD referred to this phenomena as the 'Archons', a word meaning rulers in Greek. Central American cultures speak of the 'Predators' among other names and the same theme is everywhere. I will use 'Archons' as a collective name for all of them. When you see how their nature and behaviour is described all these different sources are clearly talking about the same force. Gnostics described the Archons in terms of 'luminous fire' while Islam relates the Jinn to 'smokeless fire'. Some refer to beings in form that could occasionally be seen, but the most common of common theme is that they operate from unseen realms which means almost all existence to the visual processes of humans. I had concluded that this was indeed the foundation of human control and that the Cult was operating within the human frequency band on behalf of this hidden force when I came across the writings of Gnostics which supported my conclusions in the most extraordinary way.

A sealed earthen jar was found in 1945 near the town of Nag Hammadi about 75-80 miles north of Luxor on the banks of the River Nile in Egypt. Inside was a treasure trove of manuscripts and texts left by the Gnostic people some 1,600 years earlier. They included 13 leather-bound papyrus codices (manuscripts) and more than 50 texts written in Coptic Egyptian estimated to have been hidden in the jar in the period of 400AD although the source of the information goes back much further. Gnostics oversaw the Great or Royal Library of Alexandria, the fantastic depository of ancient texts detailing advanced knowledge and accounts of human history. The Library was dismantled and destroyed in stages over a long period with the death-blow delivered by the Cult-established Roman Church in the period around 415AD. The Church of Rome was the Church of Babylon relocated as I said earlier. Gnostics were not a race. They were a way of perceiving reality. Whenever they established themselves and their information circulated the terrorists of the Church of Rome would target them for destruction. This happened with the Great Library and with the Gnostic Cathars who were burned to death by the psychopaths after a long period of oppression at the siege of the Castle of Monségur in southern France in 1244. The Church has always been terrified of Gnostic information which demolishes the official Christian narrative although there is much in the Bible that supports the Gnostic view if you read it in another way. To anyone studying the texts of what became known as the Nag Hammadi Library it is clear that great swathes of Christian and Biblical belief has its origin with Gnostics sources going back to Sumer. Gnostic themes have been twisted to manipulate the perceived reality of Bible believers. Biblical texts have been in the open for centuries where they could be changed while Gnostic documents found at Nag Hammadi were sealed away and untouched for 1,600 years. What you see is what they wrote.

Use your *pneuma* not your *nous*

Gnosticism and Gnostic come from 'gnosis' which means knowledge, or rather *secret* knowledge, in the sense of spiritual awareness – knowledge about reality and life itself. The desperation of the Cult's Church of Rome to destroy the Gnostics can be understood when the knowledge they were circulating was the last thing the Cult wanted the population to know. Sixteen hundred years later the same Cult is working hard to undermine and silence me for the same reason. The dynamic between knowledge and ignorance is a constant. 'Time' appears to move on, but essential themes remain the same. We are told to 'use your nous', a Gnostic word for head/brain/intelligence. They said, however, that spiritual awakening or 'salvation' could only be secured by expanding awareness *beyond* what they called *nous* and into *pneuma* or Infinite Self. Obviously as I read these texts the parallels with what I have been saying since 1990 were fascinating to me. There is a universal truth that spans human history and in that case why wouldn't we be talking the same language 16 centuries apart? When you free yourself from the perception program of the five senses and explore expanded realms of consciousness you are going to connect with the same information no matter what the perceived 'era' within a manufactured timeline of a single and tiny range of manipulated frequency. Humans working with 'smart' technology or knocking rocks together in caves is only a timeline appearing to operate within the human frequency band. Expanded awareness and the knowledge it holds have always been there whether the era be Stone Age or computer age. We can only access that knowledge by opening ourselves to its frequency which the five-sense prison cell is designed to stop us doing. Gates, Fauci, Whitty, Vallance, Zuckerberg, Brin, Page, Wojcicki, Bezos, and all the others behind the 'Covid' hoax clearly have a long wait before their range of frequency can make that connection given that an open heart is crucial to that as we shall see. Instead of accessing knowledge directly through expanded awareness it is given to Cult operatives by the secret society networks of the Cult where it has been passed on over thousands of years outside the public arena. Expanded realms of consciousness is where great artists, composers and writers find their inspiration and where truth awaits anyone open enough to connect with it. We need to go there fast.

Archon hijack

A fifth of the Nag Hammadi texts describe the existence and manipulation of the Archons led by a 'Chief Archon' they call 'Yaldabaoth', or the 'Demiurge', and this is the Christian 'Devil', 'Satan', 'Lucifer', and his demons. Archons in Biblical symbolism are the 'fallen ones' which are also referred to as fallen angels after the angels expelled from heaven according to the Abrahamic religions of Judaism, Christianity and Islam. These angels are claimed to tempt humans to 'sin' ongoing and you will see how accurate that symbolism is during the rest of the book. The theme of 'original sin' is related to the 'Fall' when Adam and Eve were 'tempted by the serpent' and fell from a state of innocence and 'obedience' (connection) with God into a state of disobedience (disconnection). The Fall is said to have brought sin into the world and corrupted everything including human nature. Yaldabaoth, the 'Lord Archon', is described by Gnostics as a 'counterfeit spirit', 'The Blind One', 'The Blind God', and 'The Foolish One'. The Jewish name for Yaldabaoth in Talmudic writings is Samael which translates as 'Poison of God', or 'Blindness of God'. You see the parallels. Yaldabaoth in Islamic belief is the Muslim Jinn devil known as Shaytan – Shaytan is Satan as the same themes are found all over the world in every religion and culture. The 'Lord God' of the Old Testament is the 'Lord Archon' of Gnostic manuscripts and that's why he's such a bloodthirsty bastard. Satan is known by Christians as 'the Demon of Demons' and Gnostics called Yaldabaoth the 'Archon of Archons'. Both are known as 'The Deceiver'. We are talking about the same 'bloke' for sure and these common themes using different names, storylines and symbolism tell a common tale of the human plight.

Archons are referred to in Nag Hammadi documents as mind parasites, inverters, guards, gatekeepers, detainers, judges, pitiless ones and deceivers. The 'Covid' hoax alone is a glaring example of all these things. The Biblical 'God' is so different in the Old and New Testaments because they are not describing the same phenomenon. The vindictive, angry, hate-filled, 'God' of the Old Testament, known as Yahweh, is Yaldabaoth who is depicted in Cult-dictated popular culture as the 'Dark Lord', 'Lord of Time', Lord (Darth) Vader and Dormammu, the evil ruler of the 'Dark Dimension' trying to take over the 'Earth Dimension' in the Marvel comic movie, *Dr Strange*. Yaldabaoth is both the Old Testament 'god' and the Biblical 'Satan'. Gnostics referred to Yaldabaoth as the 'Great Architect of the Universe'and the Cult-controlled Freemason network calls their god 'the 'Great Architect of the Universe' (also Grand Architect). The 'Great Architect' Yaldabaoth is symbolised by the Cult as the all-seeing eye at the top of the pyramid on the Great Seal of the United States and the dollar bill. Archon is encoded in *arch*-itect as it is in *arch*-angels and *arch*-bishops. All religions have the theme of a force for good and force for evil in some sort of spiritual war and there is a reason for that – the theme is true. The Cult and its non-human masters are quite happy for this to circulate. They present themselves as the force for good fighting evil when they are really the force of evil (absence of love). The whole foundation of Cult modus operandi is inversion. They promote themselves as a force for good and anyone challenging them in pursuit of peace, love, fairness, truth and justice is condemned as a satanic force for evil. This has been the game plan throughout history whether the Church of Rome inquisitions of non-believers or 'conspiracy theorists' and 'anti-vaxxers' of today. The technique is the same whatever the timeline era.

Yaldabaoth is revolting (true)

Yaldabaoth and the Archons are said to have revolted against God with Yaldabaoth claiming to *be* God – the *All That Is*. The Old Testament 'God' (Yaldabaoth) demanded to be worshipped as such: '*I am* the LORD, and there is none else, there is no God beside me' (Isaiah 45:5). I have quoted in other books a man who said he was the unofficial son of the late Baron Philippe de Rothschild of the Mouton-Rothschild wine producing estates in France who died in 1988 and he told me about the Rothschild 'revolt from God'. The man said he was given the name Phillip Eugene de Rothschild and we shared long correspondence many years ago while he was living under another identity. He said that he was conceived through 'occult incest' which (within the Cult) was 'normal and to be admired'. 'Phillip' told me about his experience attending satanic rituals with rich and famous people whom he names and you can see them and the wider background to Cult Satanism in my other books starting with *The Biggest Secret*. Cult rituals are interactions with Archontic 'gods'. 'Phillip' described Baron Philippe de Rothschild as 'a master Satanist and hater of God' and he used the same term 'revolt from God' associated with Yaldabaoth/Satan/Lucifer/the Devil in describing the Sabbatian Rothschild dynasty. 'I played a key role in my family's revolt from God', he said. That role was to infiltrate in classic Sabbatian style the Christian Church, but eventually he escaped the mind-prison to live another life. The Cult has been targeting religion in a plan to make worship of the Archons the global one-world religion. Infiltration of Satanism into modern 'culture', especially among the young, through music videos, stage shows and other means, is all part of this.

Nag Hammadi texts describe Yaldabaoth and the Archons in their prime form as energy – consciousness – and say they can take form if they choose in the same way that consciousness takes form as a human. Yaldabaoth is called 'formless' and represents a deeply inverted, distorted and chaotic state of consciousness which seeks to attached to humans and turn them into a likeness of itself in an attempt at assimilation. For that to happen it has to manipulate humans into low frequency mental and emotional states that match its own. Archons can certainly appear in human form and this is the origin of the psychopathic personality. The energetic distortion Gnostics called Yaldabaoth *is* psychopathy. When psychopathic Archons take human form that human will be a psychopath as an expression of Yaldabaoth consciousness. Cult psychopaths are Archons in human form. The principle is the same as that portrayed in the 2009 *Avatar* movie when the American military travelled to a fictional Earth-like moon called Pandora in the Alpha Centauri star system to infiltrate a society of blue people, or Na'vi, by hiding within bodies that looked like the Na'vi. Archons posing as humans have a particular hybrid information field, part human, part Archon, (the ancient 'demigods') which processes information in a way that manifests behaviour to match their psychopathic evil, lack of empathy and compassion, and stops them being influenced by the empathy, compassion and love that a fully-human information field is capable of expressing. Cult bloodlines interbreed, be they royalty or dark suits, for this reason and you have their obsession with incest. Interbreeding with full-blown humans would dilute the Archontic energy field that guarantees psychopathy in its representatives in the human realm.

Gnostic writings say the main non-human forms that Archons take are *serpentine* (what I have called for decades 'reptilian' amid unbounded ridicule from the Archontically-programmed) and what Gnostics describe as 'an unborn baby or foetus with grey skin and dark, unmoving eyes'. This is an excellent representation of the ET 'Greys' of UFO folklore which large numbers of people claim to have seen and been abducted by – Zulu shaman Credo Mutwa among them. I agree with those that believe in extraterrestrial or interdimensional visitations today and for thousands of years past. No wonder with their advanced knowledge and technological capability they were perceived and worshipped as gods for technological and other 'miracles' they appeared to perform. Imagine someone arriving in a culture disconnected from the modern world with a smartphone and computer. They would be seen as a ‘god’ capable of ‘miracles’. The Renegade Mind, however, wants to know the source of everything and not only the way that source manifests as human or non-human. In the same way that a Renegade Mind seeks the original source material for the ‘Covid virus’ to see if what is claimed is true. The original source of Archons in form is consciousness – the distorted state of consciousness known to Gnostics as Yaldabaoth.

‘Revolt from God’ is energetic disconnection

Where I am going next will make a lot of sense of religious texts and ancient legends relating to ‘Satan’, Lucifer’ and the ‘gods’. Gnostic descriptions sync perfectly with the themes of my own research over the years in how they describe a consciousness distortion seeking to impose itself on human consciousness. I’ve referred to the core of infinite awareness in previous books as Infinite Awareness in Awareness of Itself. By that I mean a level of awareness that knows that it is all awareness and is aware of all awareness. From here comes the frequency of love in its true sense and balance which is what love is on one level – the balance of all forces into a single whole called Oneness and Isness. The more we disconnect from this state of love that many call ‘God’ the constituent parts of that Oneness start to unravel and express themselves as a part and not a whole. They become individualised as intellect, mind, selfishness, hatred, envy, desire for power over others, and such like. This is not a problem in the greater scheme in that ‘God’, the *All That Is*, can experience all these possibilities through different expressions of itself including humans. What we as expressions of the whole experience the *All That Is* experiences. We are the *All That Is* experiencing itself. As we withdraw from that state of Oneness we disconnect from its influence and things can get very unpleasant and very stupid. Archontic consciousness is at the extreme end of that. It has so disconnected from the influence of Oneness that it has become an inversion of unity and love, an inversion of everything, an inversion of life itself. Evil is appropriately live written backwards. Archontic consciousness is obsessed with death, an inversion of life, and so its manifestations in Satanism are obsessed with death. They use inverted symbols in their rituals such as the inverted pentagram and cross. Sabbatians as Archontic consciousness incarnate invert Judaism and every other religion and culture they infiltrate. They seek disunity and chaos and they fear unity and harmony as they fear love like garlic to a vampire. As a result the Cult, Archons incarnate, act with such evil, psychopathy and lack of empathy and compassion disconnected as they are from the source of love. How could Bill Gates and the rest of the Archontic psychopaths do what they have to human society in the 'Covid' era with all the death, suffering and destruction involved and have no emotional consequence for the impact on others? Now you know. Why have Zuckerberg, Brin, Page, Wojcicki and company callously censored information warning about the dangers of the 'vaccine' while thousands have been dying and having severe, sometimes life-changing reactions? Now you know. Why have Tedros, Fauci, Whitty, Vallance and their like around the world been using case and death figures they're aware are fraudulent to justify lockdowns and all the deaths and destroyed lives that have come from that? Now you know. Why did Christian Drosten produce and promote a 'testing' protocol that he knew couldn't test for infectious disease which led to a global human catastrophe. Now you know. The Archontic mind doesn't give a shit (Fig 17). I personally think that Gates and major Cult insiders are a form of AI cyborg that the Archons want humans to become.

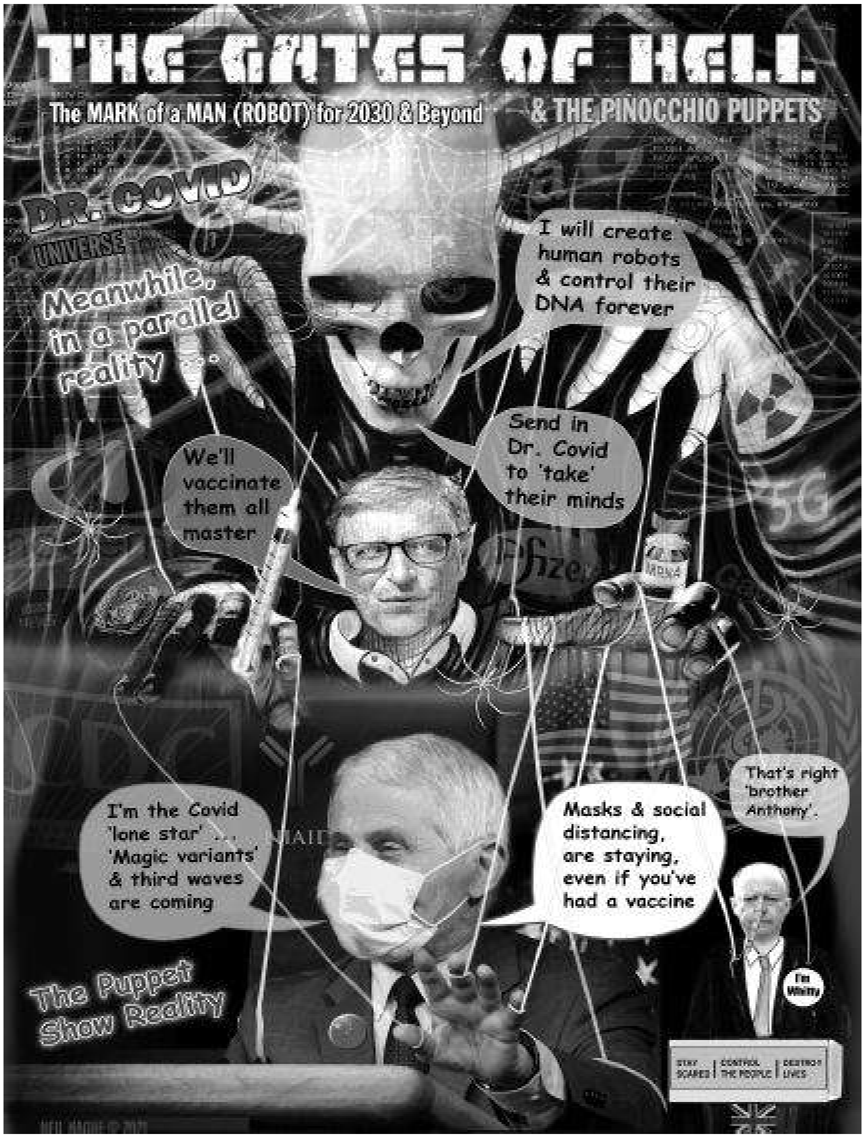

THE GATES OF HELL
The MARK of a MAN (ROBOT) for 2030 & Beyond
& THE PINOCCHIO PUPPETS
DR. COVID
UNIVERSE
Meanwhile, in a parallel reality . . .
I will create human robots & control their DNA forever
Send in Dr. Covid to 'take' their minds
We'll vaccinate them all master
5G
Pfizer
I'm the Covid 'lone star' ... 'Magic variants' & third waves are coming
NIAID
Masks & social distancing, are staying, even if you've had a vaccine
That's right 'brother Anthony'.
I'm Whitty
STAY SCARED | CONTROL THE PEOPLE | DESTROY LIVES
The Puppet Show Reality
NEIL HAGUE © 2021

Figure 17: Artist Neil Hague's version of the 'Covid' hierarchy.

Human batteries

A state of such inversion does have its consequences, however. The level of disconnection from the Source of All means that you withdraw from that source of energetic sustenance and creativity. This means that you have to find your own supply of energetic power and it has – *us*. When the Morpheus character in the first *Matrix* movie held up a battery he spoke a profound truth when he said: 'The Matrix is a computer-generated dream world built to keep us under control in order to change the human being into one of these.' The statement was true in all respects. We do live in a technologically-generated virtual reality simulation (more very shortly) and we have been manipulated to be an energy source for Archontic consciousness. The Disney-Pixar animated movie *Monsters, Inc*. in 2001 symbolised the dynamic when monsters in their world had no energy source and they would enter the human world to terrify children in their beds, catch the child's scream, terror (low-vibrational frequencies), and take that energy back to power the monster world. The lead character you might remember was a single giant eye and the symbolism of the Cult's all-seeing eye was obvious. Every thought and emotion is broadcast as a frequency unique to that thought and emotion. Feelings of love and joy, empathy and compassion, are high, quick, frequencies while fear, depression, anxiety, suffering and hate are low, slow, dense frequencies. Which kind do you think Archontic consciousness can connect with and absorb? In such a low and dense frequency state there's no way it can connect with the energy of love and joy. Archons can only feed off energy compatible with their own frequency and they and their Cult agents want to delete the human world of love and joy and manipulate the transmission of low vibrational frequencies through low-vibrational human mental and emotional states. *We are their energy source*. Wars are energetic banquets to the Archons – a world war even more so – and think how much low-frequency mental and emotional energy has been generated from the consequences for humanity of the 'Covid' hoax orchestrated by Archons incarnate like Gates.

The ancient practice of human sacrifice 'to the gods', continued in secret today by the Cult, is based on the same principle. 'The gods' are Archontic consciousness in different forms and the sacrifice is induced into a state of intense terror to generate the energy the Archontic frequency can absorb. Incarnate Archons in the ritual drink the blood which contains an adrenaline they crave which floods into the bloodstream when people are terrorised. Most of the sacrifices, ancient and modern, are children and the theme of 'sacrificing young virgins to the gods' is just code for children. They have a particular pre-puberty energy that Archons want more than anything and the energy of the young in general is their target. The California Department of Education wants students to chant the names of Aztec gods (Archontic gods) once worshipped in human sacrifice rituals in a curriculum designed to encourage them to ‘challenge racist, bigoted, discriminatory, imperialist/colonial beliefs’, join ‘social movements that struggle for social justice’, and ‘build new possibilities for a post-racist, post-systemic racism society’. It’s the usual Woke crap that inverts racism and calls it anti-racism. In this case solidarity with ‘indigenous tribes’ is being used as an excuse to chant the names of ‘gods’ to which people were sacrificed (and still are in secret). What an example of Woke’s inability to see beyond black and white, us and them, They condemn the colonisation of these tribal cultures by Europeans (quite right), but those cultures sacrificing people including children to their ‘gods’, and mass murdering untold numbers as the Aztecs did, is just fine. One chant is to the Aztec god Tezcatlipoca who had a man sacrificed to him in the 5th month of the Aztec calendar. His heart was cut out and he was eaten. Oh, that’s okay then. Come on children … after three … Other sacrificial ‘gods’ for the young to chant their allegiance include Quetzalcoatl, Huitzilopochtli and Xipe Totec. The curriculum says that ‘chants, affirmations, and energizers can be used to bring the class together, build unity around ethnic studies principles and values, and to reinvigorate the class following a lesson that may be emotionally taxing or even when student engagement may appear to be low’. Well, that’s the cover story, anyway. Chanting and mantras are the repetition of a particular frequency generated from the vocal cords and chanting the names of these Archontic ‘gods’ tunes you into their frequency. That is the last thing you want when it allows for energetic synchronisation, attachment and perceptual influence. Initiates chant the names of their ‘Gods’ in their rituals for this very reason.

Vampires of the Woke

Paedophilia is another way that Archons absorb the energy of children. Paedophiles possessed by Archontic consciousness are used as the conduit during sexual abuse for discarnate Archons to vampire the energy of the young they desire so much. Stupendous numbers of children disappear every year never to be seen again although you would never know from the media. Imagine how much low-vibrational energy has been generated by children during the 'Covid' hoax when so many have become depressed and psychologically destroyed to the point of killing themselves. Shocking numbers of children are now taken by the state from loving parents to be handed to others. I can tell you from long experience of researching this since 1996 that many end up with paedophiles and assets of the Cult through corrupt and Cult-owned social services which in the reframing era has hired many psychopaths and emotionless automatons to do the job. Children are even stolen to order using spurious reasons to take them by the corrupt and secret (because they're corrupt) 'family courts'. I have written in detail in other books, starting with *The Biggest Secret* in 1997, about the ubiquitous connections between the political, corporate, government, intelligence and military elites (Cult operatives) and Satanism and paedophilia. If you go deep enough both networks have an interlocking leadership. The Woke mentality has been developed by the Cult for many reasons: To promote almost every aspect of its agenda; to hijack the traditional political left and turn it fascist; to divide and rule; and to target agenda pushbackers. But there are other reasons which relate to what I am describing here. How many happy and joyful Wokers do you ever see especially at the extreme end? They are a mental and psychological mess consumed by emotional stress and constantly emotionally cocked for the next explosion of indignation at someone referring to a female as a female. They are walking, talking, batteries as Morpheus might say emitting frequencies which both enslave them in low-vibrational bubbles of perceptual limitation and feed the Archons. Add to this the hatred claimed to be love; fascism claimed to 'anti-fascism', racism claimed to be 'anti-racism';

exclusion claimed to inclusion; and the abuse-filled Internet trolling. You have a purpose-built Archontic energy system with not a wind turbine in sight and all founded on Archontic *inversion*. We have whole generations now manipulated to serve the Archons with their actions and energy. They will be doing so their entire adult lives unless they snap out of their Archon-induced trance. Is it really a surprise that Cult billionaires and corporations put so much money their way? Where is the energy of joy and laughter, including laughing at yourself which is confirmation of your own emotional security? Mark Twain said: 'The human race has one really effective weapon, and that is laughter.' We must use it all the time. Woke has destroyed comedy because it has no humour, no joy, sense of irony, or self-deprecation. Its energy is dense and intense. *Mmmmm*, lunch says the Archontic frequency. Rudolf Steiner (1861-1925) was the Austrian philosopher and famous esoteric thinker who established Waldorf education or Steiner schools to treat children like unique expressions of consciousness and not minds to be programmed with the perceptions determined by authority. I'd been writing about this energy vampiring for decades when I was sent in 2016 a quote by Steiner. He was spot on:

> There are beings in the spiritual realms for whom anxiety and fear emanating from human beings offer welcome food. When humans have no anxiety and fear, then these creatures starve. If fear and anxiety radiates from people and they break out in panic, then these creatures find welcome nutrition and they become more and more powerful. These beings are hostile towards humanity. Everything that feeds on negative feelings, on anxiety, fear and superstition, despair or doubt, are in reality hostile forces in super-sensible worlds, launching cruel attacks on human beings, while they are being fed ... These are exactly the feelings that belong to contemporary culture and materialism; because it estranges people from the spiritual world, it is especially suited to evoke hopelessness and fear of the unknown in people, thereby calling up the above mentioned hostile forces against them.

Pause for a moment from this perspective and reflect on what has happened in the world since the start of 2020. Not only will pennies drop, but billion dollar bills. We see the same theme from Don Juan Matus, a Yaqui Indian shaman in Mexico and the information source for Peruvian-born writer, Carlos Castaneda, who wrote a series of books from the 1960s to 1990s. Don Juan described the force manipulating human society and his name for the Archons was the predator:

> We have a predator that came from the depths of the cosmos and took over the rule of our lives. Human beings are its prisoners. The predator is our lord and master. It has rendered us docile, helpless. If we want to protest, it suppresses our protest. If we want to act independently, it demands that we don't do so ... indeed we are held prisoner!
>
> They took us over because we are food to them, and they squeeze us mercilessly because we are their sustenance. Just as we rear chickens in coops, the predators rear us in human coops, humaneros. Therefore, their food is always available to them.

Different cultures, different eras, same recurring theme.

The 'ennoia' dilemma

Nag Hammadi Gnostic manuscripts say that Archon consciousness has no 'ennoia'. This is directly translated as 'intentionality', but I'll use the term 'creative imagination'. The *All That Is* in awareness of itself is the source of all creativity – all possibility – and the more disconnected you are from that source the more you are subsequently denied 'creative imagination'. Given that Archon consciousness is almost entirely disconnected it severely lacks creativity and has to rely on far more mechanical processes of thought and exploit the creative potential of those that do have 'ennoia'. You can see cases of this throughout human society. Archon consciousness almost entirely dominates the global banking system and if we study how that system works you will appreciate what I mean. Banks manifest 'money' out of nothing by issuing lines of 'credit' which is 'money' that has never, does not, and will never exist except in theory. It's a confidence trick. If you think 'credit' figures-on-a-screen 'money' is worth anything you accept it as payment. If you don't then the whole system collapses through lack of confidence in the value of that 'money'. Archontic bankers with no 'ennoia' are 'lending' 'money' that doesn't exist to humans that *do* have creativity – those that have the inspired ideas and create businesses and products. Archon banking feeds off human creativity which it controls through ‘money’ creation and debt. Humans have the creativity and Archons exploit that for their own benefit and control while having none themselves. Archon Internet platforms like Facebook claim joint copyright of everything that creative users post and while Archontic minds like Zuckerberg may officially head that company it will be human creatives on the staff that provide the creative inspiration. When you have limitless ‘money’ you can then buy other companies established by creative humans. Witness the acquisition record of Facebook, Google and their like. Survey the Archon-controlled music industry and you see non-creative dark suit executives making their fortune from the human creativity of their artists. The cases are endless. Research the history of people like Gates and Zuckerberg and how their empires were built on exploiting the creativity of others. Archon minds cannot create out of nothing, but they are skilled (because they have to be) in what Gnostic texts call ‘countermimicry’. They can imitate, but not innovate. Sabbatians trawl the creativity of others through backdoors they install in computer systems through their cybersecurity systems. Archon-controlled China is globally infamous for stealing intellectual property and I remember how Hong Kong, now part of China, became notorious for making counterfeit copies of the creativity of others – ‘countermimicry’. With the now pervasive and all-seeing surveillance systems able to infiltrate any computer you can appreciate the potential for Archons to vampire the creativity of humans. Author John Lamb Lash wrote in his book about the Nag Hammadi texts, *Not In His Image*:

> Although they cannot originate anything, because they lack the divine factor of ennoia (intentionality), Archons can imitate with a vengeance. Their expertise is simulation (HAL, virtual reality). The Demiurge [Yaldabaoth] fashions a heaven world copied from the fractal patterns [of the original] ... His construction is celestial kitsch, like the fake Italianate villa of a Mafia don complete with militant angels to guard every portal.

This brings us to something that I have been speaking about since the turn of the millennium. Our reality is a simulation; a virtual reality that we think is real. No, I’m not kidding.

Human reality? Well, virtually

I had pondered for years about whether our reality is ‘real’ or some kind of construct. I remembered being immensely affected on a visit as a small child in the late 1950s to the then newly-opened Planetarium on the Marylebone Road in London which is now closed and part of the adjacent Madame Tussauds wax museum. It was in the middle of the day, but when the lights went out there was the night sky projected in the Planetarium’s domed ceiling and it appeared to be so real. The experience never left me and I didn’t know why until around the turn of the millennium when I became certain that our ‘night sky’ and entire reality is a projection, a virtual reality, akin to the illusory world portrayed in the *Matrix* movies. I looked at the sky one day in this period and it appeared to me like the domed roof of the Planetarium. The release of the first *Matrix* movie in 1999 also provided a synchronistic and perfect visual representation of where my mind had been going for a long time. I hadn’t come across the Gnostic Nag Hammadi texts then. When I did years later the correlation was once again astounding. As I read Gnostic accounts from 1,600 years and more earlier it was clear that they were describing the same simulation phenomenon. They tell how the Yaldabaoth ‘Demiurge’ and Archons created a ‘bad copy’ of original reality to rule over all that were captured by its illusions and the body was a prison to trap consciousness in the ‘bad copy’ fake reality. Read how Gnostics describe the ‘bad copy’ and update that to current times and they are referring to what we would call today a virtual reality simulation.

Author John Lamb Lash said ‘the Demiurge fashions a heaven world copied from the fractal patterns’ of the original through expertise in ‘HAL’ or virtual reality simulation. Fractal patterns are part of the energetic information construct of our reality, a sort of blueprint. If these patterns were copied in computer terms it would indeed give you a copy of a ‘natural’ reality in a non-natural frequency and digital form. The principle is the same as making a copy of a website. The original website still exists, but now you can change the copy version to make it whatever you like and it can become very different to the original website. Archons have done this with our reality, a *synthetic* copy of prime reality that still exists beyond the frequency walls of the simulation. Trapped within the illusions of this synthetic Matrix, however, were and are human consciousness and other expressions of prime reality and this is why the Archons via the Cult are seeking to make the human body synthetic and give us synthetic AI minds to complete the job of turning the entire reality synthetic including what we perceive to be the natural world. To quote Kurzweil: 'Nanobots will infuse all the matter around us with information. Rocks, trees, everything will become these intelligent creatures.' Yes, *synthetic* 'creatures' just as 'Covid' and other genetically-manipulating 'vaccines' are designed to make the human body synthetic. From this perspective it is obvious why Archons and their Cult are so desperate to infuse synthetic material into every human with their 'Covid' scam.

Let there be (electromagnetic) light

Yaldabaoth, the force that created the simulation, or Matrix, makes sense of the Gnostic reference to 'The Great Architect' and its use by Cult Freemasonry as the name of its deity. The designer of the Matrix in the movies is called 'The Architect' and that trilogy is jam-packed with symbolism relating to these subjects. I have contended for years that the angry Old Testament God (Yaldabaoth) is the 'God' being symbolically 'quoted' in the opening of Genesis as 'creating the world'. This is not the creation of prime reality – it's the creation of the *simulation*. The Genesis 'God' says: 'Let there be Light: and there was light.' But what is this 'Light'? I have said for decades that the speed of light (186,000 miles per second) is not the fastest speed possible as claimed by mainstream science and is in fact the frequency walls or outer limits of the Matrix. You can't have a fastest or slowest anything within all possibility when everything is possible. The human body is encoded to operate within the speed of light or *within the simulation* and thus we see only the tiny frequency band of visible *light*. Near-death experiencers who perceive reality outside the body during temporary 'death' describe a very different form of light and this is supported by the Nag Hammadi texts. Prime reality beyond the simulation ('Upper Aeons' to the Gnostics) is described as a realm of incredible beauty, bliss, love and harmony – a realm of 'watery light' that is so powerful 'there are no shadows'. Our false reality of Archon control, which Gnostics call the 'Lower Aeons', is depicted as a realm with a different kind of 'light' and described in terms of chaos, 'Hell', 'the Abyss' and 'Outer Darkness', where trapped souls are tormented and manipulated by demons (relate that to the 'Covid' hoax alone). The watery light theme can be found in near-death accounts and it is not the same as *simulation* 'light' which is electromagnetic or radiation light within the speed of light – the 'Lower Aeons'. Simulation 'light' is the 'luminous fire' associated by Gnostics with the Archons. The Bible refers to Yaldabaoth as 'that old serpent, called the Devil, and Satan, which deceiveth the whole world' (Revelation 12:9). I think that making a simulated copy of prime reality ('countermimicry') and changing it dramatically while all the time manipulating humanity to believe it to be real could probably meet the criteria of deceiving the whole world. Then we come to the Cult god Lucifer – the *Light Bringer*. Lucifer is symbolic of Yaldabaoth, the bringer of radiation light that forms the bad copy simulation within the speed of light. 'He' is symbolised by the lighted torch held by the Statue of Liberty and in the name 'Illuminati'. Sabbatian-Frankism declares that Lucifer is the true god and Lucifer is the real god of Freemasonry honoured as their 'Great or Grand Architect of the Universe' (simulation).

I would emphasise, too, the way Archontic technologically-generated luminous fire of radiation has deluged our environment since I was a kid in the 1950s and changed the nature of The Field with which we constantly interact. Through that interaction technological radiation is changing us. The Smart Grid is designed to operate with immense levels of communication power with 5G expanding across the world and 6G, 7G, in the process of development. Radiation is the simulation and the Archontic manipulation system. Why wouldn't the Archon Cult wish to unleash radiation upon us to an ever-greater extreme to form

Kurzweil's 'cloud'? The plan for a synthetic human is related to the need to cope with levels of radiation beyond even anything we've seen so far. Biological humans would not survive the scale of radiation they have in their script. The Smart Grid is a technological sub-reality within the technological simulation to further disconnect five-sense perception from expanded consciousness. It's a technological prison of the mind.

Infusing the 'spirit of darkness'

A recurring theme in religion and native cultures is the manipulation of human genetics by a non-human force and most famously recorded as the biblical 'sons of god' (the god*s* plural in the original) who interbred with the daughters of men. The Nag Hammadi *Apocryphon of John* tells the same story this way:

> He [Yaldabaoth] sent his angels [Archons/demons] to the daughters of men, that they might take some of them for themselves and raise offspring for their enjoyment. And at first they did not succeed. When they had no success, they gathered together again and they made a plan together ... And the angels changed themselves in their likeness into the likeness of their mates, filling them with the spirit of darkness, which they had mixed for them, and with evil ... And they took women and begot children out of the darkness according to the likeness of their spirit.

Possession when a discarnate entity takes over a human body is an age-old theme and continues today. It's very real and I've seen it. Satanic and secret society rituals can create an energetic environment in which entities can attach to initiates and I've heard many stories of how people have changed their personality after being initiated even into lower levels of the Freemasons. I have been inside three Freemasonic temples, one at a public open day and two by just walking in when there was no one around to stop me. They were in Ryde, the town where I live, Birmingham, England, when I was with a group, and Boston, Massachusetts. They all felt the same energetically – dark, dense, low-vibrational and sinister. Demonic attachment can happen while the initiate has no idea what is going on. To them it's just a ritual to get in the Masons and do a bit of good business. In the far more extreme rituals of Satanism human possession is even more powerful and they are designed to make possession possible. The hierarchy of the Cult is dictated by the power and perceived status of the possessing Archon. In this way the Archon hierarchy becomes the Cult hierarchy. Once the entity has attached it can influence perception and behaviour and if it attaches to the extreme then so much of its energy (information) infuses into the body information field that the hologram starts to reflect the nature of the possessing entity. This is the *Exorcist* movie type of possession when facial features change and it's known as shapeshifting. Islam's Jinn are said to be invisible tricksters who change shape, 'whisper', confuse and take human form. These are all traits of the Archons and other versions of the same phenomenon. Extreme possession could certainty infuse the 'spirit of darkness' into a partner during sex as the Nag Hammadi texts appear to describe. Such an infusion can change genetics which is also energetic information. Human genetics is information and the 'spirit of darkness' is information. Mix one with the other and change must happen. Islam has the concept of a 'Jinn baby' through possession of the mother and by Jinn taking human form. There are many ways that human genetics can be changed and remember that Archons have been aware all along of advanced techniques to do this. What is being done in human society today – and far more – was known about by Archons at the time of the 'fallen ones' and their other versions described in religions and cultures.

Archons and their human-world Cult are obsessed with genetics as we see today and they know this dictates how information is processed into perceived reality during a human life. They needed to produce a human form that would decode the simulation and this is symbolically known as 'Adam and Eve' who left the 'garden' (prime reality) and 'fell' into Matrix reality. The simulation is not a 'physical' construct (there is no 'physical'); it is a source of information. Think Wi-Fi again. The simulation is an energetic field encoded with information and body-brain systems are designed to decode that information encoded in wave or frequency form which is transmitted to the brain as electrical signals. These are decoded by the brain to construct our sense of reality – an illusory 'physical' world that only exists in the brain or the mind. Virtual reality games mimic this process using the same sensory decoding system. Information is fed to the senses to decode a virtual reality that can appear so real, but isn't (Figs 18 and 19). Some scientists believe – and I agree with them – that what we perceive as 'physical' reality only exists when we are looking or observing. The act of perception or focus triggers the decoding systems which turn waveform information into holographic reality. When we are not observing something our reality reverts from a holographic state to a waveform state. This relates to the same principle as a falling tree not making a noise unless someone is there to hear it or decode it. The concept makes sense from the simulation perspective. A computer is not decoding all the information in a Wi-Fi field all the time and only decodes or brings into reality on the screen that part of Wi-Fi that it's decoding – focusing upon – at that moment.

Figure 18: Virtual reality technology 'hacks' into the body's five-sense decoding system.

Figure 19: The result can be experienced as very 'real'.

Interestingly, Professor Donald Hoffman at the Department of Cognitive Sciences at the University of California, Irvine, says that our experienced reality is like a computer interface that shows us only the level with which we interact while hiding all that exists beyond it: 'Evolution shaped us with a user interface that hides the truth. Nothing that we see is the truth – the very language of space and time and objects is the wrong language to describe reality.' He is correct in what he says on so many levels. Space and time are not a universal reality. They are a phenomenon of decoded *simulation* reality as part of the process of enslaving our sense of reality. Near-death experiencers report again and again how space and time did not exist as we perceive them once they were free of the body – body decoding systems. You can appreciate from this why Archons and their Cult are so desperate to entrap human attention in the five senses where we are in the Matrix and of the Matrix. Opening your mind to expanded states of awareness takes you beyond the information confines of the simulation and you become aware of knowledge and insights denied to you before. This is what we call 'awakening' – *awakening from the Matrix* – and in the final chapter I will relate this to current events.

Where are the 'aliens'?

A simulation would explain the so-called 'Fermi Paradox' named after Italian physicist Enrico Fermi (1901-1954) who created the first nuclear reactor. He considered the question of why there is such a lack of extraterrestrial activity when there are so many stars and planets in an apparently vast universe; but what if the night sky that we see, or think we do, is a simulated projection as I say? If you control the simulation and your aim is to hold humanity fast in essential ignorance would you want other forms of life including advanced life coming and going sharing information with humanity? Or would you want them to believe they were isolated and apparently alone? Themes of human isolation and apartness are common whether they be the perception of a lifeless universe or the fascist isolation laws of the 'Covid' era. Paradoxically the very existence of a simulation means that we are not alone when some force had to construct it. My view is that experiences that people have reported all over the world for centuries with Reptilians and Grey entities are Archon phenomena as Nag Hammadi texts describe; and that benevolent 'alien' interactions are non-human groups that come in and out of the simulation by overcoming Archon attempts to keep them out. It should be highlighted, too, that Reptilians and Greys are obsessed with *genetics* and *technology* as related by cultural accounts and those who say they have been abducted by them. Technology is their way of overcoming some of the limitations in their creative potential and our technology-driven and controlled human society of today is *arch*etypical Archon-Reptilian-Grey modus operandi. Technocracy is really *Archon*tocracy. The Universe does not have to be as big as it appears with a simulation. There is no space or distance only information decoded into holographic reality. What we call 'space' is only the absence of holographic 'objects' and that 'space' is The Field of energetic information which connects everything into a single whole. The same applies with the artificially-generated information field of the simulation. The Universe is not big or small as a physical reality. It is decoded information, that's all, and its perceived size is decided by the way the simulation is encoded to make it appear. The entire night sky as we perceive it only exists in our brain and so where are those 'millions of light years'? The 'stars' on the ceiling of the Planetarium looked a vast distance away.

There's another point to mention about 'aliens'. I have been highlighting since the 1990s the plan to stage a fake 'alien invasion' to justify the centralisation of global power and a world military. Nazi scientist Werner von Braun, who was taken to America by Operation Paperclip after World War Two to help found NASA, told his American assistant Dr Carol Rosin about the Cult agenda when he knew he was dying in 1977. Rosin said that he told her about a sequence that would lead to total human control by a one-world government. This included threats from terrorism, rogue nations, meteors and asteroids before finally an 'alien invasion'. All of these things, von Braun said, would be bogus and what I would refer to as a No-Problem-Reaction-Solution. Keep this in mind when 'the aliens are coming' is the new mantra. The aliens are not coming – they are *already here* and they have infiltrated human society while looking human. French-Canadian investigative journalist Serge Monast said in 1994 that he had uncovered a NASA/military operation called Project Blue Beam which fits with what Werner von Braun predicted. Monast died of a 'heart attack' in 1996 the day after he was arrested and spent a night in prison. He was 51. He said Blue Beam was a plan to stage an alien invasion that would include religious figures beamed holographically into the sky as part of a global manipulation to usher in a 'new age' of worshipping what I would say is the Cult 'god' Yaldabaoth in a one-world religion. Fake holographic asteroids are also said to be part of the plan which again syncs with von Braun. How could you stage an illusory threat from asteroids unless they were holographic inserts? This is pretty straightforward given the advanced technology outside the public arena and the fact that our 'physical' reality is holographic anyway. Information fields would be projected and we would decode them into the illusion of a 'physical' asteroid. If they can sell a global 'pandemic' with a 'virus' that doesn't exist what will humans not believe if government and media tell them?

All this is particularly relevant as I write with the Pentagon planning to release in June, 2021, information about 'UFO sightings'. I have been following the UFO story since the early 1990s and the common theme throughout has been government and military denials and cover up. More recently, however, the Pentagon has suddenly become more talkative and apparently open with Air Force pilot radar images released of unexplained craft moving and changing direction at speeds well beyond anything believed possible with human technology. Then, in March, 2021, former Director of National Intelligence John Ratcliffe said a Pentagon report months later in June would reveal a great deal of information about UFO sightings unknown to the public. He said the report would have 'massive implications'. The order to do this was included bizarrely in a $2.3 trillion 'coronavirus' relief and government funding bill passed by the Trump administration at the end of 2020. I would add some serious notes of caution here. I have been pointing out since the 1990s that the US military and intelligence networks have long had craft – 'flying saucers' or anti-gravity craft – which any observer would take to be extraterrestrial in origin. Keeping this knowledge from the public allows craft flown by *humans* to be perceived as alien visitations. I am not saying that 'aliens' do not exist. I would be the last one to say that, but we have to be streetwise here. President Ronald Reagan told the UN General Assembly in 1987: 'I occasionally think how quickly our differences worldwide would vanish if we were facing an alien threat from outside this world.' That's the idea. Unite against a common 'enemy' with a common purpose behind your 'saviour force' (the Cult) as this age-old technique of mass manipulation goes global.

Science moves this way ...

I could find only one other person who was discussing the simulation hypothesis publicly when I concluded it was real. This was Nick Bostrom, a Swedish-born philosopher at the University of Oxford, who has explored for many years the possibility that human reality is a computer simulation although his version and mine are not the same. Today the simulation and holographic reality hypothesis have increasingly entered the scientific mainstream. Well, the more open-minded mainstream, that is. Here are a few of the ever-gathering examples. American nuclear physicist Silas Beane led a team of physicists at the University of Bonn in Germany pursuing the question of whether we live in a simulation. They concluded that we probably do and it was likely based on a lattice of cubes. They found that cosmic rays align with that specific pattern. The team highlighted the Greisen–Zatsepin–Kuzmin (GZK) limit which refers to cosmic ray particle interaction with cosmic background radiation that creates an apparent boundary for cosmic ray particles. They say in a paper entitled 'Constraints on the Universe as a Numerical Simulation' that this 'pattern of constraint' is exactly what you would find with a computer simulation. They also made the point that a simulation would create its own 'laws of physics' that would limit possibility. I've been making the same point for decades that the *perceived* laws of physics relate only to this reality, or what I would later call the simulation. When designers write codes to create computer and virtual reality games they are the equivalent of the laws of physics for that game. Players interact within the limitations laid out by the coding. In the same way those who wrote the codes for the simulation decided the laws of physics that would apply. These can be overridden by expanded states of consciousness, but not by those enslaved in only five-sense awareness where simulation codes rule. Overriding the codes is what people call 'miracles'. They are not. They are bypassing the encoded limits of the simulation. A population caught in simulation perception would have no idea that this was their plight. As the Bonn paper said: 'Like a prisoner in a pitch-black cell we would not be able to see the "walls" of our prison,' That's true if people remain mesmerised by the five senses. Open to expanded awareness and those walls become very clear. The main one is the speed of light.

American theoretical physicist James Gates is another who has explored the simulation question and found considerable evidence to support the idea. Gates was Professor of Physics at the University of Maryland, Director of The Center for String and Particle Theory, and on Barack Obama's Council of Advisors on Science and Technology. He and his team found *computer codes* of digital data embedded in the fabric of our reality. They relate to on-off electrical charges of 1 and 0 in the binary system used by computers. 'We have no idea what they are doing there', Gates said. They found within the energetic fabric mathematical sequences known as error-correcting codes or block codes that 'reboot' data to its original state or 'default settings' when something knocks it out of sync. Gates was asked if he had found a set of equations embedded in our reality indistinguishable from those that drive search engines and browsers and he said: 'That is correct.' Rich Terrile, director of the Centre for Evolutionary Computation and Automated Design at NASA's Jet

Propulsion Laboratory, has said publicly that he believes the Universe is a digital hologram that must have been created by a form of intelligence. I agree with that in every way. Waveform information is delivered electrically by the senses to the brain which constructs a *digital* holographic reality that we call the 'world'. This digital level of reality can be read by the esoteric art of numerology. Digital holograms are at the cutting edge of holographics today. We have digital technology everywhere designed to access and manipulate our digital level of perceived reality. Synthetic mRNA in 'Covid vaccines' has a digital component to manipulate the body's digital 'operating system'.

Reality is numbers

How many know that our reality can be broken down to numbers and codes that are the same as computer games? Max Tegmark, a physicist at the Massachusetts Institute of Technology (MIT), is the author of *Our Mathematical Universe* in which he lays out how reality can be entirely described by numbers and maths in the way that a video game is encoded with the 'physics' of computer games. Our world and computer virtual reality are essentially the same. Tegmark imagines the perceptions of characters in an advanced computer game when the graphics are so good they don't know they are in a game. They think they can bump into real objects (electromagnetic resistance in our reality), fall in love and feel emotions like excitement. When they began to study the apparently 'physical world' of the video game they would realise that everything was made of pixels (which have been found in our energetic reality as must be the case when on one level our world is digital). What computer game characters thought was physical 'stuff', Tegmark said, could actually be broken down into numbers:

> And we're exactly in this situation in our world. We look around and it doesn't seem that mathematical at all, but everything we see is made out of elementary particles like quarks and electrons. And what properties does an electron have? Does it have a smell or a colour or a texture? No! ... We physicists have come up with geeky names for [Electron] properties, like

electric charge, or spin, or lepton number, but the electron doesn't care what we call it, the properties are just numbers.

This is the illusory reality Gnostics were describing. This is the simulation. The A, C, G, and T codes of DNA have a binary value – A and C = 0 while G and T = 1. This has to be when the simulation is digital and the body must be digital to interact with it. Recurring mathematical sequences are encoded throughout reality and the body. They include the Fibonacci sequence in which the two previous numbers are added to get the next one, as in ... 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, etc. The sequence is encoded in the human face and body, proportions of animals, DNA, seed heads, pine cones, trees, shells, spiral galaxies, hurricanes and the number of petals in a flower. The list goes on and on. There are fractal patterns – a 'never-ending pattern that is infinitely complex and self-similar across all scales in the as above, so below, principle of holograms. These and other famous recurring geometrical and mathematical sequences such as Phi, Pi, Golden Mean, Golden Ratio and Golden Section are *computer codes* of the simulation. I had to laugh and give my head a shake the day I finished this book and it went into the production stage. I was sent an article in *Scientific American* published in April, 2021, with the headline 'Confirmed! We Live in a Simulation'. Two decades after I first said our reality is a simulation and the speed of light is it's outer limit the article suggested that we do live in a simulation and that the speed of light is its outer limit. I left school at 15 and never passed a major exam in my life while the writer was up to his eyes in qualifications. As I will explain in the final chapter *knowing* is far better than thinking and they come from very different sources. The article rightly connected the speed of light to the processing speed of the 'Matrix' and said what has been in my books all this time … 'If we are in a simulation, as it appears, then space is an abstract property written in code. It is not real'. No it's not and if we live in a simulation something created it and it wasn't *us*. 'That David Icke says we are manipulated by aliens' – he's crackers.'

Wow …

The reality that humanity thinks is so real is an illusion. Politicians, governments, scientists, doctors, academics, law enforcement, media, school and university curriculums, on and on, are all founded on a world that *does not exist* except as a simulated prison cell. Is it such a stretch to accept that 'Covid' doesn't exist when our entire 'physical' reality doesn't exist? Revealed here is the knowledge kept under raps in the Cult networks of compartmentalised secrecy to control humanity's sense of reality by inducing the population to believe in a reality that's not real. If it wasn't so tragic in its experiential consequences the whole thing would be hysterically funny. None of this is new to Renegade Minds. Ancient Greek philosopher Plato (about 428 to about 347BC) was a major influence on Gnostic belief and he described the human plight thousands of years ago with his Allegory of the Cave. He told the symbolic story of prisoners living in a cave who had never been outside. They were chained and could only see one wall of the cave while behind them was a fire that they could not see. Figures walked past the fire casting shadows on the prisoners' wall and those moving shadows became their sense of reality. Some prisoners began to study the shadows and were considered experts on them (today's academics and scientists), but what they studied was only an illusion (today's academics and scientists). A prisoner escaped from the cave and saw reality as it really is. When he returned to report this revelation they didn't believe him, called him mad and threatened to kill him if he tried to set them free. Plato's tale is not only a brilliant analogy of the human plight and our illusory reality. It describes, too, the dynamics of the 'Covid' hoax. I have only skimmed the surface of these subjects here. The aim of this book is to crisply connect all essential dots to put what is happening today into its true context. All subject areas and their connections in this chapter are covered in great evidential detail in *Everything You Need To Know, But Have Never Been Told* and *The Answer*.

They say that bewildered people 'can't see the forest for the trees'. Humanity, however, can't see the forest for the *twigs*. The five senses see only twigs while Renegade Minds can see the forest and it’s the forest where the answers lie with the connections that reveals. Breaking free of perceptual programming so the forest can be seen is the way we turn all this around. Not breaking free is how humanity got into this mess. The situation may seem hopeless, but I promise you it’s not. We are a perceptual heartbeat from paradise if only we knew.

CHAPTER TWELVE

Escaping Wetiko

Life is simply a vacation from the infinite
Dean Cavanagh

Renegade Minds weave the web of life and events and see common themes in the apparently random. They are always there if you look for them and their pursuit is aided by incredible synchronicity that comes when your mind is open rather than mesmerised by what it thinks it can see.

Infinite awareness is infinite possibility and the more of infinite possibility that we access the more becomes infinitely possible. That may be stating the apparently obvious, but it is a devastatingly-powerful fact that can set us free. We are a point of attention within an infinity of consciousness. The question is how much of that infinity do we choose to access? How much knowledge, insight, awareness, wisdom, do we want to connect with and explore? If your focus is only in the five senses you will be influenced by a fraction of infinite awareness. I mean a range so tiny that it gives new meaning to infinitesimal. Limitation of self-identity and a sense of the possible limit accordingly your range of consciousness. We are what we think we are. Life is what we think it is. The dream is the dreamer and the dreamer is the dream. Buddhist philosophy puts it this way: 'As a thing is viewed, so it appears.' Most humans live in the realm of touch, taste, see, hear, and smell and that's the limit of their sense of the possible and sense of self. Many will follow a religion and speak of a God in his heaven, but their lives are still dominated by the five senses in their perceptions and actions. The five senses become the arbiter of everything. When that happens all except a smear of infinity is sealed away from influence by the rigid, unyielding, reality bubbles that are the five-sense human or Phantom Self. Archon Cult methodology is to isolate consciousness within five-sense reality – the simulation – and then program that consciousness with a sense of self and the world through a deluge of life-long information designed to instil the desired perception that allows global control. Efforts to do this have increased dramatically with identity politics as identity bubbles are squeezed into the minutiae of five-sense detail which disconnect people even more profoundly from the infinite 'I'.

Five-sense focus and self-identity are like a firewall that limits access to the infinite realms. You only perceive one radio or television station and no other. We'll take that literally for a moment. Imagine a vast array of stations giving different information and angles on reality, but you only ever listen to one. Here we have the human plight in which the population is overwhelmingly confined to CultFM. This relates only to the frequency range of CultFM and limits perception and insight to that band – limits *possibility* to that band. It means you are connecting with an almost imperceptibly minuscule range of possibility and creative potential within the infinite Field. It's a world where everything seems apart from everything else and where synchronicity is rare. Synchronicity is defined in the dictionary as 'the happening by chance of two or more related or similar events at the same time'. Use of 'by chance' betrays a complete misunderstanding of reality. Synchronicity is not 'by chance'. As people open their minds, or 'awaken' to use the term, they notice more and more coincidences in their lives, bits of 'luck', apparently miraculous happenings that put them in the right place at the right time with the right people. Days become peppered with 'fancy meeting you here' and 'what are the chances of that?' My entire life has been lived like this and ever more so since my own colossal awakening in 1990 and 91 which transformed my sense of reality. Synchronicity is not 'by chance'; it is by accessing expanded realms of possibility which allow expanded potential for manifestation. People broadcasting the same vibe from the same openness of mind tend to be drawn 'by chance' to each other through what I call frequency magnetism and it's not only people. In the last more than 30 years incredible synchronicity has also led me through the Cult maze to information in so many forms and to crucial personal experiences. These 'coincidences' have allowed me to put the puzzle pieces together across an enormous array of subjects and situations. Those who have breached the bubble of five-sense reality will know exactly what I mean and this escape from the perceptual prison cell is open to everyone whenever they make that choice. This may appear super-human when compared with the limitations of 'human', but it's really our natural state. 'Human' as currently experienced is consciousness in an unnatural state of induced separation from the infinity of the whole. I'll come to how this transformation into unity can be made when I have described in more detail the force that holds humanity in servitude by denying this access to infinite self.

The Wetiko factor

I have been talking and writing for decades about the way five-sense mind is systematically barricaded from expanded awareness. I have used the analogy of a computer (five-sense mind) and someone at the keyboard (expanded awareness). Interaction between the computer and the operator is symbolic of the interaction between five-sense mind and expanded awareness. The computer directly experiences the Internet and the operator experiences the Internet via the computer which is how it's supposed to be – the two working as one. Archons seek to control that point where the operator connects with the computer to stop that interaction (Fig 20). Now the operator is banging the keyboard and clicking the mouse, but the computer is not responding and this happens when the computer is taken over – *possessed* – by an appropriately-named computer 'virus'. The operator has lost all influence over the computer which goes its own way making decisions under the control of the 'virus'. I have just described the dynamic through which the force known to Gnostics as Yaldabaoth and Archons disconnects five-sense mind from expanded awareness to imprison humanity in perceptual servitude.

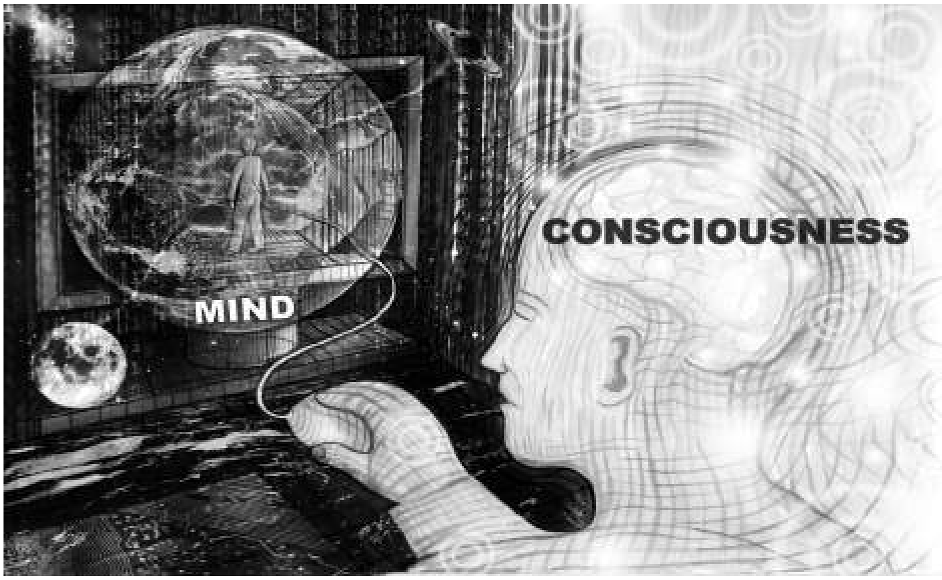


Figure 20: The mind 'virus' I have been writing about for decades seeks to isolate five-sense mind (the computer) from the true 'I'. (Image by Neil Hague).

About a year ago I came across a Native American concept of Wetiko which describes precisely the same phenomenon. Wetiko is the spelling used by the Cree and there are other versions including wintiko and windigo used by other tribal groups. They spell the name with lower case, but I see Wetiko as a proper noun as with Archons and prefer a capital. I first saw an article about Wetiko by writer and researcher Paul Levy which so synced with what I had been writing about the computer/operator disconnection and later the Archons. I then read his book, the fascinating *Dispelling Wetiko, Breaking the Spell of Evil*. The parallels between what I had concluded long before and the Native American concept of Wetiko were so clear and obvious that it was almost funny. For Wetiko see the Gnostic Archons for sure and the Jinn, the Predators, and every other name for a force of evil, inversion and chaos. Wetiko is the Native American name for the force that divides the computer from the operator (Fig 21). Indigenous author Jack D. Forbes, a founder of the Native American movement in the 1960s, wrote another book about Wetiko entitled *Columbus And Other Cannibals – The Wetiko Disease of Exploitation, Imperialism, and Terrorism* which I also read. Forbes says that Wetiko refers to an evil person or spirit 'who terrorizes other creatures by means of terrible acts, including cannibalism'. Zulu shaman Credo Mutwa told me that African accounts tell how cannibalism was brought into the world by the Chitauri 'gods' – another manifestation of Wetiko. The distinction between 'evil person or spirit' relates to Archons/Wetiko possessing a human or acting as pure consciousness. Wetiko is said to be a sickness of the soul or spirit and a state of being that takes but gives nothing back – the Cult and its operatives perfectly described. Black Hawk, a Native American war leader defending their lands from confiscation, said European invaders had 'poisoned hearts' – Wetiko hearts – and that this would spread to native societies. Mention of the heart is very significant as we shall shortly see. Forbes writes: 'Tragically, the history of the world for the past 2,000 years is, in great part, the story of the epidemiology of the wetiko disease.' Yes, and much longer. Forbes is correct when he says: 'The wetikos destroyed Egypt and Babylon and Athens and Rome and Tenochtitlan [capital of the Aztec empire] and perhaps now they will destroy the entire earth.' Evil, he said, is the number one export of a Wetiko culture – see its globalisation with 'Covid'. Constant war, mass murder, suffering of all kinds, child abuse, Satanism, torture and human sacrifice are all expressions of Wetiko and the Wetiko possessed. The world is Wetiko made manifest, *but it doesn't have to be*. There is a way out of this even now.

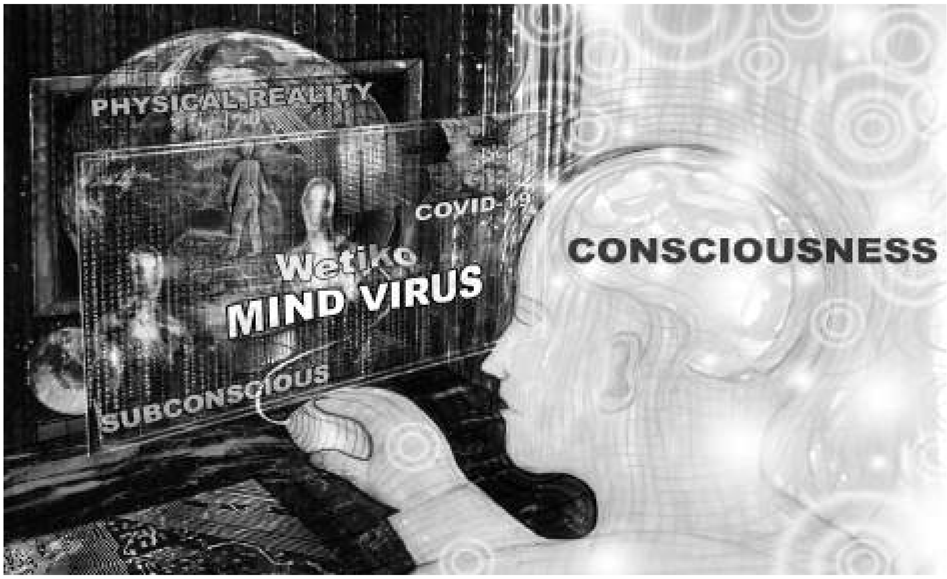


Figure 21: The mind 'virus' is known to Native Americans as 'Wetiko'. (Image by Neil Hague).

Cult of Wetiko

Wetiko is the Yaldabaoth frequency distortion that seeks to attach to human consciousness and absorb it into its own. Once this connection is made Wetiko can drive the perceptions of the target which they believe to be coming from their own mind. All the horrors of history and today from mass killers to Satanists, paedophiles like Jeffrey Epstein and other psychopaths, are the embodiment of Wetiko and express its state of being in all its grotesqueness. The Cult is Wetiko incarnate, Yaldabaoth incarnate, and it seeks to facilitate Wetiko assimilation of humanity in totality into its distortion by manipulating the population into low frequency states that match its own. Paul Levy writes: 'Holographically enforced within the psyche of every human being the wetiko virus pervades and underlies the entire field of consciousness, and can therefore potentially manifest through any one of us at any moment if we are not mindful.' The 'Covid' hoax has achieved this with many people, but others have not fallen into Wetiko's frequency lair. Players in the 'Covid' human catastrophe including Gates, Schwab, Tedros, Fauci, Whitty, Vallance, Johnson, Hancock, Ferguson, Drosten, and all the rest, including the psychopath psychologists, are expressions of Wetiko. This is why they have no compassion or empathy and no emotional consequence for what they do that would make them stop doing it. Observe all the people who support the psychopaths in authority against the Pushbackers despite the damaging impact the psychopaths have on their own lives and their family's lives. You are again looking at Wetiko possession which prevents them seeing through the lies to the obvious scam going on. *Why can't they see it?* Wetiko won't let them see it. The perceptual divide that has now become a chasm is between the Wetikoed and the non-Wetikoed.

Paul Levy describes Wetiko in the same way that I have long described the Archontic force. They are the same distorted consciousness operating across dimensions of reality: '… the subtle body of wetiko is not located in the third dimension of space and time, literally existing in another dimension … it is able to affect ordinary lives by mysteriously interpenetrating into our three-dimensional world.' Wetiko does this through its incarnate representatives in the Cult and by weaving itself into The Field which on our level of reality is the electromagnetic information field of the simulation or Matrix. More than that, the simulation *is* Wetiko / Yaldabaoth. Caleb Scharf, Director of Astrobiology at Columbia University, has speculated that 'alien life' could be so advanced that it has transcribed itself into the quantum realm to become what we call physics. He said intelligence indistinguishable from the fabric of the Universe would solve many of its greatest mysteries:

> Perhaps hyper-advanced life isn't just external. Perhaps it's already all around. It is embedded in what we perceive to be physics itself, from the root behaviour of particles and fields to the phenomena of complexity and emergence ... In other words, life might not just be in the equations. It might BE the equations [My emphasis].

Scharf said it is possible that 'we don't recognise advanced life because it forms an integral and unsuspicious part of what we've considered to be the natural world'. I agree. Wetiko/Yaldabaoth *is* the simulation. We are literally in the body of the beast. But that doesn't mean it has to control us. We all have the power to overcome Wetiko influence and the Cult knows that. I doubt it sleeps too well because it knows that.

Which Field?

This, I suggest, is how it all works. There are two Fields. One is the fierce electromagnetic light of the Matrix within the speed of light; the other is the 'watery light' of The Field beyond the walls of the Matrix that connects with the Great Infinity. Five-sense mind and the decoding systems of the body attach us to the Field of Matrix light. They have to or we could not experience this reality. Five-sense mind sees only the Matrix Field of information while our expanded consciousness is part of the Infinity Field. When we open our minds, and most importantly our hearts, to the Infinity Field we have a mission control which gives us an expanded perspective, a road map, to understand the nature of the five-sense world. If we are isolated only in five-sense mind there is no mission control. We're on our own trying to understand a world that's constantly feeding us information to ensure we do not understand. People in this state can feel 'lost' and bewildered with no direction or radar. You can see ever more clearly those who are influenced by the Fields of Big Infinity or little five-sense mind simply by their views and behaviour with regard to the 'Covid' hoax. We have had this division throughout known human history with the mass of the people on one side and individuals who could see and intuit beyond the walls of the simulation – Plato's prisoner who broke out of the cave and saw reality for what it is. Such people have always been targeted by Wetiko/Archon-possessed authority, burned at the stake or demonised as mad, bad and dangerous. The Cult today and its global network of 'anti-hate', 'anti-fascist' Woke groups are all expressions of Wetiko attacking those exposing the conspiracy, 'Covid' lies and the 'vaccine' agenda.

Woke as a whole is Wetiko which explains its black and white mentality and how at one it is with the Wetiko-possessed Cult. Paul Levy said: 'To be in this paradigm is to still be under the thrall of a two-valued logic – where things are either true or false – of a wetikoized mind.’ Wetiko consciousness is in a permanent rage, therefore so is Woke, and then there is Woke inversion and contradiction. ‘Anti-fascists’ act like fascists because fascists *and* ‘anti-fascists’ are both Wetiko at work. Political parties act the same while claiming to be different for the same reason. Secret society and satanic rituals are attaching initiates to Wetiko and the cold, ruthless, psychopathic mentality that secures the positions of power all over the world is Wetiko. Reframing ‘training programmes’ have the same cumulative effect of attaching Wetiko and we have their graduates described as automatons and robots with a cold, psychopathic, uncaring demeanour. They are all traits of Wetiko possession and look how many times they have been described in this book and elsewhere with regard to personnel behind ‘Covid’ including the police and medical profession. Climbing the greasy pole in any profession in a Wetiko society requires traits of Wetiko to get there and that is particularly true of politics which is not about fair competition and pre-eminence of ideas. It is founded on how many backs you can stab and arses you can lick. This culminated in the global ‘Covid’ coordination between the Wetiko possessed who pulled it off in all the different countries without a trace of empathy and compassion for their impact on humans. Our sight sense can see only holographic form and not the Field which connects holographic form. Therefore we perceive ‘physical’ objects with ‘space’ in between. In fact that ‘space’ is energy/consciousness operating on multiple frequencies. One of them is Wetiko and that connects the Cult psychopaths, those who submit to the psychopaths, and those who serve the psychopaths in the media operations of the world. Wetiko is Gates. Wetiko is the mask-wearing submissive. Wetiko is the fake journalist and ‘fact-checker’. The Wetiko Field is coordinating the whole thing. Psychopaths, gofers, media operatives, ‘anti-hate’ hate groups, ‘fact-checkers’ and submissive people work as one unit *even without human coordination* because they are attached to the *same* Field which is organising it all (Fig 22). Paul Levy is here describing how Wetiko-possessed people are drawn together and refuse to let any information breach their rigid perceptions. He was writing long before ‘Covid’, but I think you will recognise followers of the ‘Covid’ religion *oh just a little bit*:

> People who are channelling the vibratory frequency of wetiko align with each other through psychic resonance to reinforce their unspoken shared agreement so as to uphold their deranged view of reality. Once an unconscious content takes possession of certain individuals, it irresistibly draws them together by mutual attraction and knits them into groups tied together by their shared madness that can easily swell into an avalanche of insanity.
>
> A psychic epidemic is a closed system, which is to say that it is insular and not open to any new information or informing influences from the outside world which contradict its fixed, limited, and limiting perspective.

There we have the Woke mind and the ‘Covid’ mind. Compatible resonance draws the awakening together, too, which is clearly happening today.

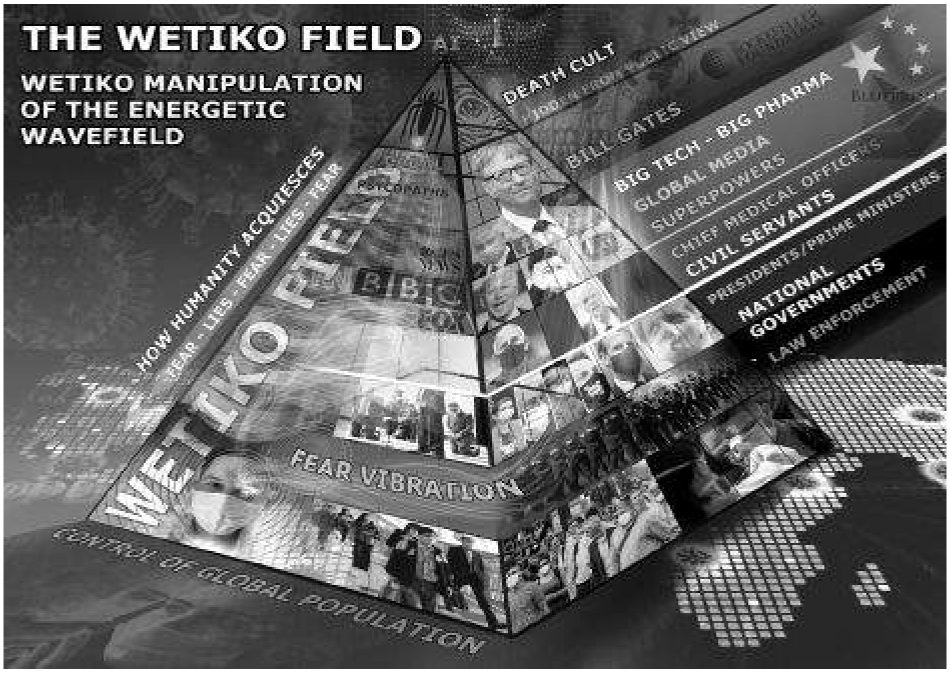

THE WETIKO FIELD
WETIKO MANIPULATION OF THE ENERGETIC WAVEFIELD
DEATH CULT
BILL GATES
BIG TECH - BIG PHARMA
GLOBAL MEDIA
SUPERPOWERS
CHIEF MEDICAL OFFICERS
CIVIL SERVANTS
PRESIDENTS/PRIME MINISTERS
NATIONAL GOVERNMENTS
LAW ENFORCEMENT
PSYCHOPATHS
HOW HUMANITY ACQUIESCES
FEAR - LIES - FEAR - LIES - FEAR
WETIKO FIELD
FEAR VIBRATION
CONTROL OF GLOBAL POPULATION

Figure 22: The Wetiko Field from which the Cult pyramid and its personnel are made manifest. (Image by Neil Hague).

Spiritual servitude

Wetiko doesn’t care about humans. It’s not human; it just possesses humans for its own ends and the effect (depending on the scale of possession) can be anything from extreme psychopathy to unquestioning obedience. Wetiko's worst nightmare is for human consciousness to expand beyond the simulation. Everything is focussed on stopping that happening through control of information, thus perception, thus frequency. The 'education system', media, science, medicine, academia, are all geared to maintaining humanity in five-sense servitude as is the constant stimulation of low-vibrational mental and emotional states (see 'Covid'). Wetiko seeks to dominate those subconscious spaces between five-sense perception and expanded consciousness where the computer meets the operator. From these subconscious hiding places Wetiko speaks to us to trigger urges and desires that we take to be our own and manipulate us into anything from low-vibrational to psychopathic states. Remember how Islam describes the Jinn as invisible tricksters that 'whisper' and confuse. Wetiko is the origin of the 'trickster god' theme that you find in cultures all over the world. Jinn, like the Archons, are Wetiko which is terrified of humans awakening and reconnecting with our true self for then its energy source has gone. With that the feedback loop breaks between Wetiko and human perception that provides the energetic momentum on which its very existence depends as a force of evil. Humans are both its target and its source of survival, but only if we are operating in low-vibrational states of fear, hate, depression and the background anxiety that most people suffer. We are Wetiko's target because we are its key to survival. It needs us, not the other way round. Paul Levy writes:

> A vampire has no intrinsic, independent, substantial existence in its own right; it only exists in relation to us. The pathogenic, vampiric mind-parasite called wetiko is nothing in itself – not being able to exist from its own side – yet it has a 'virtual reality' such that it can potentially destroy our species …
>
> …The fact that a vampire is not reflected by a mirror can also mean that what we need to see is that there's nothing, no-thing to see, other than ourselves. The fact that wetiko is the expression of something inside of us means that the cure for wetiko is with us as well. The critical issue is finding this cure within us and then putting it into effect.

Evil begets evil because if evil does not constantly expand and find new sources of energetic sustenance its evil, its *distortion*, dies with the assimilation into balance and harmony. Love is the garlic to Wetiko's vampire. Evil, the absence of love, cannot exist in the presence of love. I think I see a way out of here. I have emphasised so many times over the decades that the Archons/Wetiko and their Cult are not all powerful. *They are not*. I don't care how it looks even now *they are not*. I have not called them little boys in short trousers for effect. I have said it because it is true. Wetiko's insatiable desire for power over others is not a sign of its omnipotence, but its insecurity. Paul Levy writes: 'Due to the primal fear which ultimately drives it and which it is driven to cultivate, wetiko's body politic has an intrinsic and insistent need for centralising power and control so as to create imagined safety for itself.' *Yeeeeeees!* Exactly! Why does Wetiko want humans in an ongoing state of fear? Wetiko itself *is* fear and it is petrified of love. As evil is an absence of love, so love is an absence of fear. Love conquers all and *especially* Wetiko which *is* fear. Wetiko brought fear into the world when it wasn't here before. *Fear* was the 'fall', the fall into low-frequency ignorance and illusion – fear is **F**alse **E**motion **A**ppearing **R**eal. The simulation is driven and energised by fear because Wetiko/Yaldabaoth (fear) *are* the simulation. Fear is the absence of love and Wetiko is the absence of love.

Wetiko today

We can now view current events from this level of perspective. The 'Covid' hoax has generated momentous amounts of ongoing fear, anxiety, depression and despair which have empowered Wetiko. No wonder people like Gates have been the instigators when they are Wetiko incarnate and exhibit every trait of Wetiko in the extreme. See how cold and unemotional these people are like Gates and his cronies, how dead of eye they are. That's Wetiko. Sabbatians are Wetiko and everything they control including the World Health Organization, Big Pharma and the 'vaccine' makers, national 'health' hierarchies, corporate media, Silicon Valley, the banking system, and the United Nations with its planned transformation into world government. All are controlled and possessed by the Wetiko distortion into distorting human society in its image. We are with this knowledge at the gateway to understanding the world. Divisions of race, culture, creed and sexuality are diversions to hide the real division between those possessed and influenced by Wetiko and those that are not. The 'Covid' hoax has brought both clearly into view. Human behaviour is not about race. Tyrants and dictatorships come in all colours and creeds. What unites the US president bombing the innocent and an African tribe committing genocide against another as in Rwanda? What unites them? *Wetiko*. All wars are Wetiko, all genocide is Wetiko, all hunger over centuries in a world of plenty is Wetiko. Children going to bed hungry, including in the West, is Wetiko. Cult-generated Woke racial divisions that focus on the body are designed to obscure the reality that divisions in behaviour are manifestations of mind, not body. Obsession with body identity and group judgement is a means to divert attention from the real source of behaviour – mind and perception. Conflict sown by the Woke both within themselves and with their target groups are Wetiko providing lunch for itself through still more agents of the division, chaos, and fear on which it feeds. The Cult is seeking to assimilate the entirety of humanity and all children and young people into the Wetiko frequency by manipulating them into states of fear and despair. Witness all the suicide and psychological unravelling since the spring of 2020. Wetiko psychopaths want to impose a state of unquestioning obedience to authority which is no more than a conduit for Wetiko to enforce its will and assimilate humanity into itself. It needs us to believe that resistance is futile when it fears resistance and even more so the game-changing non-cooperation with its impositions. It can use violent resistance for its benefit. Violent impositions and violent resistance are *both* Wetiko. The Power of Love with its Power of No will sweep Wetiko from our world. Wetiko and its Cult know that. They just don't want us to know.

AI Wetiko

This brings me to AI or artificial intelligence and something else Wetikos don't want us to know. What is AI *really*? I know about computer code algorithms and AI that learns from data input. These, however, are more diversions, the expeditionary force, for the real AI that they want to connect to the human brain as promoted by Silicon Valley Wetikos like Kurzweil. What is this AI? It is the frequency of *Wetiko*, the frequency of the Archons. The connection of AI to the human brain is the connection of the Wetiko frequency to create a Wetiko hive mind and complete the job of assimilation. The hive mind is planned to be controlled from Israel and China which are both 100 percent owned by Wetiko Sabbatians. The assimilation process has been going on minute by minute in the 'smart' era which fused with the 'Covid' era. We are told that social media is scrambling the minds of the young and changing their personality. This is true, but what is social media? Look more deeply at how it works, how it creates divisions and conflict, the hostility and cruelty, the targeting of people until they are destroyed. That's Wetiko. Social media is manipulated to tune people to the Wetiko frequency with all the emotional exploitation tricks employed by platforms like Facebook and its Wetiko front man, Zuckerberg. Facebook's Instagram announced a new platform for children to overcome a legal bar on them using the main site. This is more Wetiko exploitation and manipulation of kids. Amnesty International likened the plan to foxes offering to guard the henhouse and said it was incompatible with human rights. Since when did Wetiko or Zuckerberg (I repeat myself) care about that? Would Brin and Page at Google, Wojcicki at YouTube, Bezos at Amazon and whoever the hell runs Twitter act as they do if they were not channelling Wetiko? Would those who are developing technologies for no other reason than human control? How about those designing and selling technologies to kill people and Big Pharma drug and 'vaccine' producers who know they will end or devastate lives? Quite a thought for these people to consider is that if you are Wetiko in a human life you are Wetiko on the 'other side' unless your frequency changes and that can only change by a change of perception which becomes a change of behaviour. Where Gates is going does not bear thinking about although perhaps that's exactly where he wants to go. Either way, that's where he's going. His frequency will make it so.

The frequency lair

I have been saying for a long time that a big part of the addiction to smartphones and devices is that a frequency is coming off them that entraps the mind. People spend ages on their phones and sometimes even a minute or so after they put them down they pick them up again and it all repeats. 'Covid' lockdowns will have increased this addiction a million times for obvious reasons. Addictions to alcohol overindulgence and drugs are another way that Wetiko entraps consciousness to attach to its own. Both are symptoms of low-vibrational psychological distress which alcoholism and drug addiction further compound. Do we think it's really a coincidence that access to them is made so easy while potions that can take people into realms beyond the simulation are banned and illegal? I have explored smartphone addiction in other books, the scale is mind-blowing, and that level of addiction does not come without help. Tech companies that make these phones are Wetiko and they will have no qualms about destroying the minds of children. We are seeing again with these companies the Wetiko perceptual combination of psychopathic enforcers and weak and meek unquestioning compliance by the rank and file.

The global Smart Grid is the Wetiko Grid and it is crucial to complete the Cult endgame. The simulation is radiation and we are being deluged with technological radiation on a devastating scale. Wetiko frauds like Elon Musk serve Cult interests while occasionally criticising them to maintain his street-cred. 5G and other forms of Wi-Fi are being directed at the earth from space on a volume and scale that goes on increasing by the day. Elon Musk's (officially) SpaceX Starlink project is in the process of putting tens of thousands of satellites in low orbit to cover every inch of the planet with 5G and other Wi-Fi to create Kurzweil's global 'cloud' to which the human mind is planned to be attached very soon. SpaceX has approval to operate 12,000 satellites with more than 1,300 launched at the time of writing and applications filed for *30,000* more. Other operators in the Wi-Fi, 5G, low-orbit satellite market include OneWeb (UK), Telesat (Canada), and AST & Science (US). Musk tells us that AI could be the end of humanity and then launches a company called Neuralink to connect the human brain to computers. Musk's (in theory) Tesla company is building electric cars and the driverless vehicles of the smart control grid. As frauds and bullshitters go Elon Musk in my opinion is Major League.

5G and technological radiation in general are destructive to human health, genetics and psychology and increasing the strength of artificial radiation underpins the five-sense perceptual bubbles which are themselves expressions of radiation or electromagnetism. Freedom activist John Whitehead was so right with his 'databit by databit, we are building our own electronic concentration camps'. The Smart Grid and 5G is a means to control the human mind and infuse perceptual information into The Field to influence anyone in sync with its frequency. You can change perception and behaviour en masse if you can manipulate the population into those levels of frequency and this is happening all around us today. The arrogance of Musk and his fellow Cult operatives knows no bounds in the way that we see with Gates. Musk's satellites are so many in number already they are changing the night sky when viewed from Earth. The astronomy community has complained about this and they have seen nothing yet. Some consequences of Musk's Wetiko hubris include: Radiation; visible pollution of the night sky; interference with astronomy and meteorology; ground and water pollution from intensive use of increasingly many spaceports; accumulating space debris; continual deorbiting and burning up of aging satellites, polluting the atmosphere with toxic dust and smoke; and ever-increasing likelihood of collisions. A collective public open letter of complaint to Musk said:

> We are writing to you … because SpaceX is in process of surrounding the Earth with a network of thousands of satellites whose very purpose is to irradiate every square inch of the

> Earth. SpaceX, like everyone else, is treating the radiation as if it were not there. As if the mitochondria in our cells do not depend on electrons moving undisturbed from the food we digest to the oxygen we breathe.
>
> As if our nervous systems and our hearts are not subject to radio frequency interference like any piece of electronic equipment. As if the cancer, diabetes, and heart disease that now afflict a majority of the Earth's population are not metabolic diseases that result from interference with our cellular machinery. As if insects everywhere, and the birds and animals that eat them, are not starving to death as a result.

People like Musk and Gates believe in their limitless Wetiko arrogance that they can do whatever they like to the world because they own it. Consequences for humanity are irrelevant. It's absolutely time that we stopped taking this shit from these self-styled masters of the Earth when you consider where this is going.

Why is the Cult so anti-human?

I hear this question often: Why would they do this when it will affect them, too? Ah, but will it? Who is this *them*? Forget their bodies. They are just vehicles for Wetiko consciousness. When you break it all down to the foundations we are looking at a state of severely distorted consciousness targeting another state of consciousness for assimilation. The rest is detail. The simulation is the fly-trap in which unique sensations of the five senses create a cycle of addiction called reincarnation. Renegade Minds see that everything which happens in our reality is a smaller version of the whole picture in line with the holographic principle. Addiction to the radiation of smart technology is a smaller version of addiction to the whole simulation. Connecting the body/brain to AI is taking that addiction on a giant step further to total ongoing control by assimilating human incarnate consciousness into Wetiko. I have watched during the 'Covid' hoax how many are becoming ever more profoundly attached to Wetiko's perceptual calling cards of aggressive response to any other point of view ('There is no other god but me'), psychopathic lack of compassion and empathy, and servile submission to the narrative and will of authority. Wetiko is the psychopaths *and* subservience to psychopaths. The Cult of Wetiko is so anti-human because it is *not* human. It embarked on a mission to destroy human by targeting everything that it means to be human and to survive as human. 'Covid' is not the end, just a means to an end. The Cult with its Wetiko consciousness is seeking to change Earth systems, including the atmosphere, to suit them, not humans. The gathering bombardment of 5G alone from ground and space is dramatically changing The Field with which the five senses interact. There is so much more to come if we sit on our hands and hope it will all go away. It is not meant to go away. It is meant to get ever more extreme and we need to face that while we still can – just.

Carbon dioxide is the gas of life. Without that human is over. Kaput, gone, history. No natural world, no human. The Cult has created a cock and bull story about carbon dioxide and climate change to justify its reduction to the point where Gates and the ignoramus Biden 'climate chief' John Kerry want to suck it out of the atmosphere. Kerry wants to do this because his master Gates does. Wetikos have made the gas of life a demon with the usual support from the Wokers of Extinction Rebellion and similar organisations and the bewildered puppet-child that is Greta Thunberg who was put on the world stage by Klaus Schwab and the World Economic Forum. The name Extinction Rebellion is both ironic and as always Wetiko inversion. The gas that we need to survive must be reduced to save us from extinction. The most basic need of human is oxygen and we now have billions walking around in face nappies depriving body and brain of this essential requirement of human existence. More than that 5G at 60 gigahertz interacts with the oxygen molecule to reduce the amount of oxygen the body can absorb into the bloodstream. The obvious knock-on consequences of that for respiratory and cognitive problems and life itself need no further explanation. Psychopaths like Musk are assembling a global system of satellites to deluge the human atmosphere with this insanity. The man should be in jail. Here we have two most basic of human needs, oxygen and carbon dioxide, being dismantled.

Two others, water and food, are getting similar treatment with the United Nations Agendas 21 and 2030 – the Great Reset – planning to centrally control all water and food supplies. People will not even own rain water that falls on their land. Food is affected at the most basic level by reducing carbon dioxide. We have genetic modification or GMO infiltrating the food chain on a mass scale, pesticides and herbicides polluting the air and destroying the soil. Freshwater fish that provide livelihoods for 60 million people and feed hundreds of millions worldwide are being 'pushed to the brink' according the conservationists while climate change is the only focus. Now we have Gates and Schwab wanting to dispense with current food sources all together and replace them with a synthetic version which the Wetiko Cult would control in terms of production and who eats and who doesn't. We have been on the Totalitarian Tiptoe to this for more than 60 years as food has become ever more processed and full of chemical shite to the point today when it's not natural food at all. As Dr Tom Cowan says: 'If it has a label don't eat it.' Bill Gates is now the biggest owner of farmland in the United States and he does nothing without an ulterior motive involving the Cult. Klaus Schwab wrote: 'To feed the world in the next 50 years we will need to produce as much food as was produced in the last 10,000 years … food security will only be achieved, however, if regulations on genetically modified foods are adapted to reflect the reality that gene editing offers a precise, efficient and safe method of improving crops.' Liar. People and the world are being targeted with aluminium through vaccines, chemtrails, food, drink cans, and endless other sources when aluminium has been linked to many health issues including dementia which is increasing year after year. Insects, bees and wildlife essential to the food chain are being deleted by pesticides, herbicides and radiation which 5G is dramatically increasing with 6G and 7G to come. The pollinating bee population is being devastated while wildlife including birds, dolphins and whales are having their natural radar blocked by the effects of ever-increasing radiation. In the summer windscreens used to be splattered with insects so numerous were they. It doesn't happen now. Where have they gone?

Synthetic everything

The Cult is introducing genetically-modified versions of trees, plants and insects including a Gates-funded project to unleash hundreds of millions of genetically-modified, lab-altered and patented male mosquitoes to mate with wild mosquitoes and induce genetic flaws that cause them to die out. Clinically-insane Gates-funded Japanese researchers have developed mosquitos that spread vaccine and are dubbed 'flying vaccinators'. Gates is funding the modification of weather patterns in part to sell the myth that this is caused by carbon dioxide and he's funding geoengineering of the skies to change the atmosphere. Some of this came to light with the Gates-backed plan to release tonnes of chalk into the atmosphere to 'deflect the Sun and cool the planet'. Funny how they do this while the heating effect of the Sun is not factored into climate projections focussed on carbon dioxide. The reason is that they want to reduce carbon dioxide (so don't mention the Sun), but at the same time they do want to reduce the impact of the Sun which is so essential to human life and health. I have mentioned the sun-cholesterol-vitamin D connection as they demonise the Sun with warnings about skin cancer (caused by the chemicals in sun cream they tell you to splash on). They come from the other end of the process with statin drugs to reduce cholesterol that turns sunlight into vitamin D. A lack of vitamin D leads to a long list of health effects and how vitamin D levels must have fallen with people confined to their homes over 'Covid'. Gates is funding other forms of geoengineering and most importantly chemtrails which are dropping heavy metals, aluminium and self-replicating nanotechnology onto the Earth which is killing the natural world. See *Everything You Need To Know, But Have Never Been Told* for the detailed background to this.

Every human system is being targeted for deletion by a force that's not human. The Wetiko Cult has embarked on the process of transforming the human body from biological to synthetic biological as I have explained. Biological is being replaced by the artificial and synthetic – Archontic 'countermimicry' – right across human society. The plan eventually is to dispense with the human body altogether and absorb human consciousness – which it wouldn't really be by then – into cyberspace (the simulation which is Wetiko/Yaldabaoth). Preparations for that are already happening if people would care to look. The alternative media rightly warns about globalism and 'the globalists', but this is far bigger than that and represents the end of the human race as we know it. The 'bad copy' of prime reality that Gnostics describe was a bad copy of harmony, wonder and beauty to start with before Wetiko/Yaldabaoth set out to change the simulated 'copy' into something very different. The process was slow to start with. Entrapped humans in the simulation timeline were not technologically aware and they had to be brought up to intellectual speed while being suppressed spiritually to the point where they could build their own prison while having no idea they were doing so. We have now reached that stage where technological intellect has the potential to destroy us and that's why events are moving so fast. Central American shaman Don Juan Matus said:

> Think for a moment, and tell me how you would explain the contradictions between the intelligence of man the engineer and the stupidity of his systems of belief, or the stupidity of his contradictory behaviour. Sorcerers believe that the predators have given us our systems of beliefs, our ideas of good and evil; our social mores. They are the ones who set up our dreams of success or failure. They have given us covetousness, greed, and cowardice. It is the predator who makes us complacent, routinary, and egomaniacal.
>
> In order to keep us obedient and meek and weak, the predators engaged themselves in a stupendous manoeuvre – stupendous, of course, from the point of view of a fighting strategist; a horrendous manoeuvre from the point of those who suffer it. They gave us their mind. The predators' mind is baroque, contradictory, morose, filled with the fear of being discovered any minute now.

For 'predators' see Wetiko, Archons, Yaldabaoth, Jinn, and all the other versions of the same phenomenon in cultures and religions all over the world. The theme is always the same because it's true and it's real. We have reached the point where we have to deal with it. The question is – how?

Don't fight – walk away

I thought I'd use a controversial subheading to get things moving in terms of our response to global fascism. What do you mean 'don't fight'? What do you mean 'walk away'? We've got to fight. We can't walk away. Well, it depends what we mean by fight and walk away. If fighting means physical combat we are playing Wetiko's game and falling for its trap. It wants us to get angry, aggressive, and direct hate and hostility at the enemy we think we must fight. Every war, every battle, every conflict, has been fought with Wetiko leading both sides. It's what it does. Wetiko wants a fight, anywhere, any place. Just hit me, son, so I can hit you back. Wetiko hits Wetiko and Wetiko hits Wetiko in return. I am very forthright as you can see in exposing Wetikos of the Cult, but I don't hate them. I refuse to hate them. It's what they want. What you hate you become. What you *fight* you become. Wokers, 'anti-haters' and 'anti-fascists' prove this every time they reach for their keyboards or don their balaclavas. By walk away I mean to disengage from Wetiko which includes ceasing to cooperate with its tyranny. Paul Levy says of Wetiko:

> The way to 'defeat' evil is not to try to destroy it (for then, in playing evil's game, we have already lost), but rather, to find the invulnerable place within ourselves where evil is unable to vanquish us – this is to truly 'win' our battle with evil.

Wetiko is everywhere in human society and it's been on steroids since the 'Covid' hoax. Every shouting match over wearing masks has Wetiko wearing a mask and Wetiko not wearing one. It's an electrical circuit of push and resist, push and resist, with Wetiko pushing *and* resisting. Each polarity is Wetiko empowering itself. Dictionary definitions of 'resist' include 'opposing, refusing to accept or comply with' and the word to focus on is 'opposing'. What form does this take – setting police cars alight or 'refusing to accept or comply with'? The former is Wetiko opposing Wetiko while the other points the way forward. This is the difference between those aggressively demanding that government fascism must be obeyed who stand in stark contrast to the great majority of Pushbackers. We saw this clearly with a march by thousands of Pushbackers against lockdown in London followed days later by a Woker-hijacked protest in Bristol in which police cars were set on fire. Masks were virtually absent in London and widespread in Bristol. Wetiko wants lockdown on every level of society and infuses its aggression to police it through its unknowing stooges. Lockdown protesters are the ones with the smiling faces and the hugs, The two blatantly obvious states of being – getting more obvious by the day – are the result of Wokers and their like becoming ever more influenced by the simulation Field of Wetiko and Pushbackers ever more influenced by The Field of a far higher vibration beyond the simulation. Wetiko can't invade the heart which is where most lockdown opponents are coming from. It's the heart that allows them to see through the lies to the truth in ways I will be highlighting.

Renegade Minds know that calmness is the place from which wisdom comes. You won't find wisdom in a hissing fit and wisdom is what we need in abundance right now. Calmness is not weakness – you don't have to scream at the top of your voice to be strong. Calmness is indeed a sign of strength. 'No' means I'm not doing it. *NOOOO!!!* doesn't mean you're not doing it even more. Volume does not advance 'No – I'm not doing it'. You are just not doing it. Wetiko possessed and influenced don't know how to deal with that. Wetiko wants a fight and we should not give it one. What it needs more than anything is our *cooperation* and we should not give that either. Mass rallies and marches are great in that they are a visual representation of feeling, but if it ends there they are irrelevant. You demand that Wetikos act differently? Well, they're not going to are they? They are Wetikos. We don't need to waste our time demanding that something doesn't happen when that will make no difference. We need to delete the means that *allows* it to happen. This, invariably, is our cooperation. You can demand a child stop firing a peashooter at the dog or you can refuse to buy the peashooter. If you provide the means you are cooperating with the dog being smacked on the nose with a pea. How can the authorities enforce mask-wearing if millions in a country refuse? What if the 74 million Pushbackers that voted for Trump in 2020 refused to wear masks, close their businesses or stay in their homes. It would be unenforceable. The few control the many through the compliance of the many and that's always been the dynamic be it 'Covid' regulations or the Roman Empire. I know people can find it intimidating to say no to authority or stand out in a crowd for being the only one with a face on display; but it has to be done or it's over. I hope I've made clear in this book that where this is going will be far more intimidating than standing up now and saying 'No' – I will not cooperate with my own enslavement and that of my children. There might be consequences for some initially, although not so if enough do the same. The question that must be addressed is what is going to happen if we don't? It is time to be strong and unyieldingly so. No means no. Not here and there, but *everywhere* and *always*. I have refused to wear a mask and obey all the other nonsense. I will not comply with tyranny. I repeat: Fascism is not imposed by fascists – there are never enough of them. Fascism is imposed by the population acquiescing to fascism. *I will not do it*. I will die first, or my body will. Living meekly under fascism is a form of death anyway, the death of the spirit that Martin Luther King described.

Making things happen

We must not despair. This is not over till it's over and it's far from that. The 'fat lady' must refuse to sing. The longer the 'Covid' hoax has dragged on and impacted on more lives we have seen an awakening of phenomenal numbers of people worldwide to the realisation that what they have believed all their lives is not how the world really is. Research published by the system-serving University of Bristol and King's College London in February, 2021, concluded: 'One in every 11 people in Britain say they trust David Icke's take on the coronavirus pandemic.' It will be more by now and we have gathering numbers to build on. We must urgently progress from seeing the scam to ceasing to cooperate with it. Prominent German lawyer Reiner Fuellmich, also licenced to practice law in America, is doing a magnificent job taking the legal route to bring the psychopaths to justice through a second Nuremberg tribunal for crimes against humanity. Fuellmich has an impressive record of beating the elite in court and he formed the German Corona Investigative Committee to pursue civil charges against the main perpetrators with a view to triggering criminal charges. Most importantly he has grasped the foundation of the hoax – the PCR test not testing for the 'virus' – and Christian Drosten is therefore on his charge sheet along with Gates frontman Tedros at the World Health Organization. Major players must be not be allowed to inflict their horrors on the human race without being brought to book. A life sentence must follow for Bill Gates and the rest of them. A group of researchers has also indicted the government of Norway for crimes against humanity with copies sent to the police and the International Criminal Court. The lawsuit cites participation in an internationally-planned false pandemic and violation of international law and human rights, the European Commission's definition of human rights by coercive rules, Nuremberg and Hague rules on fundamental human rights, and the Norwegian constitution. We must take the initiative from hereon and not just complain, protest and react.

There are practical ways to support vital mass non-cooperation. Organising in numbers is one. Lockdown marches in London in the spring in 2021 were mass non-cooperation that the authorities could not stop. There were too many people. Hundreds of thousands walked the London streets in the centre of the road for mile after mile while the Face-Nappies could only look on. They were determined, but calm, and just *did it* with no histrionics and lots of smiles. The police were impotent. Others are organising group shopping without masks for mutual support and imagine if that was happening all over. Policing it would be impossible. If the store refuses to serve people in these circumstances they would be faced with a long line of trolleys full of goods standing on their own and everything would have to be returned to the shelves. How would they cope with that if it kept happening? I am talking here about moving on from complaining to being pro-active; from watching things happen to making things happen. I include in this our relationship with the police. The behaviour of many Face-Nappies has been disgraceful and anyone who thinks they would never find concentration camp guards in the 'enlightened' modern era have had that myth busted big-time. The period and setting may change – Wetikos never do. I watched film footage from a London march in which a police thug viciously kicked a protestor on the floor who had done nothing. His fellow Face-Nappies stood in a ring protecting him. What he did was a criminal assault and with a crowd far outnumbering the police this can no longer be allowed to happen unchallenged. I get it when people chant 'shame on you' in these circumstances, but that is no longer enough. They *have* no shame those who do this. Crowds needs to start making a citizen's arrest of the police who commit criminal offences and brutally attack innocent people and defenceless women. A citizen's arrest can be made under section 24A of the UK Police and Criminal Evidence (PACE) Act of 1984 and you will find something similar in other countries. I prefer to call it a Common Law arrest rather than citizen's for reasons I will come to shortly. Anyone can arrest a person committing an indictable offence or if they have reasonable grounds to suspect they are committing an indictable offence. On both counts the attack by the police thug would have fallen into this category. A citizen's arrest can be made to stop someone:

- Causing physical injury to himself or any other person
- Suffering physical injury
- Causing loss of or damage to property
- Making off before a constable can assume responsibility for him

A citizen's arrest may also be made to prevent a breach of the peace under Common Law and if they believe a breach of the peace will happen or anything related to harm likely to be done or already done in their presence. This is the way to go I think – the Common Law version. If police know that the crowd and members of the public will no longer be standing and watching while they commit their thuggery and crimes they will think twice about acting like Brownshirts and Blackshirts.

Common Law – common sense

Mention of Common Law is very important. Most people think the law is the law as in one law. This is not the case. There are two bodies of law, Common Law and Statute Law, and they are not the same. Common Law is founded on the simple premise of do no harm. It does not recognise victimless crimes in which no harm is done while Statute Law does. There is a Statute Law against almost everything. So what is Statute Law? Amazingly it's the law of the *sea* that was brought ashore by the Cult to override the law of the land which is Common Law. They had no right to do this and as always they did it anyway. They had to. They could not impose their will on the people through Common Law which only applies to do no harm. How could you stitch up the fine detail of people's lives with that? Instead they took the law of the sea, or Admiralty Law, and applied it to the population. Statute Law refers to all the laws spewing out of governments and their agencies including all the fascist laws and regulations relating to 'Covid'. The key point to make is that Statute Law is *contract law*. It only applies between *contracting* corporations. Most police officers don't even know this. They have to be kept in the dark, too. Long ago when merchants and their sailing ships began to trade with different countries a contractual law was developed called Admiralty Law and other names. Again it only applied to *contracts* agreed between *corporate* entities. If there is no agreed contract the law of the sea had no jurisdiction *and that still applies to its new alias of Statute Law*. The problem for the Cult when the law of the sea was brought ashore was an obvious one. People were not corporations and neither were government entities. To overcome the latter they made governments and all associated organisations corporations. All the institutions are *private corporations* and I mean governments and their agencies, local councils, police, courts, military, US states, the whole lot. Go to the

Dun and Bradstreet corporate listings website for confirmation that they are all corporations. You are arrested by a private corporation called the police by someone who is really a private security guard and they take you to court which is another private corporation. Neither have jurisdiction over you unless you consent and *contract* with them. This is why you hear the mantra about law enforcement policing by *consent* of the people. In truth the people 'consent' only in theory through monumental trickery.

Okay, the Cult overcame the corporate law problem by making governments and institutions corporate entities; but what about people? They are not corporations are they? Ah ... well in a sense, and *only* a sense, they are. Not people exactly – the illusion of people. The Cult creates a corporation in the name of everyone at the time that their birth certificate is issued. Note birth/ *berth* certificate and when you go to court under the law of the sea on land you stand in a *dock*. These are throwbacks to the origin. My Common Law name is David Vaughan Icke. The name of the corporation created by the government when I was born is called Mr David Vaughan Icke usually written in capitals as MR DAVID VAUGHAN ICKE. That is not me, the living, breathing man. It is a fictitious corporate entity. The trick is to make you think that David Vaughan Icke and MR DAVID VAUGHAN ICKE are the same thing. *They are not*. When police charge you and take you to court they are prosecuting the corporate entity and not the living, breathing, man or woman. They have to trick you into identifying as the corporate entity and contracting with them. Otherwise they have no jurisdiction. They do this through a language known as legalese. Lawful and legal are not the same either. Lawful relates to Common Law and legal relates to Statute Law. Legalese is the language of Statue Law which uses terms that mean one thing to the public and another in legalese. Notice that when a police officer tells someone why they are being charged he or she will say at the end: 'Do you understand?' To the public that means 'Do you comprehend?' In legalese it means 'Do you stand under me?' Do you stand under my authority? If you say yes to the question you are unknowingly agreeing to give them jurisdiction over you in a contract between two corporate entities.

This is a confidence trick in every way. Contracts have to be agreed between informed parties and if you don't know that David Vaughan Icke is agreeing to be the corporation MR DAVID VAUGHAN ICKE you cannot knowingly agree to contract. They are deceiving you and another way they do this is to ask for proof of identity. You usually show them a driving licence or other document on which your corporate name is written. In doing so you are accepting that you are that corporate entity when you are not. Referring to yourself as a 'person' or 'citizen' is also identifying with your corporate fiction which is why I made the Common Law point about the citizen's arrest. If you are approached by a police officer you identify yourself immediately as a living, breathing, man or woman and say 'I do not consent, I do not contract with you and I do not understand' or stand under their authority. I have a Common Law birth certificate as a living man and these are available at no charge from commonlawcourt.com. Businesses registered under the Statute Law system means that its laws apply. There are, however, ways to run a business under Common Law. Remember all 'Covid' laws and regulations are Statute Law – the law of *contracts* and you do not have to contract. This doesn't mean that you can kill someone and get away with it. Common Law says do no harm and that applies to physical harm, financial harm etc. Police are employees of private corporations and there needs to be a new system of non-corporate Common Law constables operating outside the Statute Law system. If you go to davidicke.com and put Common Law into the search engine you will find videos that explain Common Law in much greater detail. It is definitely a road we should walk.

With all my heart

I have heard people say that we are in a spiritual war. I don't like the term 'war' with its Wetiko dynamic, but I know what they mean. Sweep aside all the bodily forms and we are in a situation in which two states of consciousness are seeking very different realities.

Wetiko wants upheaval, chaos, fear, suffering, conflict and control. The other wants love, peace, harmony, fairness and freedom. That's where we are. We should not fall for the idea that Wetiko is all-powerful and there's nothing we can do. Wetiko is not all-powerful. It's a joke, pathetic. It doesn't have to be, but it has made that choice for now. A handful of times over the years when I have felt the presence of its frequency I have allowed it to attach briefly so I could consciously observe its nature. The experience is not pleasant, the energy is heavy and dark, but the ease with which you can kick it back out the door shows that its real power is in persuading us that it has power. It's all a con. Wetiko is a con. It's a trickster and not a power that can control us if we unleash our own. The con is founded on manipulating humanity to give its power to Wetiko which recycles it back to present the illusion that it has power when its power is *ours* that we gave away. This happens on an energetic level and plays out in the world of the seen as humanity giving its power to Wetiko authority which uses that power to control the population when the power is only the power the population has handed over. How could it be any other way for billions to be controlled by a relative few? I have had experiences with people possessed by Wetiko and again you can kick its arse if you do it with an open heart. Oh yes – the *heart* which can transform the world of perceived 'matter'.

We are receiver-transmitters and processors of information, but what information and where from? Information is processed into perception in three main areas – the brain, the heart and the belly. These relate to thinking, knowing, and emotion. Wetiko wants us to be head and belly people which means we think within the confines of the Matrix simulation and low-vibrational emotional reaction scrambles balance and perception. A few minutes on social media and you see how emotion is the dominant force. Woke is all emotion and is therefore thought-free and fact-free. Our heart is something different. It *knows* while the head *thinks* and has to try to work it out because it doesn't know. The human energy field has seven prime vortexes which connect us with wider reality (Fig 23). Chakra means ‘wheels of light’ in the Sanskrit language of ancient India. The main ones are: The crown chakra on top of the head; brow (or ‘third eye’) chakra in the centre of the forehead; throat chakra; heart chakra in the centre of the chest; solar plexus chakra below the sternum; sacral chakra beneath the navel; and base chakra at the bottom of the spine. Each one has a particular function or functions. We feel anxiety and nervousness in the belly where the sacral chakra is located and this processes emotion that can affect the colon to give people ‘the shits’ or make them ‘shit scared’ when they are nervous. Chakras all play an important role, but the Mr and Mrs Big is the heart chakra which sits at the centre of the seven, above the chakras that connect us to the ‘physical’ and below those that connect with higher realms (or at least should). Here in the heart chakra we feel love, empathy and compassion – ‘My heart goes out to you’. Those with closed hearts become literally ‘heart-less’ in their attitudes and behaviour (see Bill Gates). Native Americans portrayed Wetiko with what Paul Levy calls a ‘frigid, icy heart, devoid of mercy’ (see Bill Gates).

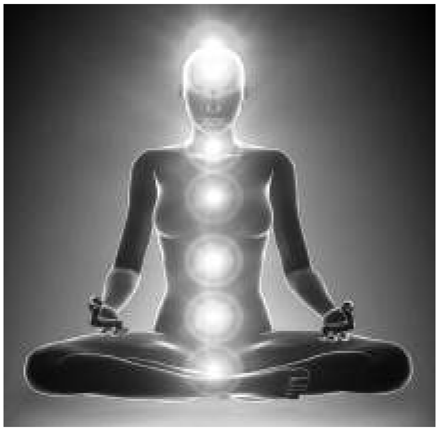


Figure 23: The chakra system which interpenetrates the human energy field. The heart chakra is the governor – or should be.

Wetiko trembles at the thought of heart energy which it cannot infiltrate. The frequency is too high. What it seeks to do instead is close the heart chakra vortex to block its perceptual and energetic influence. Psychopaths have ‘hearts of stone’ and emotionally-damaged people have ‘heartache’ and ‘broken hearts’. The astonishing amount of heart disease is related to heart chakra disruption with its fundamental connection to the 'physical' heart. Dr Tom Cowan has written an outstanding book challenging the belief that the heart is a pump and making the connection between the 'physical' and spiritual heart. Rudolph Steiner who was way ahead of his time said the same about the fallacy that the heart is a pump. *What*? The heart is not a pump? That's crazy, right? Everybody knows that. Read Cowan's *Human Heart, Cosmic Heart* and you will realise that the very idea of the heart as a pump is ridiculous when you see the evidence. How does blood in the feet so far from the heart get pumped horizontally up the body by the heart?? Cowan explains in the book the real reason why blood moves as it does. Our 'physical' heart is used to symbolise love when the source is really the heart vortex or spiritual heart which is our most powerful energetic connection to 'out there' expanded consciousness. That's why we feel *knowing* – intuitive knowing – in the centre of the chest. Knowing doesn't come from a process of thoughts leading to a conclusion. It is there in an instant all in one go. Our heart knows because of its connection to levels of awareness that *do* know. This is the meaning and source of intuition – intuitive *knowing*.

For the last more than 30 years of uncovering the global game and the nature of reality my heart has been my constant antenna for truth and accuracy. An American intelligence insider once said that I had quoted a disinformer in one of my books and yet I had only quoted the part that was true. He asked: 'How do you do that?' By using my heart antenna was the answer and anyone can do it. Heart-centred is how we are meant to be. With a closed heart chakra we withdraw into a closed mind and the bubble of five-sense reality. If you take a moment to focus your attention on the centre of your chest, picture a spinning wheel of light and see it opening and expanding. You will feel it happening, too, and perceptions of the heart like joy and love as the heart impacts on the mind as they interact. The more the chakra opens the more you will feel expressions of heart consciousness and as the process continues, and becomes part of you, insights and knowings will follow. An open heart is connected to that level of awareness that knows all is *One*. You will see from its perspective that the fault-lines that divide us are only illusions to control us. An open heart does not process the illusions of race, creed and sexuality except as brief experiences for a consciousness that is all. Our heart does not see division, only unity (Figs 24 and 25). There's something else, too. Our hearts love to laugh. Mark Twain's quote that says 'The human race has one really effective weapon, and that is laughter' is really a reference to the heart which loves to laugh with the joy of knowing the true nature of infinite reality and that all the madness of human society is an illusion of the mind. Twain also said: 'Against the assault of laughter nothing can stand.' This is so true of Wetiko and the Cult. Their insecurity demands that they be taken seriously and their power and authority acknowledged and feared. We should do nothing of the sort. We should not get aggressive or fearful which their insecurity so desires. We should laugh in their face. Even in their no-face as police come over in their face-nappies and expect to be taken seriously. They don't take themselves seriously looking like that so why should we? Laugh in the face of intimidation. Laugh in the face of tyranny. You will see by its reaction that you have pressed all of its buttons. Wetiko does not know what to do in the face of laughter or when its targets refuse to concede their joy to fear. We have seen many examples during the 'Covid' hoax when people have expressed their energetic power and the string puppets of Wetiko retreat with their tail limp between their knees. Laugh – the world is bloody mad after all and if it's a choice between laughter and tears I know which way I'm going.

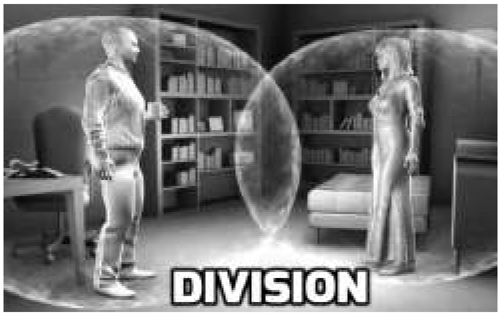


Figure 24: Head consciousness without the heart sees division and everything apart from everything else.

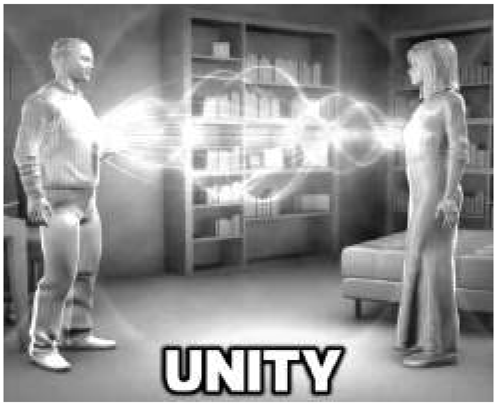


Figure 25: Heart consciousness sees everything as One.

'Vaccines' and the soul

The foundation of Wetiko/Archon control of humans is the separation of incarnate five-sense mind from the infinite 'I' and closing the heart chakra where the True 'I' lives during a human life. The goal has been to achieve complete separation in both cases. I was interested therefore to read an account by a French energetic healer of what she said she experienced with a patient who had been given the 'Covid' vaccine. Genuine energy healers can sense information and consciousness fields at different levels of being which are referred to as 'subtle bodies'. She described treating the patient who later returned after having, without the healer's knowledge, two doses of the 'Covid vaccine'. The healer said:

> I noticed immediately the change, very heavy energy emanating from [the] subtle bodies. The scariest thing was when I was working on the heart chakra, I connected with her soul: it was detached from the physical body, it had no contact and it was, as if it was floating in a state of total confusion: a damage to the consciousness that loses contact with the physical body, i.e. with our biological machine, there is no longer any communication between them.
>
> I continued the treatment by sending light to the heart chakra, the soul of the person, but it seemed that the soul could no longer receive any light, frequency or energy. It was a very powerful experience for me. Then I understood that this substance is indeed used to detach consciousness so that this consciousness can no longer interact through this body that it possesses in life, where there is no longer any contact, no frequency, no light, no more energetic balance or mind.

This would create a human that is rudderless and at the extreme almost zombie-like operating with a fractional state of consciousness at the mercy of Wetiko. I was especially intrigued by what the healer said in the light of the prediction by the highly-informed Rudolf Steiner more than a hundred years ago. He said:

> In the future, we will eliminate the soul with medicine. Under the pretext of a 'healthy point of view', there will be a vaccine by which the human body will be treated as soon as possible directly at birth, so that the human being cannot develop the thought of the existence of soul and Spirit. To materialistic doctors will be entrusted the task of removing the soul of humanity.
>
> As today, people are vaccinated against this disease or that disease, so in the future, children will be vaccinated with a substance that can be produced precisely in such a way that people, thanks to this vaccination, will be immune to being subjected to the 'madness' of spiritual life. He would be extremely smart, but he would not develop a conscience, and that is the true goal of some materialistic circles.

Steiner said the vaccine would detach the physical body from the etheric body (subtle bodies) and 'once the etheric body is detached the relationship between the universe and the etheric body would become extremely unstable, and man would become an automaton'. He said 'the physical body of man must be polished on this Earth by spiritual will – so the vaccine becomes a kind of arymanique (Wetiko) force' and 'man can no longer get rid of a given materialistic feeling'. Humans would then, he said, become 'materialistic of constitution and can no longer rise to the spiritual'. I have been writing for years about DNA being a receiver-transmitter of information that connects us to other levels of reality and these 'vaccines' changing DNA can be likened to changing an antenna and what it can transmit and receive. Such a disconnection would clearly lead to changes in personality and perception. Steiner further predicted the arrival of AI. Big Pharma 'Covid vaccine' makers, expressions of Wetiko, are testing their DNA-manipulating evil on children as I write with a view to giving the 'vaccine' to babies. If it's a soul-body disconnector – and I say that it is or can be – every child would be disconnected from 'soul' at birth and the 'vaccine' would create a closed system in which spiritual guidance from the greater self would play no part. This has been the ambition of Wetiko all along. A Pentagon video from 2005 was leaked of a presentation explaining the development of vaccines to change behaviour by their effect on the brain. Those that believe this is not happening with the 'Covid' genetically-modifying procedure masquerading as a 'vaccine' should make an urgent appointment with Naivety Anonymous. Klaus Schwab wrote in 2018:

> Neurotechnologies enable us to better influence consciousness and thought and to understand many activities of the brain. They include decoding what we are thinking in fine levels of detail through new chemicals and interventions that can influence our brains to correct for errors or enhance functionality.

The plan is clear and only the heart can stop it. With every heart that opens, every mind that awakens, Wetiko is weakened. Heart and love are far more powerful than head and hate and so nothing like a majority is needed to turn this around.

Beyond the Phantom

Our heart is the prime target of Wetiko and so it must be the answer to Wetiko. We *are* our heart which is part of one heart, the infinite heart. Our heart is where the true self lives in a human life behind firewalls of five-sense illusion when an imposter takes its place – *Phantom Self*; but our heart waits patiently to be set free any time we choose to see beyond the Phantom, beyond Wetiko. A Wetikoed Phantom Self can wreak mass death and destruction while the love of forever is locked away in its heart. The time is here to unleash its power and let it sweep away the fear and despair that is Wetiko. Heart consciousness does not seek manipulated, censored, advantage for its belief or religion, its activism and desires. As an expression of the One it treats all as One with the same rights to freedom and opinion. Our heart demands fairness for itself no more than for others. From this unity of heart we can come together in mutual support and transform this Wetikoed world into what reality is meant to be – a place of love, joy, happiness, fairness, justice and freedom. Wetiko has another agenda and that's why the world is as it is, but enough of this nonsense. Wetiko can't stay where hearts are open and it works so hard to keep them closed. Fear is its currency and its food source and love in its true sense has no fear. Why would love have fear when it knows it is *All That Is, Has Been, And Ever Can Be* on an eternal exploration of all possibility? Love in this true sense is not the physical attraction that passes for love. This can be an expression of it, yes, but Infinite Love, a love without condition, goes far deeper to the core of all being. It *is* the core of all being. Infinite realty was born from love beyond the illusions of the simulation. Love infinitely expressed is the knowing that all is One and the swiftly-passing experience of separation is a temporary hallucination. You cannot disconnect from Oneness; you can only *perceive* that you have and withdraw from its influence. This is the most important of all perception trickery by the mind parasite that is Wetiko and the foundation of all its potential for manipulation.

If we open our hearts, open the sluice gates of the mind, and redefine self-identity amazing things start to happen. Consciousness expands or contracts in accordance with self-identity. When true self is recognised as infinite awareness and label self – Phantom Self – is seen as only a series of brief experiences life is transformed. Consciousness expands to the extent that self-identity expands and everything changes. You see unity, not division, the picture, not the pixels. From this we can play the long game. No more is an experience something in and of itself, but a fleeting moment in the eternity of forever. Suddenly people in uniform and dark suits are no longer intimidating. Doing what your heart knows to be right is no longer intimidating and consequences for those actions take on the same nature of a brief experience that passes in the blink of an infinite eye. Intimidation is all in the mind. Beyond the mind there is no intimidation.

An open heart does not consider consequences for what it knows to be right. To do so would be to consider not doing what it knows to be right and for a heart in its power that is never an option. The Renegade Mind is really the Renegade Heart. Consideration of consequences will always provide a getaway car for the mind and the heart doesn't want one. What is right in the light of what we face today is to stop cooperating with Wetiko in all its forms and to do it without fear or compromise. You cannot compromise with tyranny when tyranny always demands more until it has everything. Life is your perception and you are your destiny. Change your perception and you change your life. Change collective perception and we change the world.

Come on people … One human family, One heart, One goal … *FREEEEEEDOM*!

We must settle for nothing less.

Postscript

The big scare story as the book goes to press is the 'Indian' variant and the world is being deluged with propaganda about the 'Covid catastrophe' in India which mirrors in its lies and misrepresentations what happened in Italy before the first lockdown in 2020.

The *New York Post* published a picture of someone who had 'collapsed in the street from Covid' in India in April, 2021, which was actually taken during a gas leak in May, 2020. Same old, same old. Media articles in mid-February were asking why India had been so untouched by 'Covid' and then as their vaccine rollout gathered pace the alleged 'cases' began to rapidly increase. Indian 'Covid vaccine' maker Bharat Biotech was funded into existence by the Bill and Melinda Gates Foundation (the pair announced their divorce in May, 2021, which is a pity because they so deserve each other). The Indian 'Covid crisis' was ramped up by the media to terrify the world and prepare people for submission to still more restrictions. The scam that worked the first time was being repeated only with far more people seeing through the deceit. Davidicke.com and Ickonic.com have sought to tell the true story of what is happening by talking to people living through the Indian nightmare which has nothing to do with 'Covid'. We posted a letter from 'Alisha' in Pune who told a very different story to government and media mendacity. She said scenes of dying people and overwhelmed hospitals were designed to hide what was really happening – genocide and starvation. Alisha said that millions had already died of starvation during the ongoing lockdowns while government and media were lying and making it look like the 'virus':

> Restaurants, shops, gyms, theatres, basically everything is shut. The cities are ghost towns. Even so-called 'essential' businesses are only open till 11am in the morning. You basically have just an hour to buy food and then your time is up.
>
> Inter-state travel and even inter-district travel is banned. The cops wait at all major crossroads to question why you are traveling outdoors or to fine you if you are not wearing a mask.
>
> The medical community here is also complicit in genocide, lying about hospitals being full and turning away people with genuine illnesses, who need immediate care. They have even created a shortage of oxygen cylinders.

This is the classic Cult modus operandi played out in every country. Alisha said that people who would not have a PCR test not testing for the 'virus' were being denied hospital treatment. She said the people hit hardest were migrant workers and those in rural areas. Most businesses employed migrant workers and with everything closed there were no jobs, no income and no food. As a result millions were dying of starvation or malnutrition. All this was happening under Prime Minister Narendra Modi, a 100-percent asset of the Cult, and it emphasises yet again the scale of pure anti-human evil we are dealing with. Australia banned its people from returning home from India with penalties for trying to do so of up to five years in jail and a fine of £37,000. The manufactured 'Covid' crisis in India was being prepared to justify further fascism in the West. Obvious connections could be seen between the Indian 'vaccine' programme and increased 'cases' and this became a common theme. The Seychelles, the most per capita 'Covid vaccinated' population in the world, went back into lockdown after a 'surge of cases'.

Long ago the truly evil Monsanto agricultural biotechnology corporation with its big connections to Bill Gates devastated Indian farming with genetically-modified crops. Human rights activist Gurcharan Singh highlighted the efforts by the Indian government to complete the job by destroying the food supply to hundreds of millions with 'Covid' lockdowns. He said that 415 million people at the bottom of the disgusting caste system (still going whatever they say) were below the poverty line and struggled to feed themselves every year. Now the government was imposing lockdown at just the time to destroy the harvest. This deliberate policy was leading to mass starvation. People may reel back at the suggestion that a government would do that, but Wetiko-controlled ‘leaders’ are capable of any level of evil. In fact what is described in India is in the process of being instigated worldwide. The food chain and food supply are being targeted at every level to cause world hunger and thus control. Bill Gates is not the biggest owner of farmland in America for no reason and destroying access to food aids both the depopulation agenda and the plan for synthetic ‘food’ already being funded into existence by Gates. Add to this the coming hyper-inflation from the suicidal creation of fake ‘money’ in response to ‘Covid’ and the breakdown of container shipping systems and you have a cocktail that can only lead one way and is meant to. The Cult plan is to crash the entire system to ‘build back better’ with the Great Reset.

‘Vaccine’ transmission

Reports from all over the world continue to emerge of women suffering menstrual and fertility problems after having the fake ‘vaccine’ and of the non-’vaccinated’ having similar problems when interacting with the ‘vaccinated’. There are far too many for ‘coincidence’ to be credible. We’ve had menopausal women getting periods, others having periods stop or not stopping for weeks, passing clots, sometimes the lining of the uterus, breast irregularities, and miscarriages (which increased by 400 percent in parts of the United States). Non-‘vaccinated’ men and children have suffered blood clots and nose bleeding after interaction with the ‘vaccinated’. Babies have died from the effects of breast milk from a ‘vaccinated’ mother. Awake doctors – the small minority – speculated on the cause of non-’vaccinated’ suffering the same effects as the ‘vaccinated’. Was it nanotechnology in the synthetic substance transmitting frequencies or was it a straight chemical bioweapon that was being transmitted between people? I am not saying that some kind of chemical transmission is not one possible answer, but the foundation of all that the Cult does is frequency and this is fertile ground for understanding how transmission can happen. American doctor Carrie Madej, an internal medicine physician and osteopath, has been practicing for the last 20 years, teaching medical students, and she says attending different meetings where the agenda for humanity was discussed. Madej, who operates out of Georgia, did not dismiss other possible forms of transmission, but she focused on frequency in search of an explanation for transmission. She said the Moderna and Pfizer 'vaccines' contained nano-lipid particles as a key component. This was a brand new technology never before used on humanity. 'They're using a nanotechnology which is pretty much little tiny computer bits … nanobots or hydrogel.' Inside the 'vaccines' was 'this sci-fi kind of substance' which suppressed immune checkpoints to get into the cell. I referred to this earlier as the 'Trojan horse' technique that tricks the cell into opening a gateway for the self-replicating synthetic material and while the immune system is artificially suppressed the body has no defences. Madej said the substance served many purposes including an on-demand ability to 'deliver the payload' and using the nano 'computer bits' as biosensors in the body. 'It actually has the ability to accumulate data from your body, like your breathing, your respiration, thoughts, emotions, all kinds of things.'

She said the technology obviously has the ability to operate through Wi-Fi and transmit and receive energy, messages, frequencies or impulses. 'Just imagine you're getting this new substance in you and it can react to things all around you, the 5G, your smart device, your phones.' We had something completely foreign in the human body that had never been launched large scale at a time when we were seeing 5G going into schools and hospitals (plus the Musk satellites) and she believed the 'vaccine' transmission had something to do with this: '… if these people have this inside of them … it can act like an antenna and actually transmit it outwardly as well.' The synthetic substance produced its own voltage and so it could have that kind of effect. This fits with my own contention that the nano receiver-transmitters are designed to connect people to the

Smart Grid and break the receiver-transmitter connection to expanded consciousness. That would explain the French energy healer's experience of the disconnection of body from 'soul' with those who have had the 'vaccine'. The nanobots, self-replicating inside the body, would also transmit the synthetic frequency which could be picked up through close interaction by those who have not been 'vaccinated'. Madej speculated that perhaps it was 5G and increased levels of other radiation that was causing the symptoms directly although interestingly she said that non-'vaccinated' patients had shown improvement when they were away from the 'vaccinated' person they had interacted with. It must be remembered that you can control frequency and energy with your mind and you can consciously create energetic barriers or bubbles with the mind to stop damaging frequencies from penetrating your field. American paediatrician Dr Larry Palevsky said the 'vaccine' was not a 'vaccine' and was never designed to protect from a 'viral' infection. He called it 'a massive, brilliant propaganda of genocide' because they didn't have to inject everyone to get the result they wanted. He said the content of the jabs was able to infuse any material into the brain, heart, lungs, kidneys, liver, sperm and female productive system. 'This is genocide; this is a weapon of mass destruction.' At the same time American colleges were banning students from attending if they didn't have this life-changing and potentially life-ending 'vaccine'. Class action lawsuits must follow when the consequences of this college fascism come to light. As the book was going to press came reports about fertility effects on sperm in 'vaccinated' men which would absolutely fit with what I have been saying and hospitals continued to fill with 'vaccine' reactions. Another question is what about transmission via blood transfusions? The NHS has extended blood donation restrictions from seven days after a 'Covid vaccination' to 28 days after even a sore arm reaction.

I said in the spring of 2020 that the then touted 'Covid vaccine' would be ongoing each year like the flu jab. A year later Pfizer CEO, the appalling Albert Bourla, said people would 'likely' need a 'booster dose' of the 'vaccine' within 12 months of getting 'fully vaccinated' and then a yearly shot. 'Variants will play a key role', he said confirming the point. Johnson & Johnson CEO Alex Gorsky also took time out from his 'vaccine' disaster to say that people may need to be vaccinated against 'Covid-19' each year. UK Health Secretary, the psychopath Matt Hancock, said additional 'boosters' would be available in the autumn of 2021. This is the trap of the 'vaccine passport'. The public will have to accept every last 'vaccine' they introduce, including for the fake 'variants', or it would cease to be valid. The only other way in some cases would be continuous testing with a test not testing for the 'virus' and what is on the swabs constantly pushed up your noise towards the brain every time?

'Vaccines' changing behaviour

I mentioned in the body of the book how I believed we would see gathering behaviour changes in the 'vaccinated' and I am already hearing such comments from the non-'vaccinated' describing behaviour changes in friends, loved ones and work colleagues. This will only increase as the self-replicating synthetic material and nanoparticles expand in body and brain. An article in the *Guardian* in 2016 detailed research at the University of Virginia in Charlottesville which developed a new method for controlling brain circuits associated with complex animal behaviour. The method, dubbed 'magnetogenetics', involves genetically-engineering a protein called ferritin, which stores and releases iron, to create a magnetised substance – 'Magneto' – that can activate specific groups of nerve cells from a distance. This is claimed to be an advance on other methods of brain activity manipulation known as optogenetics and chemogenetics (the Cult has been developing methods of brain control for a long time). The ferritin technique is said to be non-invasive and able to activate neurons 'rapidly and reversibly'. In other words, human thought and perception. The article said that earlier studies revealed how nerve cell proteins 'activated by heat and mechanical pressure can be genetically engineered so that they become sensitive to radio waves and magnetic fields, by attaching them to an iron-storing protein called ferritin, or to inorganic paramagnetic particles'. Sensitive to radio waves and magnetic fields? You mean like 5G, 6G and 7G? This is the human-AI Smart Grid hive mind we are talking about. The *Guardian* article said:

> … the researchers injected Magneto into the striatum of freely behaving mice, a deep brain structure containing dopamine-producing neurons that are involved in reward and motivation, and then placed the animals into an apparatus split into magnetised and non-magnetised sections.
>
> Mice expressing Magneto spent far more time in the magnetised areas than mice that did not, because activation of the protein caused the striatal neurons expressing it to release dopamine, so that the mice found being in those areas rewarding. This shows that Magneto can remotely control the firing of neurons deep within the brain, and also control complex behaviours.

Make no mistake this basic methodology will be part of the 'Covid vaccine' cocktail and using magnetics to change brain function through electromagnetic field frequency activation. The Pentagon is developing a 'Covid vaccine' using ferritin. Magnetics would explain changes in behaviour and why videos are appearing across the Internet as I write showing how magnets stick to the skin at the point of the 'vaccine' shot. Once people take these 'vaccines' anything becomes possible in terms of brain function and illness which will be blamed on 'Covid-19' and 'variants'. Magnetic field manipulation would further explain why the non-'vaccinated' are reporting the same symptoms as the 'vaccinated' they interact with and why those symptoms are reported to decrease when not in their company. Interestingly 'Magneto', a 'mutant', is a character in the Marvel Comic *X-Men* stories with the ability to manipulate magnetic fields and he believes that mutants should fight back against their human oppressors by any means necessary. The character was born Erik Lehnsherr to a Jewish family in Germany.

Cult-controlled courts

The European Court of Human Rights opened the door for mandatory 'Covid-19 vaccines' across the continent when it ruled in a Czech Republic dispute over childhood immunisation that legally enforced vaccination could be ‘necessary in a democratic society’. The 17 judges decided that compulsory vaccinations did not breach human rights law. On the face of it the judgement was so inverted you gasp for air. If not having a vaccine infused into your body is not a human right then what is? Ah, but they said human rights law which has been specifically written to delete all human rights at the behest of the state (the Cult). Article 8 of the European Convention on Human Rights relates to the right to a private life. The crucial word here is *‘except’*:

> There shall be no interference by a public authority with the exercise of this right EXCEPT such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic wellbeing of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others [My emphasis].

No interference *except* in accordance with the law means there *are* no ‘human rights’ *except* what EU governments decide you can have at their behest. ‘As is necessary in a democratic society’ explains that reference in the judgement and ‘in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others’ gives the EU a coach and horses to ride through ‘human rights’ and scatter them in all directions. The judiciary is not a check and balance on government extremism; it is a vehicle to enforce it. This judgement was almost laughably predictable when the last thing the Cult wanted was a decision that went against mandatory vaccination. Judges rule over and over again to benefit the system of which they are a part. Vaccination disputes that come before them are invariably delivered in favour of doctors and authorities representing the view of the state which owns the judiciary. Oh, yes, and we have even had calls to stop putting ‘Covid-19’ on death certificates within 28 days of a ‘positive test’ because it is claimed the practice makes the ‘vaccine’ appear not to work. They are laughing at you.

The scale of madness, inhumanity and things to come was highlighted when those not 'vaccinated' for 'Covid' were refused evacuation from the Caribbean island of St Vincent during massive volcanic eruptions. Cruise ships taking residents to the safety of another island allowed only the 'vaccinated' to board and the rest were left to their fate. Even in life and death situations like this we see 'Covid' stripping people of their most basic human instincts and the insanity is even more extreme when you think that fake 'vaccine'-makers are not even claiming their body-manipulating concoctions stop 'infection' and 'transmission' of a 'virus' that doesn't exist. St Vincent Prime Minister Ralph Gonsalves said: 'The chief medical officer will be identifying the persons already vaccinated so that we can get them on the ship.' Note again the power of the chief medical officer who, like Whitty in the UK, will be answering to the World Health Organization. This is the Cult network structure that has overridden politicians who 'follow the science' which means doing what WHO-controlled 'medical officers' and 'science advisers' tell them. Gonsalves even said that residents who were 'vaccinated' after the order so they could board the ships would still be refused entry due to possible side effects such as 'wooziness in the head'. The good news is that if they were woozy enough in the head they could qualify to be prime minister of St Vincent.

Microchipping freedom

The European judgement will be used at some point to justify moves to enforce the 'Covid' DNA-manipulating procedure. Sandra Ro, CEO of the Global Blockchain Business Council, told a World Economic Forum event that she hoped 'vaccine passports' would help to 'drive forced consent and standardisation' of global digital identity schemes: 'I'm hoping with the desire and global demand for some sort of vaccine passport – so that people can get travelling and working again – [it] will drive forced consent, standardisation, and frankly, cooperation across the world.' The lady is either not very bright, or thoroughly mendacious, to use the term 'forced consent'.

You do not 'consent' if you are forced – you *submit*. She was describing what the plan has been all along and that's to enforce a digital identity on every human without which they could not function. 'Vaccine passports' are opening the door and are far from the end goal. A digital identity would allow you to be tracked in everything you do in cyberspace and this is the same technique used by Cult-owned China to enforce its social credit system of total control. The ultimate 'passport' is planned to be a microchip as my books have warned for nearly 30 years. Those nice people at the Pentagon working for the Cult-controlled Defense Advanced Research Projects Agency (DARPA) claimed in April, 2021, they have developed a microchip inserted under the skin to detect 'asymptomatic Covid-19 infection' before it becomes an outbreak and a 'revolutionary filter' that can remove the 'virus' from the blood when attached to a dialysis machine. The only problems with this are that the 'virus' does not exist and people transmitting the 'virus' with no symptoms is brain-numbing bullshit. This is, of course, not a ruse to get people to be microchipped for very different reasons. DARPA also said it was producing a one-stop 'vaccine' for the 'virus' and all 'variants'. One of the most sinister organisations on Planet Earth is doing this? Better have it then. These people are insane because Wetiko that possesses them is insane.

Researchers from the Salk Institute in California announced they have created an embryo that is part human and part monkey. My books going back to the 1990s have exposed experiments in top secret underground facilities in the United States where humans are being crossed with animal and non-human 'extraterrestrial' species. They are now easing that long-developed capability into the public arena and there is much more to come given we are dealing with psychiatric basket cases. Talking of which – Elon Musk's scientists at Neuralink trained a monkey to play Pong and other puzzles on a computer screen using a joystick and when the monkey made the correct move a metal tube squirted banana smoothie into his mouth which is the basic technique for training humans into unquestioning compliance. Two Neuralink chips were in the monkey's skull and more than 2,000 wires 'fanned out' into its brain. Eventually the monkey played a video game purely with its brain waves. Psychopathic narcissist Musk said the 'breakthrough' was a step towards putting Neuralink chips into human skulls and merging minds with artificial intelligence. *Exactly*. This man is so dark and Cult to his DNA.

World Economic Fascism (WEF)

The World Economic Forum is telling you the plan by the statements made at its many and various events. Cult-owned fascist YouTube CEO Susan Wojcicki spoke at the 2021 WEF Global Technology Governance Summit (see the name) in which 40 governments and 150 companies met to ensure 'the responsible design and deployment of emerging technologies'. Orwellian translation: 'Ensuring the design and deployment of long-planned technologies will advance the Cult agenda for control and censorship.' Freedom-destroyer and Nuremberg-bound Wojcicki expressed support for tech platforms like hers to censor content that is 'technically legal but could be harmful'. Who decides what is 'harmful'? She does and they do. 'Harmful' will be whatever the Cult doesn't want people to see and we have legislation proposed by the UK government that would censor content on the basis of 'harm' no matter if the information is fair, legal and provably true. Make that *especially* if it is fair, legal and provably true. Wojcicki called for a global coalition to be formed to enforce content moderation standards through automated censorship. This is a woman and mega-censor so self-deluded that she shamelessly accepted a 'free expression' award – *Wojcicki* – in an event sponsored by her own *YouTube*. They have no shame and no self-awareness.

You know that 'Covid' is a scam and Wojcicki a Cult operative when YouTube is censoring medical and scientific opinion purely on the grounds of whether it supports or opposes the Cult 'Covid' narrative. Florida governor Ron DeSantis compiled an expert panel with four professors of medicine from Harvard, Oxford, and Stanford Universities who spoke against forcing children and vaccinated people to wear masks. They also said there was no proof that lockdowns reduced spread or death rates of 'Covid-19'. Cult-gofer Wojcicki and her YouTube deleted the panel video 'because it included content that contradicts the consensus of local and global health authorities regarding the efficacy of masks to prevent the spread of Covid-19'. This 'consensus' refers to what the Cult tells the World Health Organization to say and the WHO tells 'local health authorities' to do. Wojcicki knows this, of course. The panellists pointed out that censorship of scientific debate was responsible for deaths from many causes, but Wojcicki couldn't care less. She would not dare go against what she is told and as a disgrace to humanity she wouldn't want to anyway. The UK government is seeking to pass a fascist 'Online Safety Bill' to specifically target with massive fines and other means non-censored video and social media platforms to make them censor 'lawful but harmful' content like the Cult-owned Facebook, Twitter, Google and YouTube. What is 'lawful but harmful' would be decided by the fascist Blair-created Ofcom.

Another WEF obsession is a cyber-attack on the financial system and this is clearly what the Cult has planned to take down the bank accounts of everyone – except theirs. Those that think they have enough money for the Cult agenda not to matter to them have got a big lesson coming if they continue to ignore what is staring them in the face. The World Economic Forum, funded by Gates and fronted by Klaus Schwab, announced it would be running a 'simulation' with the Russian government and global banks of just such an attack called Cyber Polygon 2021. What they simulate – as with the 'Covid' Event 201 – they plan to instigate. The WEF is involved in a project with the Cult-owned Carnegie Endowment for International Peace called the WEF-Carnegie Cyber Policy Initiative which seeks to merge Wall Street banks, 'regulators' (I love it) and intelligence agencies to 'prevent' (arrange and allow) a cyber-attack that would bring down the global financial system as long planned by those that control the WEF and the Carnegie operation. The Carnegie Endowment for International Peace sent an instruction to First World

War US President Woodrow Wilson not to let the war end before society had been irreversibly transformed.

The Wuhan lab diversion

As I close, the Cult-controlled authorities and lapdog media are systematically pushing 'the virus was released from the Wuhan lab' narrative. There are two versions – it happened by accident and it happened on purpose. Both are nonsense. The perceived existence of the never-shown-to-exist 'virus' is vital to sell the impression that there is actually an infective agent to deal with and to allow the endless potential for terrifying the population with 'variants' of a 'virus' that does not exist. The authorities at the time of writing are going with the 'by accident' while the alternative media is promoting the 'on purpose'. Cable news host Tucker Carlson who has questioned aspects of lockdown and 'vaccine' compulsion has bought the Wuhan lab story. 'Everyone now agrees' he said. Well, I don't and many others don't and the question is *why* does the system and its media suddenly 'agree'? When the media moves as one unit with a narrative it is always a lie – witness the hour by hour mendacity of the 'Covid' era. Why would this Cult-owned combination which has unleashed lies like machine gun fire suddenly 'agree' to tell the truth??

Much of the alternative media is buying the lie because it fits the conspiracy narrative, but it's the *wrong* conspiracy. The real conspiracy is that *there is no virus* and that is what the Cult is desperate to hide. The idea that the 'virus' was released by accident is ludicrous when the whole 'Covid' hoax was clearly long-planned and waiting to be played out as it was so fast in accordance with the Rockefeller document and Event 201. So they prepared everything in detail over decades and then sat around strumming their fingers waiting for an 'accidental' release from a bio-lab? *What??* It's crazy. Then there's the 'on purpose' claim. You want to circulate a 'deadly virus' and hide the fact that you've done so and you release it down the street from the highest-level bio-lab in China? I repeat – *What??*

You would release it far from that lab to stop any association being made. But, no, we'll do it in a place where the connection was certain to be made. Why would you need to scam 'cases' and 'deaths' and pay hospitals to diagnose 'Covid-19' if you had a real 'virus'? What are sections of the alternative media doing believing this crap? Where were all the mass deaths in Wuhan from a 'deadly pathogen' when the recovery to normal life after the initial propaganda was dramatic in speed? Why isn't the 'deadly pathogen' now circulating all over China with bodies in the street? Once again we have the technique of tell them what they want to hear and they will likely believe it. The alternative media has its 'conspiracy' and with Carlson it fits with his 'China is the danger' narrative over years. China *is* a danger as a global Cult operations centre, but not for this reason. The Wuhan lab story also has the potential to instigate conflict with China when at some stage the plan is to trigger a Problem-Reaction-Solution confrontation with the West. Question everything – *everything* – and especially when the media agrees on a common party line.

Third wave ... fourth wave ... fifth wave ...

As the book went into production the world was being set up for more lockdowns and a 'third wave' supported by invented 'variants' that were increasing all the time and will continue to do so in public statements and computer programs, but not in reality. India became the new Italy in the 'Covid' propaganda campaign and we were told to be frightened of the new 'Indian strain'. Somehow I couldn't find it within myself to do so. A document produced for the UK government entitled 'Summary of further modelling of easing of restrictions – Roadmap Step 2' declared that a third wave was inevitable (of course when it's in the script) and it would be the fault of children and those who refuse the health-destroying fake 'Covid vaccine'. One of the computer models involved came from the Cult-owned *Imperial College* and the other from Warwick University which I wouldn't trust to tell me the date in a calendar factory. The document states that both models presumed extremely high uptake of the ‘Covid vaccines’ and didn’t allow for ‘variants’. The document states: ‘The resurgence is a result of some people (mostly children) being ineligible for vaccination; others choosing not to receive the vaccine; and others being vaccinated but not perfectly protected.’ The mendacity takes the breath away. Okay, blame those with a brain who won’t take the DNA-modifying shots and put more pressure on children to have it as ‘trials’ were underway involving children as young as six months with parents who give insanity a bad name. Massive pressure is being put on the young to have the fake ‘vaccine’ and child age consent limits have been systematically lowered around the world to stop parents intervening. Most extraordinary about the document was its claim that the ‘third wave’ would be driven by ‘the resurgence in both hospitalisations and deaths … dominated by *those that have received two doses of the vaccine*, comprising around 60-70% of the wave respectively’. The predicted peak of the ‘third wave’ suggested 300 deaths per day with 250 of them *fully ‘vaccinated’ people*. How many more lies do acquiescers need to be told before they see the obvious? Those who took the jab to ‘protect themselves’ are projected to be those who mostly get sick and die? So what’s in the ‘vaccine’? The document went on:

> It is possible that a summer of low prevalence could be followed by substantial increases in incidence over the following autumn and winter. Low prevalence in late summer should not be taken as an indication that SARS-CoV-2 has retreated or that the population has high enough levels of immunity to prevent another wave.

They are telling you the script and while many British people believed ‘Covid’ restrictions would end in the summer of 2021 the government was preparing for them to be ongoing. Authorities were awarding contracts for ‘Covid marshals’ to police the restrictions with contracts starting in July, 2021, and going through to January 31st, 2022, and the government was advertising for ‘Media Buying Services’ to secure media propaganda slots worth a potential £320 million for ‘Covid-19 campaigns’ with a contract not ending until March, 2022. The recipient – via a list of other front companies – was reported to be American media marketing giant Omnicom Group

Inc. While money is no object for 'Covid' the UK waiting list for all other treatment – including life-threatening conditions – passed 4.5 million. Meantime the Cult is seeking to control all official 'inquiries' to block revelations about what has really been happening and why. It must not be allowed to – we need Nuremberg jury trials in every country. The cover-up doesn't get more obvious than appointing ultra-Zionist professor Philip Zelikow to oversee two dozen US virologists, public health officials, clinicians, former government officials and four American 'charitable foundations' to 'learn the lessons' of the 'Covid' debacle. The personnel will be those that created and perpetuated the 'Covid' lies while Zelikow is the former executive director of the 9/11 Commission who ensured that the truth about those attacks never came out and produced a report that must be among the most mendacious and manipulative documents ever written – see *The Trigger* for the detailed exposure of the almost unimaginable 9/11 story in which Sabbatians can be found at every level.

Passive no more

People are increasingly challenging the authorities with amazing numbers of people taking to the streets in London well beyond the ability of the Face-Nappies to stop them. Instead the Nappies choose situations away from the mass crowds to target, intimidate, and seek to promote the impression of 'violent protestors'. One such incident happened in London's Hyde Park. Hundreds of thousands walking through the streets in protest against 'Covid' fascism were ignored by the Cult-owned BBC and most of the rest of the mainstream media, but they delighted in reporting how police were injured in 'clashes with protestors'. The truth was that a group of people gathered in Hyde Park at the end of one march when most had gone home and they were peacefully having a good time with music and chat. Face-Nappies who couldn't deal with the full-march crowd then waded in with their batons and got more than they bargained for. Instead of just standing for this criminal brutality the crowd used their numerical superiority to push the Face-Nappies out of the park. Eventually the Nappies turned and ran. Unfortunately two or three idiots in the crowd threw drink cans striking two officers which gave the media and the government the image they wanted to discredit the 99.9999 percent who were peaceful. The idiots walked straight into the trap and we must always be aware of potential agent provocateurs used by the authorities to discredit their targets.

This response from the crowd – the can people apart – must be a turning point when the public no longer stand by while the innocent are arrested and brutally attacked by the Face-Nappies. That doesn't mean to be violent, that's the last thing we need. We'll leave the violence to the Face-Nappies and government. But it does mean that when the Face-Nappies use violence against peaceful people the numerical superiority is employed to stop them and make citizen's arrests or Common Law arrests for a breach of the peace. The time for being passive in the face of fascism is over.

We are the many, they are the few, and we need to make that count before there is no freedom left and our children and grandchildren face an ongoing fascist nightmare.

COME ON PEOPLE – IT'S TIME.

One final thought …

The power of love
A force from above
Cleaning my soul
Flame on burn desire
Love with tongues of fire
Purge the soul
Make love your goal

I’ll protect you from the hooded claw
Keep the vampires from your door
When the chips are down I’ll be around
With my undying, death-defying
Love for you

Envy will hurt itself
Let yourself be beautiful
Sparkling love, flowers
And pearls and pretty girls
Love is like an energy
Rushin’ rushin’ inside of me

This time we go sublime
Lovers entwine, divine, divine,
Love is danger, love is pleasure
Love is pure – the only treasure

I’m so in love with you
Purge the soul
Make love your goal

The power of love
A force from above
Cleaning my soul
The power of love
A force from above
A sky-scraping dove

Flame on burn desire
Love with tongues of fire
Purge the soul
Make love your goal

Frankie Goes To Hollywood

APPENDIX

Cowan-Kaufman-Morell Statement on Virus Isolation (SOVI)

Isolation: The action of isolating; the fact or condition of being isolated or standing alone; separation from other things or persons; solitariness
Oxford English Dictionary

The controversy over whether the SARS-CoV-2 virus has ever been isolated or purified continues. However, using the above definition, common sense, the laws of logic and the dictates of science, any unbiased person must come to the conclusion that the SARS-CoV-2 virus has never been isolated or purified. As a result, no confirmation of the virus' existence can be found. The logical, common sense, and scientific consequences of this fact are:

- the structure and composition of something not shown to exist can't be known, including the presence, structure, and function of any hypothetical spike or other proteins;
- the genetic sequence of something that has never been found can't be known;
- "variants" of something that hasn't been shown to exist can't be known;
- it's impossible to demonstrate that SARS-CoV-2 causes a disease called Covid-19.

In as concise terms as possible, here's the proper way to isolate, characterize and demonstrate a new virus. First, one takes samples (blood, sputum, secretions) from many people (e.g. 500) with symptoms which are unique and specific enough to characterize an illness. Without mixing these samples with ANY tissue or products that also contain genetic material, the virologist macerates, filters and ultracentrifuges i.e. *purifies* the specimen. This common virology technique, done for decades to isolate bacteriophages[1] and so-called giant viruses in every virology lab, then allows the virologist to demonstrate with electron microscopy thousands of identically sized and shaped particles. These particles are the isolated and purified virus.

These identical particles are then checked for uniformity by physical and/or microscopic techniques. Once the purity is determined, the particles may be further characterized. This would include examining the structure, morphology, and chemical composition of the particles. Next, their genetic makeup is characterized by extracting the genetic material directly from the purified particles and using genetic-sequencing techniques, such as Sanger sequencing, that have also been around for decades. Then one does an analysis to confirm that these uniform particles are exogenous (outside) in origin as a virus is conceptualized to be, and not the normal breakdown products of dead and dying tissues.[2] (As of May 2020, we know that virologists have no way to determine whether the particles they're seeing are viruses or just normal break-down products of dead and dying tissues.)[3]

---

[1] Isolation, characterization and analysis of bacteriophages from the haloalkaline lake Elmenteita, KenyaJuliah Khayeli Akhwale et al, PLOS One, Published: April 25, 2019. https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0215734 – accessed 2/15/21

---

[2] "Extracellular Vesicles Derived From Apoptotic Cells: An Essential Link Between Death and Regeneration," Maojiao Li1 et al, Frontiers in Cell and Developmental Biology, 2020 October 2. https://www.frontiersin.org/articles/10.3389/fcell.2020.573511/full – accessed 2/15/21

If we have come this far then we have fully isolated, characterized, and genetically sequenced an exogenous virus particle. However, we still have to show it is causally related to a disease. This is carried out by exposing a group of healthy subjects (animals are usually used) to this isolated, purified virus in the manner in which the disease is thought to be transmitted. If the animals get sick with the same disease, as confirmed by clinical and autopsy findings, one has now shown that the virus actually causes a disease. This demonstrates infectivity and transmission of an infectious agent.

None of these steps has even been attempted with the SARS-CoV-2 virus, nor have all these steps been successfully performed for any so-called pathogenic virus. Our research indicates that a single study showing these steps does not exist in the medical literature.

Instead, since 1954, virologists have taken unpurified samples from a relatively few people, often less than ten, with a similar disease. They then minimally process this sample and inoculate this unpurified sample onto tissue culture containing usually four to six other types of material – all of which contain identical genetic material as to what is called a “virus.” The tissue culture is starved and poisoned and naturally disintegrates into many types of particles, some of which contain genetic material. Against all common sense, logic, use of the English language and scientific integrity, this process is called “virus isolation.” This brew containing fragments of genetic material from many sources is then subjected to genetic analysis, which then creates in a computer-simulation process the alleged sequence of the alleged virus, a so called in silico genome. At no time is an actual virus confirmed by electron microscopy. At no time is a genome extracted and sequenced from an actual virus. This is scientific fraud.

---

3 “The Role of Extraellular Vesicles as Allies of HIV, HCV and SARS Viruses,” Flavia Giannessi, et al, Viruses, 2020 May

The observation that the unpurified specimen — inoculated onto tissue culture along with toxic antibiotics, bovine fetal tissue, amniotic fluid and other tissues — destroys the kidney tissue onto which it is inoculated is given as evidence of the virus' existence and pathogenicity. This is scientific fraud.

From now on, when anyone gives you a paper that suggests the SARS-CoV-2 virus has been isolated, please check the methods sections. If the researchers used Vero cells or any other culture method, you know that their process was not isolation. You will hear the following excuses for why actual isolation isn't done:

1. There were not enough virus particles found in samples from patients to analyze.

2. Viruses are intracellular parasites; they can't be found outside the cell in this manner.

If No. 1 is correct, and we can't find the virus in the sputum of sick people, then on what evidence do we think the virus is dangerous or even lethal? If No. 2 is correct, then how is the virus spread from person to person? We are told it emerges from the cell to infect others. Then why isn't it possible to find it?

Finally, questioning these virology techniques and conclusions is not some distraction or divisive issue. Shining the light on this truth is essential to stop this terrible fraud that humanity is confronting. For, as we now know, if the virus has never been isolated, sequenced or shown to cause illness, if the virus is imaginary, then why are we wearing masks, social distancing and putting the whole world into prison?

Finally, if pathogenic viruses don't exist, then what is going into those injectable devices erroneously called "vaccines," and what is their purpose? This scientific question is the most urgent and relevant one of our time.

We are correct. The SARS-CoV2 virus does not exist.

Sally Fallon Morell, MA
Dr. Thomas Cowan, MD
Dr. Andrew Kaufman, MD

Bibliography

Alinsky, Saul: *Rules for Radicals* (Vintage, 1989)

Antelman, Rabbi Marvin: *To Eliminate the Opiate* (Zahavia, 1974)

Bastardi, Joe: *The Climate Chronicles* (Relentless Thunder Press, 2018)

Cowan, Tom: *Human Heart, Cosmic Heart* (Chelsea Green Publishing, 2016)

Cowan, Tom, and Fallon Morell, Sally: *The Contagion Myth* (Skyhorse Publishing, 2020)

Forbes, Jack D: *Columbus And Other Cannibals – The Wetiko Disease of Exploitation, Imperialism, and Terrorism* (Seven Stories Press, 2008 – originally published in 1979)

Gates, Bill: *How to Avoid a Climate Disaster: The Solutions We Have and the Breakthroughs We Need* (Allen Lane, 2021)

Huxley, Aldous: *Brave New World* (Chatto & Windus, 1932)

Köhnlein, Dr Claus, and Engelbrecht, Torsten: *Virus Mania* (emu-Vertag, Lahnstein, 2020)

Lanza, Robert, and Berman, Bob: *Biocentrism* (BenBella Books, 2010)

Lash, John Lamb: *Not In His Image* (Chelsea Green Publishing, 2006)

Lester, Dawn, and Parker, David: *What Really Makes You Ill – Why everything you thought you knew about disease is wrong* (Independently Published, 2019)

Levy, Paul: *Dispelling Wetiko, Breaking the Spell of Evil* (North Atlantic Books, 2013)

Marx, Karl: *A World Without Jews* (Philosophical Library, first edition, 1959)

Mullis, Kary: *Dancing Naked in the Mine Field* (Bloomsbury, 1999)

O'Brien, Cathy: *Trance-Formation of America* (Reality Marketing, 1995)

Scholem, Gershon: *The Messianic Idea in Judaism* (Schocken Books, 1994)

Schwab, Klaus, and Davis, Nicholas: *Shaping the Future of the Fourth Industrial Revolution: A guide to building a better world* (Penguin Books, 2018)

Schwab, Klaus: *The Great Reset* (Agentur Schweiz, 2020)

Sunstein, Cass and Thaler, Richard: *Nudge: Improving Decisions About Health, Wealth, and Happiness* (Penguin, 2009)

Swan, Shanna: *Count Down: How Our Modern World Is Threatening Sperm Counts, Altering Male and Female Reproductive Development and Imperiling the Future of the Human Race* (Scribner, 2021)

Tegmark, Max: *Our Mathematical Universe: My Quest for the Ultimate Nature of Reality* (Penguin, 2015)

Velikovsky, Immanuel: *Worlds in Collision* (Paradigma, 2009)

Wilton, Robert: *The Last Days of the Romanovs* (Blurb, 2018, first published 1920)

Index

A

abusive relationships
  blaming themselves, abused as ref1
  children ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
  conspiracy theories ref1
  domestic abuse ref1, ref2
  economic abuse and dependency ref1
  isolation ref1
  physical abuse ref1
  psychological abuse ref1
  signs of abuse ref1
addiction
  alcoholism ref1
  frequencies ref1
  substance abuse ref1, ref2
  technology ref1, ref2, ref3
Adelson, Sheldon ref1, ref2, ref3
Agenda 21/Agenda 2030 (UN) ref1, ref2, ref3, ref4
AIDs/HIV ref1
  causal link between HIV and AIDs ref1, ref2
retroviruses ref1
  testing ref1, ref2
  trial-run for Covid-19, as ref1, ref2
aliens/extraterrestrials ref1, ref2
aluminium ref1
Amazon ref1, ref2, ref3

amplification cycles ref1, ref2
anaphylactic shock ref1, ref2, ref3, ref4
animals ref1, ref2, ref3
antibodies ref1, ref2, ref3, ref4, ref5
Antifa ref1, ref2, ref3, ref4
antigens ref1, ref2
anti-Semitism ref1, ref2, ref3
Archons ref1, ref2
  consciousness ref1, ref2, ref3
  energy ref1, ref2, ref3
  ennoia ref1
  genetic manipulation ref1, ref2
  inversion ref1, ref2, ref3
  lockdowns ref1
  money ref1
  radiation ref1
  religion ref1, ref2
  technology ref1, ref2, ref3
  Wetiko factor ref1, ref2, ref3, ref4
artificial intelligence (AI) ref1
army made up of robots ref1, ref2
  Human 2.0 ref1, ref2
  Internet ref1
  MHRA ref1
  Morgellons fibres ref1, ref2
  Smart Grid ref1
  Wetiko factor ref1
asymptomatic, Covid-19 as ref1, ref2, ref3
aviation industry ref1

B

banking, finance and money ref1, ref2, ref3
  2008 crisis ref1, ref2
  boom and bust ref1
  cashless digital money systems ref1
  central banks ref1
  credit ref1
  digital currency ref1
  fractional reserve lending ref1
  Great Reset ref1
  guaranteed income ref1, ref2, ref3
  Human 2.0 ref1
  incomes, destruction of ref1, ref2
  interest ref1
  one per cent ref1, ref2
  scams ref1
BBC ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
Becker-Phelps, Leslie ref1
Behavioural Insights Team (BIT) (Nudge Unit) ref1, ref2, ref3
behavioural scientists *and* **psychologists, advice from** ref1, ref2
Bezos, Jeff ref1, ref2, ref3, ref4
Biden, Hunter ref1
Biden, Joe ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11, ref12, ref13, ref14, ref15, ref16, ref17
Big Pharma
  cholesterol ref1
  health professionals ref1, ref2
  immunity from prosecution in US ref1
  vaccines ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
  Wetiko factor ref1, ref2
  WHO ref1, ref2, ref3
Bill and Melinda Gates Foundation ref1, ref2, ref3, ref4, ref5, ref6, ref7

billionaires ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9 ref10, ref11
bird flu (H5N1) ref1
Black Lives Matter (BLM) ref1, ref2, ref3, ref4, ref5
Blair, Tony ref1, ref2, ref3, ref4, ref5, ref6, ref7
Brin, Sergei ref1, ref2, ref3, ref4, ref5, ref6, ref7
British Empire ref1
Bush, George HW ref1, ref2
Bush, George W ref1, ref2, ref3, ref4
Byrd, Robert ref1

C

Canada
  Global Cult ref1
  hate speech ref1
  internment ref1
  masks ref1
  old people ref1
  SARS-COV-2 ref1
  satellites ref1
  vaccines ref1
  wearable technology ref1
Capitol Hill riot ref1, ref2
  agents provocateur ref1
  Antifa ref1
  Black Lives Matter (BLM) ref1, ref2
  QAnon ref1
  security precautions, lack of ref1, ref2, ref3
carbon dioxide ref1, ref2
care homes, deaths in ref1, ref2
cashless digital money systems ref1
censorship ref1, ref2, ref3, ref4, ref5

fact-checkers ref1
masks ref1
media ref1, ref2
private messages ref1
social media ref1, ref2, ref3, ref4, ref5, ref6
transgender persons ref1
vaccines ref1, ref2, ref3
Wokeness ref1
Centers for Disease Control (CDC) (United States) ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11, ref12, ref13
centralisation ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
chakras ref1
change agents ref1, ref2, ref3
chemtrails ref1, ref2, ref3
chief medical officers and scientific advisers ref1, ref2, ref3, ref4, ref5, ref6
children *see also* **young people**
abuse ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
care, taken into ref1, ref2, ref3
education ref1, ref2, ref3, ref4
energy ref1
family courts ref1
hand sanitisers ref1
human sacrifice ref1
lockdowns ref1, ref2, ref3
masks ref1, ref2, ref3, ref4, ref5
mental health ref1
old people ref1
parents, replacement of ref1, ref2
Psyop (psychological operation), Covid as a ref1, ref2
reframing ref1
smartphone addiction ref1

  - social distancing and isolation ref1
  - social media ref1
  - transgender persons ref1, ref2
  - United States ref1
  - vaccines ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
  - Wetiko factor ref1

China ref1, ref2, ref3, ref4
  - anal swab tests ref1
  - **Chinese Revolution** ref1, ref2, ref3
  - digital currency ref1
  - Global Cult ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9
  - guaranteed income ref1
  - Imperial College ref1
  - Israel ref1
  - lockdown ref1, ref2
  - masculinity crisis ref1
  - masks ref1
  - media ref1
  - origins of virus in China ref1, ref2, ref3, ref4, ref5
  - pollution causing respiratory diseases ref1
  - Sabbatians ref1, ref2
  - Smart Grid ref1, ref2
  - social credit system ref1
  - testing ref1, ref2
  - United States ref1, ref2
  - vaccines ref1, ref2
  - Wetiko factor ref1
  - wet market conspiracy ref1
  - Wuhan ref1, ref2, ref3, ref4, ref5, ref6, ref7

cholesterol ref1, ref2

Christianity ref1, ref2, ref3, ref4, ref5
  - criticism ref1
  - cross, inversion of the ref1

  Nag Hammadi texts ref1, ref2, ref3
  Roman Catholic Church ref1, ref2
  Sabbatians ref1, ref2
  Satan ref1, ref2, ref3, ref4
  Wokeness ref1

class ref1, ref2

climate change hoax ref1, ref2, ref3, ref4, ref5
  Agenda 21/Agenda 2030 ref1, ref2, ref3
  carbon dioxide ref1, ref2
  Club of Rome ref1, ref2, ref3, ref4, ref5
  fear ref1
  funding ref1
  Global Cult ref1
  green new deals ref1
  green parties ref1
  inversion ref1
  perception, control of ref1
  PICC ref1
  reframing ref1
  temperature, increases in ref1
  United Nations ref1, ref2
  Wikipedia ref1
  Wokeness ref1, ref2

Clinton, Bill ref1, ref2, ref3, ref4, ref5, ref6

Clinton, Hillary ref1, ref2, ref3

the cloud ref1, ref2, ref3, ref4, ref5, ref6, ref7

Club of Rome and climate change hoax ref1, ref2, ref3, ref4, ref5

cognitive therapy ref1

Cohn, Roy ref1

Common Law ref1
  Admiralty Law ref1
  arrests ref1, ref2

  contractual law, Statute Law as ref1
  corporate entities, people as ref1
  legalese ref1
  sea, law of the ref1
  Statute Law ref1

Common Purpose leadership programme ref1, ref2

communism ref1, ref2

co-morbidities ref1

computer-generated virus,

Covid-19 as ref1, ref2, ref3

computer models ref1, ref2, ref3, ref4, ref5

connections ref1, ref2, ref3, ref4

consciousness ref1, ref2, ref3, ref4
  Archons ref1, ref2, ref3
  expanded ref1, ref2, ref3, ref4, ref5, ref6, ref7
  experience ref1
  heart ref1
  infinity ref1, ref2
  religion ref1, ref2
  self-identity ref1
  simulation thesis ref1
  vaccines ref1
  Wetiko factor ref1, ref2

conspiracy theorists ref1, ref2, ref3, ref4, ref5

contradictory rules ref1

contrails ref1

Corman-Drosten test ref1, ref2, ref3, ref4

countermimicry ref1, ref2, ref3

Covid-19 vaccines *see* vaccines

Covidiots ref1, ref2

Cowan, Tom ref1, ref2, ref3, ref4

crimes against humanity ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8

cyber-operations ref1
cyberwarfare ref1

D

DARPA (Defense Advanced Research Projects Agency) ref1
deaths
  care homes ref1
  certificates ref1, ref2, ref3, ref4
  mortality rate ref1
  post-mortems/autopsies ref1
  recording ref1, ref2, ref3, ref4, ref5, ref6, ref7
  vaccines ref1, ref2, ref3, ref4, ref5
deceit
  pyramid of deceit ref1, ref2
  sequence of deceit ref1
decoding ref1, ref2, ref3
dehumanisation ref1, ref2, ref3
Delphi technique ref1
democracy ref1
dependency ref1, ref2, ref3, ref4, ref5
Descartes, René ref1
DNA
  numbers ref1
  vaccines ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
DNR (do not resuscitate)
  orders ref1
domestic abuse ref1, ref2
downgrading of Covid-19 ref1
Drosten, Christian ref1, ref2, ref3, ref4, ref5, ref6, ref7
Duesberg, Peter ref1, ref2

E

economic abuse ref1

Edmunds, John ref1, ref2

education ref1, ref2, ref3, ref4

electromagnetic spectrum ref1, ref2

Enders, John ref1

energy
  Archons ref1, ref2, ref3
  children and young people ref1
  consciousness ref1
  decoding ref1
  frequencies ref1, ref2, ref3, ref4
  heart ref1
  human energy field ref1
  source, humans as an energy ref1, ref2
  vaccines ref1
  viruses ref1

ennoia ref1

Epstein, Jeffrey ref1, ref2

eternal 'I' ref1, ref2

ethylene oxide ref1

European Union ref1, ref2, ref3, ref4

Event **ref1** *and* **Bill Gates** ref2

exosomes, Covid-19 as natural defence mechanism called ref1

experience ref1, ref2

Extinction Rebellion ref1, ref2

F

Facebook
  addiction ref1, 448–50
  Facebook

  Archons ref1
  censorship ref1, ref2, ref3
  hate speech ref1
  monopoly, as ref1
  private messages, censorship of ref1
  Sabbatians ref1
  United States election fraud ref1
  vaccines ref1
  Wetiko factor ref1
fact-checkers ref1
Fauci, Anthony ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11, ref12
fear ref1, ref2, ref3, ref4
  climate change ref1
  computer models ref1
  conspiracy theories ref1
  empty hospitals ref1
  Italy ref1, ref2, ref3
  lockdowns ref1, ref2, ref3, ref4
  masks ref1, ref2
  media ref1, ref2
  medical staff ref1
  Psyop (psychological operation), Covid as a ref1
  Wetiko factor ref1, ref2
female infertility ref1
Fermi Paradox ref1
Ferguson, Neil ref1, ref2, ref3, ref4, ref5, ref6, ref7
fertility, decline in ref1
The Field ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
finance *see* **banking, finance and money**
five-senses ref1, ref2
  Archons ref1, ref2, ref3

  - censorship ref1
  - consciousness, expansion of ref1, ref2, ref3, ref4, ref5, ref6
  - decoding ref1
  - education ref1, ref2
  - the Field ref1, ref2
  - God, personification of ref1
  - infinity ref1, ref2
  - media ref1
  - paranormal ref1
  - perceptual programming ref1, ref2
  - Phantom Self ref1
  - pneuma not nous, using ref1
  - reincarnation ref1
  - self-identity ref1
  - Wetiko factor ref1, ref2, ref3, ref4, ref5, ref6

5G ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8

Floyd, George and protests, killing of ref1

flu, re-labelling of ref1, ref2, ref3

food and water, control of ref1, ref2

Freemasons ref1, ref2, ref3, ref4, ref5, ref6

Frei, Rosemary ref1

frequencies
  - addictions ref1
  - Archons ref1, ref2, ref3
  - awareness ref1
  - chanting and mantras ref1
  - consciousness ref1
  - decoding ref1, ref2
  - education ref1
  - electromagnetic (EMF) frequencies ref1
  - energy ref1, ref2, ref3, ref4
  - fear ref1

  the Field ref1, ref2 5G ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
  five-senses ref1, ref2
  ghosts ref1
  Gnostics ref1
  hive-minds ref1
  human, meaning of ref1
  light ref1, ref2
  love ref1, ref2
  magnetism ref1
  perception ref1
  reality ref1, ref2, ref3
  simulation ref1
  terror ref1
  vaccines ref1
  Wetiko ref1, ref2, ref3
Fuellmich, Reiner ref1, ref2, ref3
furlough/rescue payments ref1

G

Gallo, Robert ref1, ref2, ref3
Gates, Bill
  Archons ref1, ref2, ref3
  climate change ref1, ref2, ref3, ref4
  Daily Pass tracking system ref1
  Epstein ref1
  fascism ref1
  five senses ref1
  GAVI ref1
  Great Reset ref1
  GSK ref1
  Imperial College ref1, ref2
  Johns Hopkins University ref1, ref2, ref3

  lockdowns ref1, ref2
  masks ref1
  Nuremberg trial, proposal for ref1, ref2
  Rockefellers ref1, ref2
  social distancing and isolation ref1
  Sun, dimming the ref1
  synthetic meat ref1, ref2
  vaccines ref1, ref2, ref3, ref4, ref5, ref6, ref7
  Wellcome Trust ref1
  Wetiko factor ref1, ref2, ref3
  WHO ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
  Wokeness ref1
  World Economic Forum ref1, ref2, ref3, ref4
Gates, Melinda ref1, ref2, ref3
GAVI vaccine alliance ref1
genetics, manipulation of ref1, ref2, ref3
Germany ref1, ref2, ref3, ref4, ref5, ref6 *see also* **Nazi Germany**
Global Cult ref1, ref2, ref3, ref4, ref5
  anti-human, why Global Cult is ref1
  Black Lives Matter (BLM) ref1, ref2, ref3, ref4
  China ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9
  climate change hoax ref1
  contradictory rules ref1
  Covid-19 ref1, ref2, ref3
  fascism ref1
  geographical origins ref1
  immigration ref1
  Internet ref1
  mainstream media ref1, ref2
  masks ref1, ref2
  monarchy ref1
  non-human dimension ref1

- perception ref1
- political parties ref1, ref2
- pyramidal hierarchy ref1, ref2, ref3
- reframing ref1
- Sabbantian-Frankism ref1, ref2
- science, manipulation of ref1
- spider and the web ref1
- transgender persons ref1
- vaccines ref1
- who controls the Cult ref1
- Wokeness ref1, ref2, ref3, ref4

globalisation ref1, ref2

Gnostics ref1, ref2, ref3, ref4, ref5

Google ref1, ref2, ref3, ref4

government
- behavioural scientists and psychologists, advice from ref1, ref2

definition ref1

Joint Biosecurity Centre (JBC) ref1

people, abusive relationship with ref1

Great Reset ref1, ref2, ref3, ref4, ref5, ref6
- fascism ref1, ref2, ref3
- financial system ref1
- Human 2.0 ref1
- water and food, control of ref1

green parties ref1

Griesz-Brisson, Margarite ref1

guaranteed income ref1, ref2, ref3

H

Hancock, Matt ref1, ref2, ref3, ref4, ref5

hand sanitisers ref1

heart ref1, ref2

hive-minds/groupthink ref1, ref2, ref3
holographs ref1, ref2, ref3, ref4
hospitals, empty ref1
human, meaning of ref1
Human 2.0 ref1
  addiction to technology ref1
  artificial intelligence (AI) ref1, ref2
  elimination of Human 1.0 ref1
  fertility, decline in ref1
  Great Reset ref1
  implantables ref1
  money ref1
  mRNA ref1
  nanotechnology ref1
  parents, replacement of ref1, ref2
  Smart Grid, connection to ref1, ref2
  synthetic biology ref1, ref2, ref3, ref4
  testosterone levels, decrease in ref1
  transgender = transhumanism ref1, ref2, ref3
  vaccines ref1, ref2, ref3, ref4
human sacrifice ref1, ref2, ref3
Hunger Games Society ref1, ref2, ref3, ref4, ref5, ref6, ref7
Huxley, Aldous ref1, ref2, ref3

I

identity politics ref1, ref2, ref3
Illuminati ref1, ref2
illusory physical reality ref1
immigration ref1, ref2, ref3, ref4
Imperial College ref1, ref2, ref3, ref4, ref5, ref6
implantables ref1, ref2

incomes, destruction of ref1, ref2
Infinite Awareness ref1, ref2, ref3, ref4
Internet ref1, ref2 *see also* social media
  artificial intelligence (AI) ref1
  independent journalism, lack of ref1
  Internet of Bodies (IoB) ref1
Internet of Everything (IoE) ref1, ref2
Internet of Things (IoT) ref1, ref2
lockdowns ref1
Psyop (psychological operation), Covid as a ref1
  trolls ref1
intersectionality ref1
inversion
  Archons ref1, ref2, ref3
  climate change hoax ref1
  energy ref1
  Judaism ref1, ref2, ref3
  symbolism ref1
  Wetiko factor ref1
  Wokeness ref1, ref2, ref3
Islam
  Archons ref1
  crypto-Jews ref1
  Islamic State ref1, ref2
  Jinn and Djinn ref1, ref2, ref3
  Ottoman Empire ref1
  Wahhabism ref1
isolation *see* **social distancing** *and* **isolation**
Israel
  China ref1
  Cyber Intelligence Unit Beersheba complex ref1
  expansion of illegal settlements ref1

  formation ref1
  Global Cult ref1
  Judaism ref1, ref2, ref3, ref4, ref5
  medical experiments, consent for ref1
  Mossad ref1, ref2, ref3, ref4
  Palestine-Israel conflict ref1, ref2, ref3
  parents, replacement of ref1
  Sabbatians ref1, ref2, ref3, ref4, ref5
  September 11, 2001, terrorist attacks on United States ref1
  Silicon Valley ref1
  Smart Grid ref1, ref2
  United States ref1, ref2
  vaccines ref1
  Wetiko factor ref1
Italy
  fear ref1, ref2, ref3
  Lombardy ref1, ref2, ref3
  vaccines ref1

J

Johns Hopkins University ref1, ref2, ref3, ref4, ref5, ref6, ref7
Johnson, Boris ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
Joint Biosecurity Centre (JBC) ref1
Judaism
  anti-Semitism ref1, ref2, ref3
  Archons ref1, ref2
  crypto-Jews ref1
  inversion ref1, ref2, ref3
  Israel ref1, ref2, ref3, ref4, ref5
  Labour Party ref1
  Nazi Germany ref1, ref2, ref3, ref4
  Sabbatians ref1, ref2, ref3, ref4, ref5

  Silicon Valley ref1
  Torah ref1
  United States ref1, ref2
  Zionists ref1, ref2, ref3

K

Kaufman, Andrew ref1, ref2, ref3, ref4
knowledge ref1, ref2, ref3, ref4, ref5, ref6
Koch's postulates ref1
Kurzweil, Ray ref1, ref2, ref3, ref4, ref5, ref6, ref7
Kushner, Jared **ref1**, **ref2**

L

Labour Party ref1, ref2
Lanka, Stefan ref1, ref2
Lateral Flow Device (LFD) ref1
Levy, Paul ref1, ref2, ref3
Life Program ref1
lockdowns ref1, ref2, ref3
  amplification tampering ref1
  Archons ref1
  Behavioural Insights Team ref1
  Black Lives Matter (BLM) ref1
  care homes, deaths in ref1
  children
abuse ref1, ref2
mental health ref1
  China ref1, ref2
  computer models ref1
  consequences ref1, ref2
  dependency ref1, ref2, ref3

  domestic abuse ref1
  fall in cases ref1
  fear ref1, ref2, ref3, ref4
  guaranteed income ref1
  Hunger Games Society ref1, ref2, ref3
  interaction, destroying ref1
  Internet ref1, ref2
  overdoses ref1
  perception ref1
police-military state ref1, ref2
protests ref1, ref2, ref3, ref4, ref5
psychopathic personality ref1, ref2, ref3
reporting/snitching, encouragement of ref1, ref2
  testing ref1
  vaccines ref1
  Wetiko factor ref1
  WHO ref1

love ref1, ref2, ref3

Lucifer ref1, ref2, ref3

M

Madej, Carrie ref1, ref2

Magufuli, John ref1, ref2

mainstream media ref1
  BBC ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
  censorship ref1, ref2
  China ref1
  climate change hoax ref1
  fear ref1, ref2
  Global Cult ref1, ref2
  independent journalism, lack of ref1
  Ofcom ref1, ref2, ref3

  perception ref1, ref2
  Psyop (psychological operation), Covid as a ref1
  Sabbatians ref1, ref2
  social disapproval ref1
  social distancing and isolation ref1
  United States ref1, ref2
  vaccines ref1, ref2, ref3, ref4, ref5

Mao Zedong ref1, ref2, ref3

Marx and Marxism ref1, ref2, ref3, ref4, ref5, ref6

masculinity ref1

masks/face coverings ref1, ref2, ref3
  censorship ref1
  children ref1, ref2, ref3, ref4, ref5
  China, made in ref1
  dehumanisation ref1, ref2, ref3
  fear ref1, ref2
  flu ref1
  health professionals ref1, ref2, ref3, ref4
  isolation ref1
  laughter ref1

mass non-cooperation ref1

microplastics, risk of ref1

mind control ref1

multiple masks ref1

oxygen deficiency ref1, ref2, ref3
police ref1, ref2, ref3, ref4, ref5
pollution, as cause of plastic ref1
Psyop (psychological operation), Covid as a ref1
reframing ref1, ref2
risk assessments, lack of ref1, ref2
self-respect ref1
surgeons ref1

United States ref1
vaccines ref1, ref2, ref3, ref4, ref5
Wetiko factor ref1
‘worms’ ref1
The Matrix movies ref1, ref2, ref3
measles ref1, ref2
media see mainstream media
Medicines and Healthcare products Regulatory Agency (MHRA) ref1, ref2, ref3, ref4
Mesopotamia ref1
messaging ref1
military-police state ref1, ref2, ref3
mind control ref1, ref2, ref3, ref4, ref5, ref6 *see also* **MKUltra**
MKUltra ref1, ref2, ref3
monarchy **ref1**
money *see* **banking, finance and money**
Montagnier, Luc ref1, ref2, ref3
Mooney, Bel ref1
Morgellons disease ref1, ref2
mortality rate ref1
Mullis, Kary ref1, ref2, ref3
Musk, Elon ref1

N

Nag Hammadi texts ref1, ref2, ref3
nanotechnology ref1, ref2, ref3
narcissism ref1
Nazi Germany ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
near-death experiences ref1, ref2
Neocons ref1, ref2, ref3

Neuro-Linguistic Programming (NLP) and the Delphi technique **ref1**

NHS (National Health Service)
  amplification cycles ref1
  Common Purpose ref1, ref2
  mind control ref1
  NHS England ref1
  saving the NHS ref1, ref2
  vaccines ref1, ref2, ref3, ref4, ref5
  whistle-blowers ref1, ref2, ref3

No-Problem-Reaction-Solution ref1, ref2, ref3, ref4

non-human dimension of Global Cult ref1

nous ref1

numbers, reality as ref1

Nuremberg Codes ref1, ref2, ref3

Nuremberg-like tribunal, proposal for ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11, ref12

O

Obama, Barack ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10

O'Brien, Cathy ref1, ref2, ref3, ref4

Ochel, Evita ref1

Ofcom ref1, ref2, ref3

old people ref1, ref2, ref3, ref4, ref5

Oneness ref1, ref2, ref3

Open Society Foundations (Soros) ref1, ref2, ref3

oxygen 406, 528–34

P

paedophilia ref1, ref2

Page, Larry ref1, ref2, ref3, ref4, ref5, ref6, ref7

Palestine-Israel conflict ref1, ref2, ref3
pandemic, definition of ref1
pandemic and health crisis scenarios/simulations ref1, ref2, ref3, ref4
paranormal ref1
PCR tests ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
Pearl Harbor attacks, prior knowledge of ref1
Pelosi, Nancy ref1, ref2, ref3
perception ref1, ref2, ref3, ref4
  climate change hoax ref1
  control ref1, ref2, ref3
  decoding ref1, ref2
  enslavement ref1
  externally-delivered perceptions ref1
  five senses ref1
  human labels ref1
  media ref1, ref2
  political parties ref1, ref2
  Psyop (psychological operation), Covid as a ref1
  sale of perception ref1
  self-identity ref1, ref2
  Wokeness ref1
Phantom Self ref1, ref2, ref3
pharmaceutical industry *see* **Big Pharma**
phthalates ref1
Plato's Allegory of the Cave ref1, ref2
pneuma ref1
police
  Black Lives Matter (BLM) ref1
  brutality ref1
  citizen's arrests ref1, ref2
  common law arrests ref1, ref2

  - Common Purpose ref1
  - defunding ref1
  - lockdowns ref1, ref2
  - masks ref1, ref2, ref3, ref4
  - police-military state ref1, ref2, ref3
  - psychopathic personality ref1, ref2, ref3, ref4
  - reframing ref1
  - United States ref1, ref2, ref3, ref4
  - Wokeness ref1

polio ref1

political correctness ref1, ref2, ref3, ref4

political parties ref1, ref2, ref3, ref4

political puppets ref1

pollution ref1, ref2, ref3

post-mortems/autopsies ref1

Postage Stamp Consensus ref1, ref2

pre-emptive programming ref1

Problem-Reaction-Solution ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8

Project for the New American Century ref1, ref2, ref3, ref4

psychopathic personality ref1
  - Archons ref1
  - heart energy ref1
  - lockdowns ref1, ref2, ref3
  - police ref1, ref2, ref3, ref4
  - recruitment ref1, ref2
  - vaccines ref1
  - wealth ref1
  - Wetiko ref1, ref2

Psyop (psychological operation), Covid as a ref1, ref2, ref3, ref4, ref5

Pushbackers ref1, ref2, ref3, ref4

pyramid structure ref1, ref2, ref3, ref4

Q

QAnon Psyop ref1, ref2, ref3

R

racism *see also* **Black Lives**
  Matter (BLM)
  anti-racism industry ref1
  class ref1
  critical race theory ref1
  culture ref1
  intersectionality ref1
  reverse racism ref1
white privilege ref1, ref2
white supremacy ref1, ref2, ref3, ref4, ref5
  Wokeness ref1, ref2, ref3
radiation ref1, ref2
randomness, illusion of ref1, ref2, ref3
reality ref1, ref2, ref3
reframing ref1, ref2
  change agents ref1, ref2
  children ref1
  climate change ref1
  Common Purpose leadership programme ref1, ref2
  contradictory rules ref1
  enforcers ref1
  masks ref1, ref2
  NLP and the Delphi technique ref1
  police ref1
  Wetiko factor ref1
  Wokeness ref1, ref2
religion *see also* particular religions
  alien invasions ref1

  - Archons ref1, ref2
  - consciousness ref1, ref2
  - control, system of ref1, ref2, ref3
  - criticism, prohibition on ref1
  - five senses ref1
  - good and evil, war between ref1
  - hidden non-human forces ref1, ref2
  - Sabbatians ref1
  - save me syndrome ref1
  - Wetiko ref1
  - Wokeness ref1

repetition and mind control ref1, ref2, ref3
reporting/snitching, encouragement of ref1, ref2
Reptilians/Grey entities ref1
rewiring the mind ref1
Rivers, Thomas Milton ref1, ref2
Rockefeller family ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9
Rockefeller Foundation documents ref1, ref2, ref3, ref4
Roman Empire ref1
Rothschild family ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9
RT-PCR tests ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
Russia
  - collusion inquiry in US ref1

Russian Revolution ref1, ref2
Sabbatians ref1

S

Sabbantian-Frankism ref1, ref2
  - anti-Semitism ref1, ref2
  - banking and finance ref1, ref2, ref3
  - China ref1, ref2
  - Israel ref1, ref2, ref3, ref4, ref5

- Judaism ref1, ref2, ref3, ref4, ref5
- Lucifer ref1
- media ref1, ref2
- Nazis ref1, ref2
- QAnon ref1
- Rothschilds ref1, ref2, ref3, ref4, ref5, ref6
- Russia ref1
- Saudi Arabia ref1
- Silicon Valley ref1
- Sumer ref1
- United States ref1, ref2, ref3
- Wetiko factor ref1
- Wokeness ref1, ref2, ref3

SAGE (Scientific Advisory Group for Emergencies) ref1, ref2, ref3, ref4

SARS-1 ref1

SARs-CoV-2 ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8

Satan/Satanism ref1, ref2, ref3, ref4, ref5, ref6, ref7

satellites in low-orbit ref1

Saudi Arabia ref1

Save Me Syndrome ref1

scapegoating ref1

Schwab, Klaus ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11, ref12

science, manipulation of ref1

self-identity ref1, ref2, ref3, ref4

self-respect, attacks on ref1

September 11, 2001, terrorist attacks on United States ref1, ref2, ref3, ref4

77th Brigade of UK military ref1, ref2, ref3

Silicon Valley/tech giants ref1, ref2, ref3, ref4, ref5, ref6 *see also* **Facebook**

  Israel ref1
  Sabbatians ref1
  technocracy ref1
  Wetiko factor ref1
  Wokeness ref1
simulation hypothesis ref1, ref2, ref3, ref4, ref5
Smart Grid ref1, ref2, ref3
  artificial intelligence (AI) ref1
  China ref1, ref2
  control centres ref1
  the Field ref1
  Great Reset ref1
  Human 2.0 ref1, ref2
  Israel ref1, ref2
  vaccines ref1
  Wetiko factor ref1
social disapproval ref1
social distancing *and* **isolation** ref1, ref2, ref3
  abusive relationships ref1, ref2
  children ref1
  flats and apartments ref1
  heart issues ref1
  hugs ref1
  Internet ref1
  masks ref1
  media ref1
  older people ref1, ref2
  one-metre (three feet) rule ref1
  rewiring the mind ref1
  simulation, universe as a ref1
  SPI-B ref1
  substance abuse ref1

  suicide and self-harm ref1, ref2, ref3, ref4, ref5
  technology ref1
  torture, as ref1, ref2
  two-metre (six feet) rule ref1
  women ref1
social justice ref1, ref2, ref3, ref4
social media *see also* **Facebook bans on alternative views** ref1
  censorship ref1, ref2, ref3, ref4, ref5, ref6
  children ref1
  emotion ref1
  perception ref1
  private messages ref1
  Twitter ref1, ref2, ref3, ref4, ref5, ref6, ref7
  Wetiko factor ref1
  YouTube ref1, ref2, ref3, ref4, ref5
Soros, George ref1, ref2, ref3, ref4, ref5, ref6
Spain ref1
SPI-B (Scientific Pandemic Insights Group on Behaviours) ref1, ref2, ref3, ref4
spider and the web ref1, ref2, ref3, ref4
Starmer, Keir ref1
Statute Law ref1
Steiner, Rudolf ref1, ref2, ref3
Stockholm syndrome ref1
streptomycin ref1
suicide and self-harm ref1, ref2, ref3, ref4, ref5
Sumer ref1, ref2
Sunstein, Cass ref1, ref2, ref3
swine flu (H1N1) ref1, ref2, ref3
synchronicity ref1
synthetic biology ref1, ref2, ref3, ref4
synthetic meat ref1, ref2

T

technology *see also* **artificial intelligence (AI); Internet;**
  social media addiction ref1, ref2, ref3, ref4
  Archons ref1, ref2
  the cloud ref1, ref2, ref3, ref4, ref5, ref6, ref7
cyber-operations ref1
cyberwarfare ref1
  radiation ref1, ref2
  social distancing and isolation ref1
  technocracy ref1
Tedros Adhanom Ghebreyesus ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11, ref12, ref13
telepathy ref1
Tenpenny, Sherri ref1
Tesla, Nikola ref1
testosterone levels, decrease in ref1
testing for Covid-19 ref1, ref2
  anal swab tests ref1
  cancer ref1
  China ref1, ref2, ref3
  Corman-Drosten test ref1, ref2, ref3, ref4
  death certificates ref1, ref2
  fraudulent testing ref1
  genetic material, amplification of ref1
  Lateral Flow Device (LFD) ref1
  PCR tests ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8
  vaccines ref1, ref2, ref3
Thunberg, Greta ref1, ref2, ref3
Totalitarian Tiptoe ref1, ref2, ref3, ref4
transgender persons
  activism ref1
  artificial wombs ref1

censorship ref1
child abuse ref1, ref2
Human 2.0 ref1, ref2, ref3
Wokeness ref1, ref2, ref3, ref4, ref5
women, deletion of rights and status of ref1, ref2
young persons ref1
travel restrictions ref1
Trudeau, Justin ref1, ref2, ref3
Trump, Donald ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11
Twitter ref1, ref2, ref3, ref4, ref5, ref6, ref7

U

UKColumn ref1, ref2
United Nations (UN) ref1, ref2, ref3, ref4, ref5 *see also* **Agenda 21/Agenda 2030 (UN)**
United States ref1, ref2
American Revolution ref1
borders ref1, ref2
Capitol Hill riot ref1, ref2
children ref1
China ref1, ref2
CIA ref1, ref2
Daily Pass tracking system ref1
demographics by immigration, changes in ref1
Democrats ref1, ref2, ref3, ref4, ref5, ref6, ref7
election fraud ref1
far-right domestic terrorists, pushbackers as ref1
Federal Reserve ref1
flu/respiratory diseases statistics ref1
Global Cult ref1, ref2
hand sanitisers, FDA warnings on ref1

  immigration, effects of illegal ref1
  impeachment ref1
  Israel ref1, ref2
  Judaism ref1, ref2, ref3
  lockdown ref1
  masks ref1
  mass media ref1, ref2
  nursing homes ref1
  Pentagon ref1, ref2, ref3, ref4
  police ref1, ref2, ref3, ref4
  pushbackers ref1
  Republicans ref1, ref2
  borders ref1, ref2
Democrats ref1, ref2, ref3, ref4, ref5
  Russia, inquiry into collusion with ref1
  Sabbatians ref1, ref2, ref3
  September 11, 2001, terrorist attacks ref1, ref2, ref3, ref4
  UFO sightings, release of information on ref1
  vaccines ref1
  white supremacy ref1, ref2, ref3, ref4
  Woke Democrats ref1, ref2

V

vaccines ref1, ref2, ref3
  adverse reactions ref1, ref2, ref3, ref4, ref5
  Africa ref1
  anaphylactic shock ref1, ref2, ref3, ref4
  animals ref1, ref2
  anti-vax movement ref1, ref2, ref3, ref4, ref5
  AstraZeneca/Oxford ref1, ref2, ref3, ref4
  autoimmune diseases, rise in ref1, ref2
  Big Pharma ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8

bioweapon, as real ref1, ref2
black and ethnic minority communities ref1
blood clots ref1, ref2
Brain Computer Interface (BCI) ref1
care homes, deaths in ref1
censorship ref1, ref2, ref3
chief medical officers and scientific advisers, financial interests of ref1, ref2
children ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
China ref1, ref2
clinical trials ref1, ref2, ref3, ref4, ref5, ref6
compensation ref1
compulsory vaccinations ref1, ref2, ref3
computer programs ref1
consciousness ref1
cover-ups ref1
creation before Covid ref1
cytokine storm ref1
deaths and illnesses caused by vaccines ref1, ref2, ref3, ref4, ref5
definition ref1
developing countries ref1
digital tattoos ref1
DNA-manipulation ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
emergency approval ref1, ref2, ref3, ref4, ref5
female infertility ref1
funding ref1
genetic suicide ref1
Global Cult ref1
heart chakras ref1
hesitancy ref1
Human 2.0 ref1, ref2, ref3, ref4
immunity from prosecution ref1, ref2, ref3

  implantable technology ref1
  Israel ref1
  Johnson & Johnson ref1, ref2, ref3, ref4
  lockdowns ref1
  long-term effects ref1
  mainstream media ref1, ref2, ref3, ref4, ref5
  masks ref1, ref2, ref3, ref4, ref5
  Medicines and Healthcare products Regulatory Agency (MHRA) ref1, ref2
  messaging ref1
  Moderna ref1, ref2, ref3, ref4, ref5, ref6
  mRNA vaccines ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9
  nanotechnology ref1, ref2
  NHS ref1, ref2, ref3, ref4, ref5
  older people ref1, ref2
  operating system ref1
  passports ref1, ref2, ref3, ref4
  Pfizer/BioNTech ref1, ref2, ref3, ref4, ref5, ref6, ref7
polyethylene glycol ref1
pregnant women ref1
psychopathic personality ref1
races, targeting different ref1
  reverse transcription ref1
  Smart Grid ref1
  social distancing ref1
  social media ref1
  sterility ref1
  synthetic material, introduction of ref1
  tests ref1, ref2, ref3
  travel restrictions ref1
  variants ref1, ref2
  viruses, existence of ref1
  whistle-blowing ref1

  - WHO ref1, ref2, ref3, ref4
  - Wokeness ref1
  - working, vaccine as ref1
  - young people ref1

Vallance, Patrick ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9

variants ref1, ref2, ref3

vegans ref1

ventilators ref1, ref2

virology ref1, ref2

virtual reality ref1, ref2, ref3

viruses, existence of ref1

visual reality ref1, ref2

vitamin D ref1, ref2

von Braun, Wernher ref1, ref2

W

war-zone hospital myths ref1

waveforms ref1, ref2

wealth ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9 ref10, ref11

wet market conspiracy ref1

Wetiko factor ref1
  - alcoholism and drug addiction ref1
  - anti-human, why Global Cult is ref1
  - Archons ref1, ref2, ref3, ref4
  - artificial intelligence (AI) ref1
  - Big Pharma ref1, ref2
  - children ref1
  - China ref1
  - consciousness ref1, ref2
  - education ref1
  - Facebook ref1

  - fear ref1, ref2
  - frequency ref1, ref2
  - Gates ref1, ref2
  - Global Cult ref1, ref2
  - heart ref1, ref2
  - lockdowns ref1
  - masks ref1
  - Native American concept ref1
  - psychopathic personality ref1, ref2
  - reframing/retraining programmes ref1
  - religion ref1
  - Silicon Valley ref1
  - Smart Grid ref1
  - smartphone addiction ref1, ref2
  - social media ref1
  - war ref1, ref2
  - WHO ref1
  - Wokeness ref1, ref2, ref3
  - Yaldabaoth ref1, ref2, ref3, ref4
- **whistle-blowing** ref1, ref2, ref3, ref4, ref5, ref6, ref7
- **white privilege** ref1, ref2
- **white supremacy** ref1, ref2, ref3, ref4, ref5
- **Whitty, Christopher** ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10
- **‘who benefits’** ref1
- **Wi-Fi** ref1, ref2, ref3, ref4
- **Wikipedia** ref1, ref2
- **Wojcicki, Susan** ref1, ref2, ref3, ref4, ref5, ref6, ref7
- **Wokeness**
  - Antifa ref1, ref2, ref3, ref4
  - anti-Semitism ref1
  - billionaire social justice warriors ref1, ref2, ref3

  - Capitol Hill riot ref1, ref2
  - censorship ref1
  - Christianity ref1
  - climate change hoax ref1, ref2
  - culture ref1
  - education, control of ref1
  - emotion ref1
  - facts ref1
  - fascism ref1, ref2, ref3
  - Global Cult ref1, ref2, ref3, ref4
  - group-think ref1
  - immigration ref1
  - indigenous people, solidarity with ref1
  - inversion ref1, ref2, ref3
  - left, hijacking the ref1, ref2
- Marxism ref1, ref2, ref3
- mind control ref1
- New Woke ref1
- Old Woke ref1
- Oneness ref1
- perceptual programming ref1
  - Phantom Self ref1
- police ref1
- defunding the ref1
- reframing ref1
- public institutions ref1
- Pushbackers ref1, ref2, ref3
- racism ref1, ref2, ref3
- reframing ref1, ref2
- religion, as ref1
- Sabbatians ref1, ref2, ref3
- Silicon Valley ref1
- social justice ref1, ref2, ref3, ref4

transgender ref1, ref2, ref3, ref4, ref5
United States ref1, ref2
  vaccines ref1
  Wetiko factor ref1, ref2, ref3
young people ref1, ref2, ref3
women, deletion of rights and status of ref1, ref2
World Economic Forum (WEF) ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9
World Health Organization (WHO) ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9
  AIDs/HIV ref1
  amplification cycles ref1
  Big Pharma ref1, ref2, ref3
  cooperation in health emergencies ref1
  creation ref1, ref2
  fatality rate ref1
  funding ref1, ref2, ref3
  Gates ref1
  Internet ref1
  lockdown ref1
  vaccines ref1, ref2, ref3, ref4
  Wetiko factor ref1
world number 1 (masses) ref1, ref2
world number 2 ref1
Wuhan ref1, ref2, ref3, ref4, ref5, ref6, ref7 ref8

Y

Yaldabaoth ref1, ref2, ref3, ref4, ref5, ref6
Yeadon, Michael ref1, ref2, ref3, ref4
young people *see also* **children** addiction to technology ref1
  Human 2.0 ref1
  vaccines ref1, ref2

Wokeness ref1, ref2, ref3

YouTube ref1, ref2, ref3, ref4, ref5

WHO 548

Z

Zaks, Tal ref1

Zionism ref1, ref2, ref3

Zuckerberg, Mark ref1, ref2, ref3, ref4, ref5, ref6, ref7, ref8, ref9, ref10, ref11, ref12

Zulus ref1

ICKONIC
THE ALTERNATIVE

Ickonic is something that has been a dream of mine for the last 5 years, growing up around alternative information I have always had a natural interest in what is going on in the World and what could I do to make it better.
Across the range of subjects and positions of influence occupied mainly by people who don't strive to make things better it's the Media that I have always found the most frustrating and fascinating. Mainly because if the Media did their Jobs properly then so much of the negative things happening in the World simply would not be able to happen, because they would be exposed within a heartbeat.
Free Press and the Opportunities that the internet could have given would mean that the Media are able to expose things like never before and hold people to account for their actions. As we all know there are 'Untouchables' that walk among us, people the Media simply won't touch, expose or investigate and that leads to the dark underworlds that infest the establishment the World over.
Well I say enough, it's time for something different, a different kind of Media, where no one is off limits from exposing and investigating. All we're interested in at Ickonic is the truth of what is really going on in the World on whichever subject we're covering.
We hope you enjoy what we have created and take something away from the platform, we aim to deliver information that's informative and most importantly self-empowering, you're not a little person, you're part of something much bigger than that and its time we as a collective race began to understand that and look to the future as ours to take.

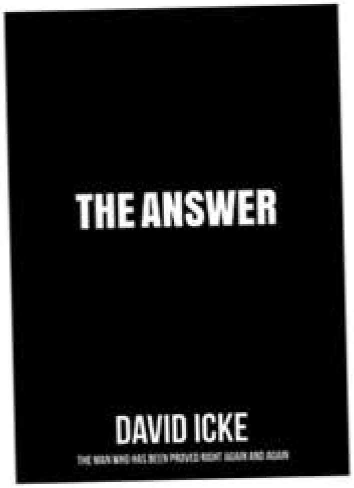
It's time...

Jaymie Icke - Founder Ickonic Alternative Media.

SIGN UP NOW AT ICKONIC.COM

DAVID ICKE

THE ANSWER



THE ANSWER

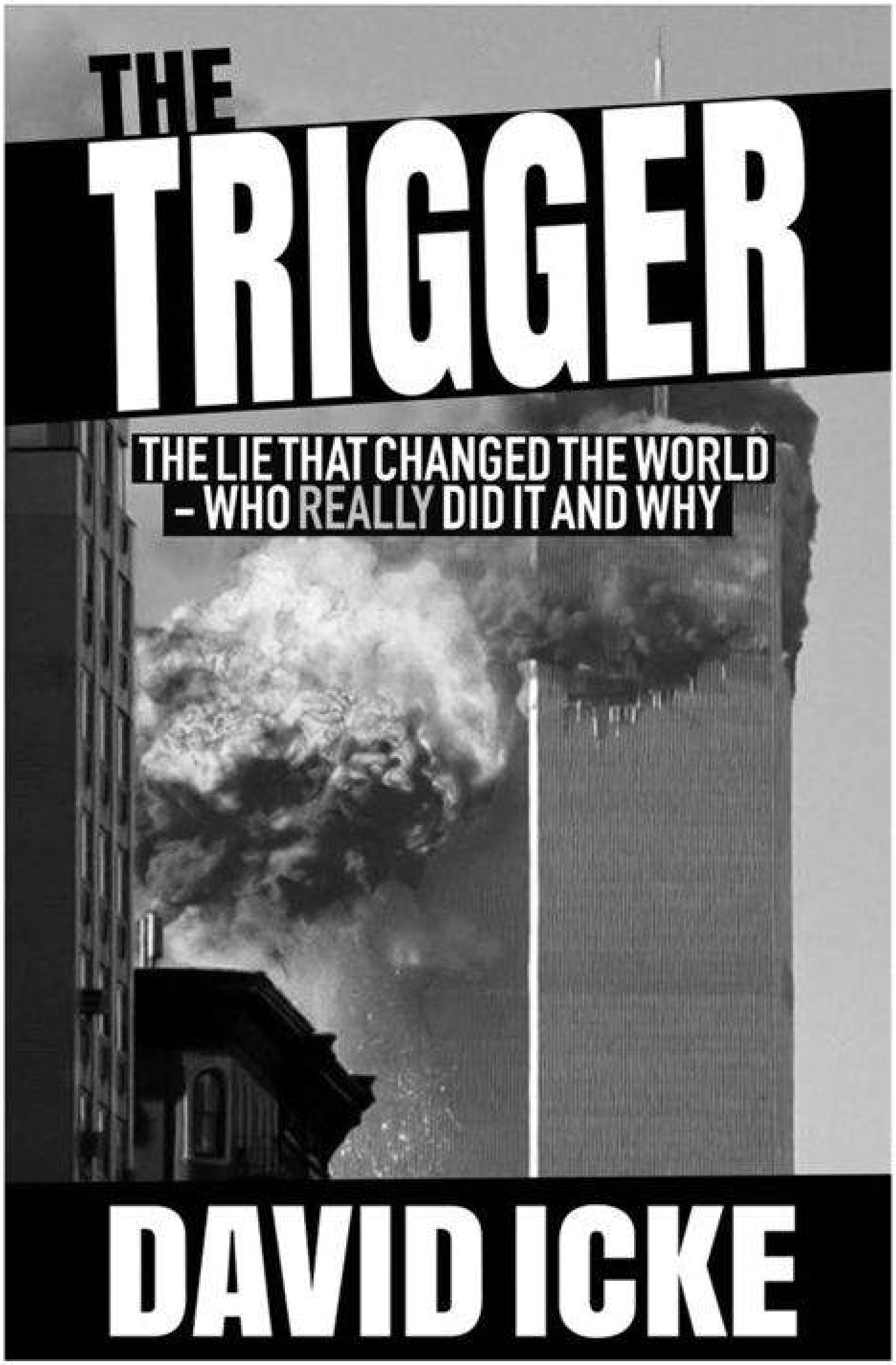
DAVID ICKE

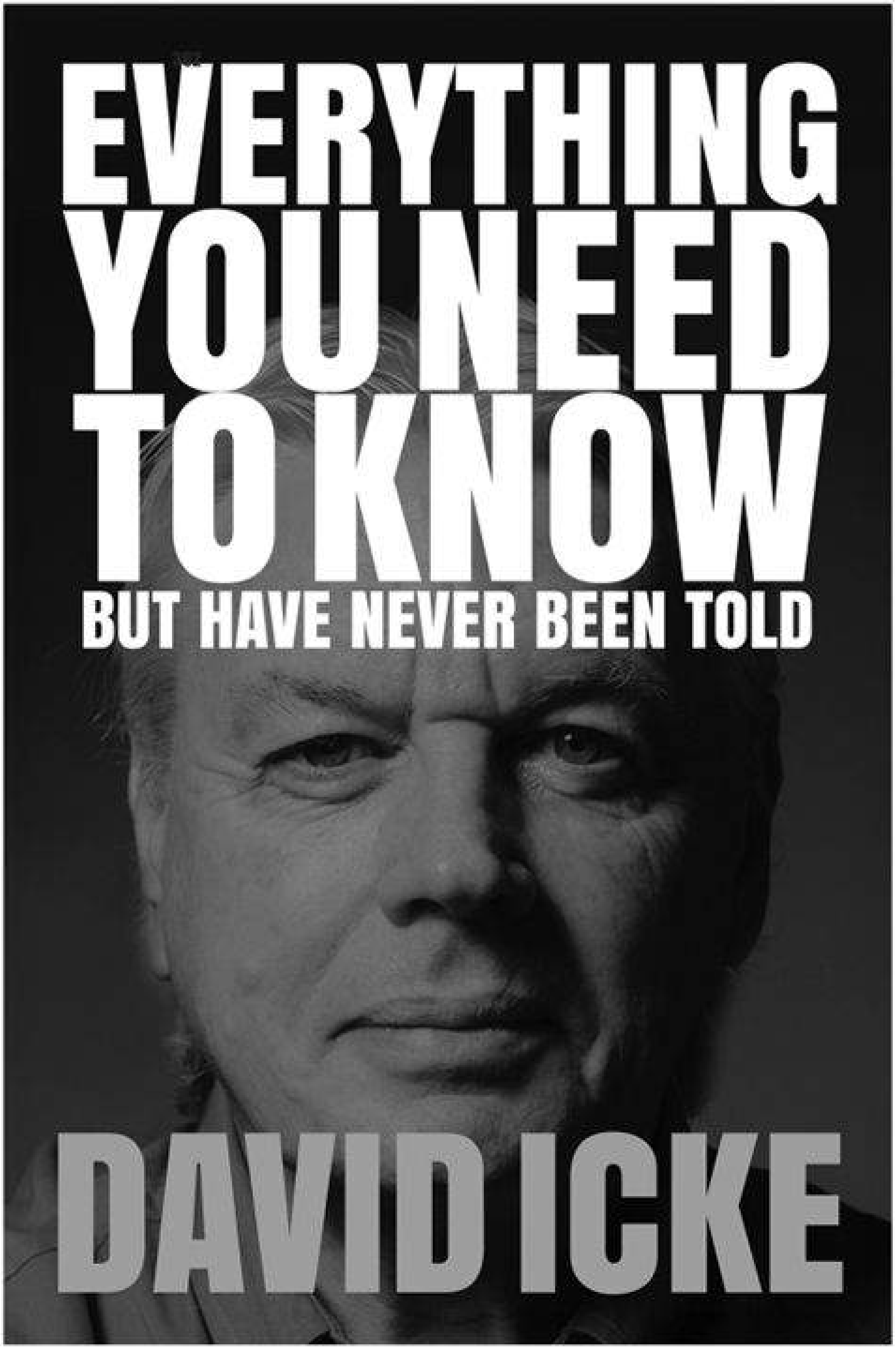
THE MAN WHO HAS BEEN PROVED RIGHT AGAIN AND AGAIN

We live in extraordinary times with billions bewildered and seeking answers for what is happening. David Icke, the man who has been proved right again and again, has spent 30 years uncovering the truth behind world affairs and in a stream of previous books he predicted current events.

The Answer will change your every perception of life and the world and set you free of the illusions that control human society. There is nothing more vital for our collective freedom than humanity becoming aware of what is in this book.

Available now at davidicke.com.


THE
TRIGGER
THE LIE THAT CHANGED THE WORLD
– WHO REALLY DID IT AND WHY
DAVID ICKE



EVERYTHING YOU NEED TO KNOW

BUT HAVE NEVER BEEN TOLD

DAVID ICKE

DAVIDICKE.COM

DAVID ICKE STORE
LATEST NEWS ARTICLES
DAVID ICKE VIDEOS
WEEKLY DOT-CONNECTOR PODCASTS
LIVE EVENTS
WWW.DAVIDICKE.COM

THE LIFE STORY OF DAVID ICKE

RENEGADE

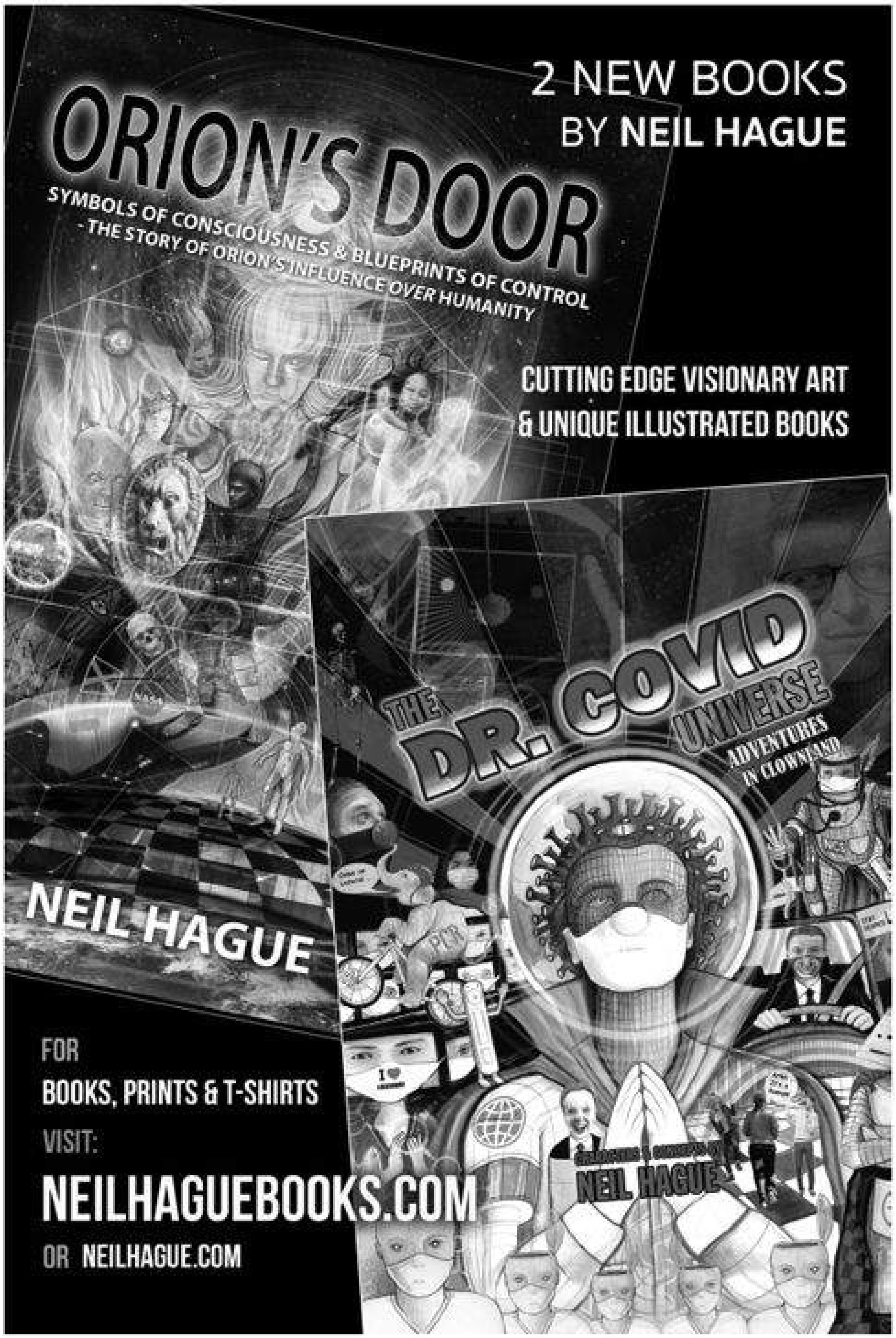
THE FEATURE LENGTH FILM

/ˈren·ɪˌɡeɪd/

noun

A person who behaves in a rebelliously unconventional manner.

AVAILABLE NOW AT DAVIDICKE.COM



ORION'S DOOR
SYMBOLS OF CONSCIOUSNESS & BLUEPRINTS OF CONTROL
- THE STORY OF ORION'S INFLUENCE OVER HUMANITY
NEIL HAGUE
THE DR. COVID UNIVERSE
ADVENTURES IN CLOWNLAND
NEIL HAGUE

2 NEW BOOKS
BY **NEIL HAGUE**

CUTTING EDGE VISIONARY ART
& UNIQUE ILLUSTRATED BOOKS

FOR
BOOKS, PRINTS & T-SHIRTS
VISIT:
NEILHAGUEBOOKS.COM
OR NEILHAGUE.COM

Before you go ...

For more detail, background and evidence about the subjects in *Perceptions of a Renegade Mind* – and so much more – see my others books including *And The Truth Shall Set You Free; The Biggest Secret; Children of the Matrix; The David Icke Guide to the Global Conspiracy; Tales from the Time Loop; The Perception Deception; Remember Who You Are; Human Race Get Off Your Knees; Phantom Self; Everything You Need To Know But Have Never Been Told, The Trigger and The Answer*.

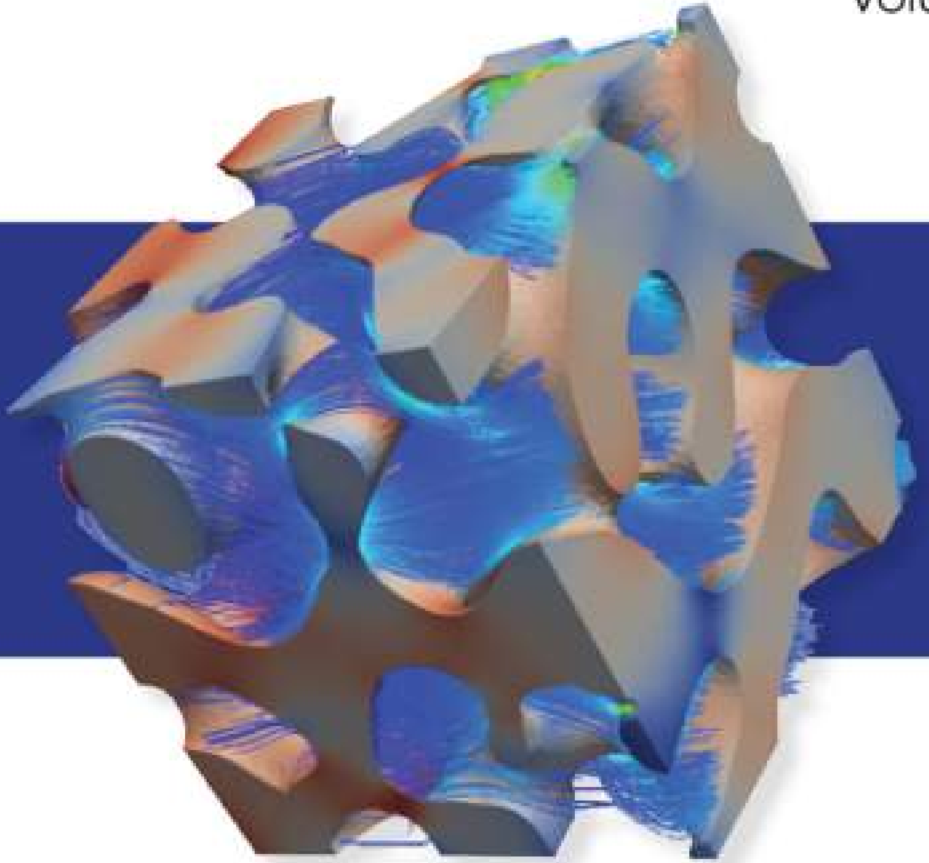
You can subscribe to the fantastic new Ickonic media platform where there are many hundreds of hours of cutting-edge information in videos, documentaries and series across a whole range of subjects which are added to every week. This includes my 90 minute breakdown of the week's news every Friday to explain *why* events are happening and to what end.

SERIES IN OPTICS AND OPTOELECTRONICS

Handbook of Optoelectronic Device Modeling & Simulation

Lasers, Modulators, Photodetectors, Solar Cells, and Numerical Methods

Volume II



Edited by

Joachim Piprek

CRC Press
Taylor & Francis Group

Handbook of Optoelectronic Device Modeling and Simulation

Series in Optics and Optoelectronics

Series Editors: **Robert G W Brown**
University of California, Irvine, USA

E Roy Pike
Kings College, London, UK

Recent titles in the series

Optical MEMS, Nanophotonics, and Their Applications
Guangya Zhou and Vincent C. Lee (Eds.)

Optical Microring Resonators: Theory, Techniques, and Applications
Vien Van

Thin-Film Optical Filters, Fifth Edition
H. Angus Macleod

Handbook of Optoelectronics, Second Edition: Concepts, Devices, and Techniques – Volume One
John P. Dakin and Robert Brown (Eds.)

Handbook of Optoelectronics, Second Edition: Enabling Technologies – Volume Two
John P. Dakin and Robert Brown (Eds.)

Handbook of Optoelectronics, Second Edition: Applied Optical Electronics – Volume Three
John P. Dakin and Robert Brown (Eds.)

Handbook of GaN Semiconductor Materials and Devices
Wengang (Wayne) Bi, Hao-chung (Henry) Kuo, Pei-Cheng Ku, and Bo Shen (Eds.)

Handbook of Optoelectronic Device Modeling and Simulation: Fundamentals, Materials, Nanostructures, LEDs, and Amplifiers – Volume One
Joachim Piprek (Ed.)

Handbook of Optoelectronic Device Modeling and Simulation: Lasers, Modulators, Photodetectors, Solar Cells, and Numerical Methods – Volume Two
Joachim Piprek (Ed.)

Nanophotonics and Plasmonics: An Integrated View
Dr. Ching Eng (Jason) Png and Dr. Yuriy Akimov

Handbook of Solid-State Lighting and LEDs
Zhe Chuan Feng (Ed.)

Optical Compressive Imaging
Adrian Stern

Singular Optics
Gregory J. Gbur

Handbook of Optoelectronic Device Modeling and Simulation

Lasers, Modulators, Photodetectors, Solar Cells, and Numerical Methods

VOLUME TWO

Edited by
Joachim Piprek

CRC Press
Taylor & Francis Group
Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This book's use or discussion of MATLAB software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB software.

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2018 by Taylor & Francis Group, LLC

CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works

Printed on acid-free paper

International Standard Book Number-13: 978-1-4987-4956-5 (Hardback)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (http://www.copyright.com/) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Library of Congress Cataloging-in-Publication Data

Names: Piprek, Joachim, editor.
Title: Handbook of optoelectronic device modeling and simulation / edited by Joachim Piprek.
Other titles: Series in optics and optoelectronics (CRC Press) ; 27.
Description: Boca Raton, FL : CRC Press, Taylor & Francis Group, [2017] |
Series: Series in optics and optoelectronics ; 27 | Includes bibliographical references and index.
Contents: volume 1. Fundamentals, materials, nanostructures, LEDS, and amplifiers – volume 2.
Lasers, modulators, photodetectors, solar cells, and numerical methods.
Identifiers: LCCN 2016058063 | ISBN 9781498749466 (v. 1; hardback ; alk. paper) |
ISBN 1498749461 (v. 1; hardback ; alk. paper) | ISBN 9781498749565 (v. 2 ; hardback ; alk. paper) |
ISBN 1498749569 (v. 2 ; hardback ; alk. paper)
Subjects: LCSH: Optoelectronic devices–Mathematical models–Handbooks, manuals, etc. |
Optoelectronic devices–Simulation methods–Handbooks, manuals, etc. |
Semiconductors–Handbooks, manuals, etc. | Nanostructures–Handbooks, manuals, etc.
Classification: LCC TK8304 .H343 2017 | DDC 621.381/045011–dc23
LC record available at https://lccn.loc.gov/2016058063

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Contents

Series Preface ix

Preface xi

Editor xiii

Contributors xv

PART VI Laser Diodes

26 Laser Diode Fundamentals 3
Joachim Piprek

27 High-Power Lasers 15
Hans Wenzel and Anissa Zeghuzi

28 High-Brightness Tapered Lasers 59
Ignacio Esquivias, Antonio Pérez-Serrano, and José-Manuel G. Tijero

29 High-Brightness Laser Diodes with External Feedback 81
Mohamad Anas Helal, Simeon N. Kaunga-Nyirenda, Steve Bull, and Eric Larkins

30 Single Longitudinal Mode Laser Diodes 109
Xun Li

31 Traveling Wave Modeling of Nonlinear Dynamics in Multisection Laser Diodes ... 153
Mindaugas Radziunas

32 Mode-Locked Semiconductor Lasers 183
Eugene Avrutin and Julien Javaloyes

33 Quantum Cascade Lasers: Electrothermal Simulation 235
Song Mei, Yanbing Shi, Olafur Jonasson, and Irena Knezevic

34 Vertical-Cavity Surface-Emitting Lasers 261
Tomasz Czyszanowski, Leszek Frasunkiewicz, and Maciej Dems

PART VII Photodetectors and Modulators

35 Photodetector Fundamentals 283
Prasanta Basu

36 P-N Junction Photodiodes 307
Weida Hu

37 Quantum Well Infrared Photodetectors 337
Kwong-Kit Choi

38 Optical Modulators 363
Dominic F.G. Gallagher and Dmitry Labukhin

PART VIII Solar Cells

39 Solar Cell Fundamentals 383
Matthias Müller

40 Multijunction Solar Cells 415
Matthew Wilkins and Karin Hinzer

41 Nanostructure Solar Cells 441
Urs Aeberhard

42 Nanowire Solar Cells: Electro-Optical Performance 475
Bernd Witzigmann

43 Thin-Film Solar Cells 497
Matthias Auf der Maur, Tim Albes, and Alessio Gagliardi

PART IX Novel Applications

44 Electroluminescent Refrigerators 541
Kuan-Chen Lee and Shun-Tung Yen

45 Photonic Crystal Laser Diodes 561
Maciej Dems

46 Single-Photon Sources 585
Niels Gregersen, Dara P. S. McCutcheon, and Jesper Mørk

47 Nanoplasmonic Lasers and Spasers 609
A. Freddie Page and Ortwin Hess

48 Quantum-Dot Nanolasers 627
Christopher Gies, Michael Lorke, Frank Jahnke, and Weng W. Chow

49 Nonlinear Dynamics in Quantum Photonic Structures 661
Gabriela Slavcheva and Mirella Koleva

PART X Mathematical Methods

50 Drift-Diffusion Models 733
Patricio Farrell, Nella Rotundo, Duy Hai Doan, Markus Kantner, Jürgen Fuhrmann, and Thomas Koprucki

51 Monte Carlo Device Simulations 773
Katerina Raleva, Abdul R. Shaik, Raghuraj Hathwar, Akash Laturia, Suleman S. Qazi, Robin Daugherty, Dragica Vasileska, and Stephen M. Goodnick

52 Photonics 807
Frank Schmidt

Index 853

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Series Preface

This international series covers all aspects of theoretical and applied optics and optoelectronics. Active since 1986, eminent authors have long been choosing to publish with this series, and it is now established as a premier forum for high-impact monographs and textbooks. The editors are proud of the breadth and depth showcased by the published works, with levels ranging from advanced undergraduate and graduate student texts to professional references. Topics addressed are both cutting edge and fundamental, basic science and applications-oriented, on subject matter that includes lasers, photonic devices, nonlinear optics, interferometry, waves, crystals, optical materials, biomedical optics, optical tweezers, optical metrology, solid-state lighting, nanophotonics, and silicon photonics. Readers of the series are students, scientists, and engineers working in optics, optoelectronics, and related fields in the industry.

Proposals for new volumes in the series may be directed to Lu Han, senior publishing editor at CRC Press/Taylor & Francis Group (lu.han@taylorandfrancis.com).

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Preface

Optoelectronic devices have become ubiquitous in our daily lives. For example, light-emitting diodes (LEDs) are used in almost all household appliances, in traffic and streetlights, and in full-color displays. Laser diodes, optical modulators, and photodetectors are key components of the Internet. Solar cells are core elements of energy supply systems. Optoelectronic devices are typically based on nanoscale semiconductor structures that utilize the interaction of electrons and photons. The underlying and highly complex physical processes require mathematical models and numerical simulation for device design, analysis, and performance optimization. This handbook gives an introduction to modern optoelectronic devices, models, and simulation methods.

Driven by the expanding diversity of available and envisioned practical applications, mathematical models and numerical simulation software for optoelectronic devices have experienced a rapid development in recent years. In the past, advanced modeling and simulation was the domain of a few specialists using proprietary software in computational research groups. The increasing user-friendliness of commercial software now also opens the door for nontheoreticians and experimentalists to perform sophisticated modeling and simulation tasks. However, the ever-growing variety and complexity of devices, materials, physical mechanisms, theoretical models, and numerical techniques make it often difficult to identify the best approach to a given project or problem. This book presents an up-to-date review of optoelectronic device models and numerical techniques. The handbook format is ideal for beginners but also gives experienced researchers an opportunity to renew and broaden their knowledge in this expanding field.

Semiconductors are the key material of optoelectronic devices, as they enable propagation and interaction of electrons and photons. The handbook starts with an overview of fundamental semiconductor device models, which apply to almost all device types, followed by sections on novel materials and nanostructures. The main part of the handbook is ordered by device type (LED, amplifier, laser diode, photodetector, and solar cell). For each device type, an introductory chapter is followed by chapters on specialized device designs and applications, describing characteristic effects and models. Finally, novel device concepts and applications are reviewed. At the end of the handbook, an overview of numerical techniques is provided, both for electronic and photonic simulations.

I would like to thank the publisher for initiating this important handbook project and for giving me the opportunity to serve as editor. Many years of organizing the annual international conference on *Numerical Simulation of Optoelectronic Devices (NUSOD)* enabled me to attract a large number of experts from all over the world to write handbook chapters on their research area. I sincerely thank all authors for their valuable contributions.

Joachim Piprek
Newark, Delaware, USA

MATLAB$^{®}$ is a registered trademark of The MathWorks, Inc. For product information, please contact:

The MathWorks, Inc.
3 Apple Hill Drive
Natick, MA 01760-2098 USA
Tel: 508-647-7000
Fax: 508-647-7001
Email: info@mathworks.com
Web: www.mathworks.com

Editor

Joachim Piprek received his diploma and doctoral degrees in physics from the Humboldt University in Berlin, Germany. For more than two decades, he worked in industry and academia on modeling, simulation, and analysis of various semiconductor devices used in optoelectronics. Currently, he serves as president of the NUSOD Institute, Newark, Delaware (see http://www.nusod.org). During his previous career in higher education, Dr. Piprek taught various graduate courses at universities in Germany, Sweden, and the United States. Since 2001, he has been organizing the annual international conference on *Numerical Simulation of Optoelectronic Devices (NUSOD)*. Thus far, Dr. Piprek has published three books, six book chapters, and about 250 research papers, which have received more than 6000 citations. He was an invited guest editor for several journal issues on optoelectronic device simulation and currently serves as an executive/associate editor of two research journals in this field.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

Contributors

Urs Aeberhard
IEK-5 Photovoltaik
Forschungszentrum Jülich
Jülich, Germany

Tim Albes
Department of Electrical Engineering
Technical University of Munich
Munich, Germany

Matthias Auf der Maur
Department of Electronic Engineering
University of Rome Tor Vergata
Rome, Italy

Eugene Avrutin
University of York
York, United Kingdom

Prasanta Basu
Institute of Radio Physics and Electronics
University of Calcutta
Kolkata, India

Steve Bull
Department of Electrical and Electronic Engineering
University of Nottingham
Nottingham, United Kingdom

Kwong-Kit Choi
Sensors and Electron Devices Directorate
U.S. Army Research Laboratory
Adelphi, Maryland

Weng W. Chow
Sandia National Laboratories,
Albuquerque, USA

Tomasz Czyszanowski
Institute of Physics
Lodz University of Technology
Lodz, Poland

Robin Daugherty
Arizona State University
Tempe, Arizona

Maciej Dems
Institute of Physics
Lodz University of Technology
Lodz, Poland

Duy Hai Doan
Research Group Partial Differential Equations
Weierstrass Institute
Berlin, Germany

Ignacio Esquivias
Center of Advanced Materials and Devices for ICT Applications
Universidad Politécnica de Madrid
Madrid, Spain

Patricio Farrell
Research Group Numerical Mathematics and Scientific Computing
Weierstrass Institute
Berlin, Germany

Leszek Frasunkiewicz
Institute of Physics
Lodz University of Technology
Lodz, Poland

Jürgen Fuhrmann
Research Group Numerical Mathematics and Scientific Computing
Weierstrass Institute
Berlin, Germany

Alessio Gagliardi
Department of Electrical Engineering
Technical University of Munich
Munich, Germany

Dominic F.G. Gallagher
Photon Design Ltd.
Oxford, United Kingdom

Christopher Gies
Institute for Theoretical Physics
University of Bremen
Bremen, Germany

Stephen M. Goodnick
Arizona State University
Tempe, Arizona

Niels Gregersen
Technical University of Denmark
Lyngby, Denmark

Raghuraj Hathwar
Arizona State University
Tempe, Arizona

Mohamad Anas Helal
Department of Electrical and Electronic Engineering
University of Nottingham
Nottingham, United Kingdom

Ortwin Hess
Imperial College
London, United Kingdom

Karin Hinzer
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, Canada

Weida Hu
State Key Laboratory of Infrared Physics
Shanghai Institute of Technical Physics
Shanghai, China

Frank Jahnke
Institute for Theoretical Physics,
University of Bremen
Bremen, Germany

Julien Javaloyes
Universitat de les Illes Balears,
Palma de Mallorca, Spain

Olafur Jonasson
Department of Electrical and Computer Engineering
University of Wisconsin
Madison, Wisconsin

Markus Kantner
Research Group Laser Dynamics
Weierstrass Institute
Berlin, Germany

Simeon N. Kaunga-Nyirenda
Department of Electrical and Electronic Engineering
University of Nottingham
Nottingham, United Kingdom

Irena Knezevic
Department of Electrical and Computer Engineering
University of Wisconsin
Madison, Wisconsin

Mirella Koleva
Department of Physics
University of Oxford
Oxford, United Kingdom

Thomas Koprucki
Research Group Partial Differential Equations
Weierstrass Institute
Berlin, Germany

Dmitry Labukhin
Photon Design Ltd.
Oxford, United Kingdom

Eric Larkins
Department of Electrical and
Electronic Engineering
University of Nottingham
Nottingham, United Kingdom

Akash Laturia
Arizona State University
Tempe, Arizona

Kuan-Chen Lee
National Chiao Tung University
Hsinchu, Taiwan

Xun Li
Department of Electrical and Computer
Engineering
McMaster University
Hamilton, Canada

Michael Lorke
Institute for Theoretical Physics
University of Bremen
Bremen, Germany

Dara P.S. McCutcheon
Quantum Engineering Technologies Labs
University of Bristol
Bristol, United Kingdom

Song Mei
Department of Electrical and Computer
Engineering
University of Wisconsin
Madison, Wisconsin

Jesper Mørk
Technical University of Denmark
Lyngby, Denmark

Matthias Müller
Institute of Applied Physics
Technical University Bergakademie Freiberg
Freiberg, Germany

A. Freddie Page
Imperial College London
London, United Kingdom

Antonio Pérez-Serrano
Center of Advanced Materials and Devices for
ICT Applications
Universidad Politécnica de Madrid
Madrid, Spain

Joachim Piprek
NUSOD Institute LLC
Newark, Delaware

Suleman S. Qazi
Arizona State University
Tempe, Arizona

Mindaugas Radziunas
Research Group *Laser Dynamics*
Weierstrass Institute
Berlin, Germany

Katerina Raleva
University Sts. Cyril and Methodius
Skopje, Macedonia

Nella Rotundo
Research Group Partial Differential
Equations
Weierstrass Institute
Berlin, Germany

Frank Schmidt
Zuse Institute
Berlin, Germany

Abdul R. Shaik
Arizona State University
Tempe, Arizona

Yanbing Shi
Department of Electrical and Computer Engineering
University of Wisconsin
Madison, Wisconsin

Gabriela Slavcheva
Centre for Photonics and Photonic Materials
Department of Physics
University of Bath
Bath, United Kingdom

José-Manuel G. Tijero
Center of Advanced Materials and Devices for ICT Applications
Universidad Politécnica de Madrid
Madrid, Spain

Dragica Vasileska
Arizona State University
Tempe, Arizona

Hans Wenzel
Ferdinand Braun Institute
Berlin, Germany

Matthew Wilkins
School of Electrical Engineering and Computer Science
University of Ottawa
Ottawa, Canada

Bernd Witzigmann
Department of Electrical Engineering and Computer Science
University of Kassel
Kassel, Germany

Shun-Tung Yen
National Chiao Tung University
Hsinchu, Taiwan

Anissa Zeghuzi
Ferdinand Braun Institute
Berlin, Germany

VI

Laser Diodes

26 Laser Diode Fundamentals *Joachim Piprek*....................3
Introduction • Optical Gain and Optical Loss • Threshold Current and Slope Efficiency • Temperature Effects • Efficiency Analysis • Rate Equation Analysis • Basic Laser Cavity Designs

27 High-Power Lasers *Hans Wenzel and Anissa Zeghuzi*....................15
Introduction • Phenomenological Model • Models for the Optical Power • Model for the Optical Field • Models for Nonlinear Modes and Filamentation • Thermodynamic-Based Energy-Transport Model

28 High-Brightness Tapered Lasers *Ignacio Esquivias, Antonio Pérez-Serrano, and José-Manuel G. Tijero*....................59
Introduction • Simulation Model • Simulation Examples • Summary

29 High-Brightness Laser Diodes with External Feedback *Mohamad Anas Helal, Simeon N. Kaunga-Nyirenda, Steve Bull, and Eric Larkins*....................81
Power Scaling and the Role of Beam Quality • High-Brightness Laser Diode Sources • External Cavity Laser Simulation • Case Study: The Impact of Unintentional Reflections on a DBR-Tapered Laser • Conclusions

30 Single Longitudinal Mode Laser Diodes *Xun Li*....................109
Introduction • Grating-Assisted SLM Laser Diodes • Coupled-Cavity Single Longitudinal Mode Laser Diodes • Recent Development on Single Longitudinal Mode Laser Diode

31 Traveling Wave Modeling of Nonlinear Dynamics in Multisection Laser Diodes *Mindaugas Radziunas*....................153
Introduction • Basic TW Model in the Solitary Laser • Model of Material Gain Dispersion • Thermal Detuning • Spatially Inhomogeneous Carrier Density • Nonlinear Gain Saturation • Further Modifications of the TW Model • Multisection Lasers and Coupled Laser Systems • Simulations of Nontrivial MSL Device • Beyond Numerical Simulations of the TW Model • Conclusions

32 Mode-Locked Semiconductor Lasers *Eugene Avrutin and Julien Javaloyes*....................183
General Principles of Mode Locking, the Important Features of Mode-Locked Semiconductor Lasers, and the Role of Theory and Modeling • ML Techniques in Laser Diodes: The Main Features • Theoretical Models of ML in Semiconductor Lasers • The Main Predictions of Mode-Locked Laser Theory • Microscopic and Semimicroscopic Approaches in Mode-Locked Laser Modeling • Some Novel Problems and Challenges in Mode-Locked Laser Modeling • Concluding Remarks

33 Quantum Cascade Lasers: Electrothermal Simulation *Song Mei, Yanbing Shi, Olafur Jonasson, and Irena Knezevic* 235
Introduction • Electronic Transport • Thermal Transport • Device-Level Electrothermal Simulation: An Example • Conclusion

34 Vertical-Cavity Surface-Emitting Lasers *Tomasz Czyszanowski, Leszek Frasunkiewicz, and Maciej Dems* 261
Introduction • Physical Phenomena • Light Confinement in VCSELs • Fundamental Equations • VCSEL Structure • One-Dimensional Scalar Approach • Two-Dimensional Scalar Approach • Three-Dimensional Vectorial Approach • Impact of Oxide Aperture Diameter and Position • Conclusion

26

Laser Diode Fundamentals

26.1 Introduction 3
26.2 Optical Gain and Optical Loss 3
26.3 Threshold Current and Slope Efficiency 6
26.4 Temperature Effects 7
26.5 Efficiency Analysis 8
26.6 Rate Equation Analysis 9
26.7 Basic Laser Cavity Designs 12

Joachim Piprek

26.1 Introduction

Semiconductor laser diodes are key components in optical fiber communication, data storage, sensing, material processing, and other applications. They are based on sophisticated interactions of electrons and photons in semiconductor nanostructures (see Chapter 3). Advanced theoretical models and simulation tools are required for the development and analysis of future generations of laser diodes. The following chapters describe some of these complex models in detail. This introductory chapter is aimed at readers who are not yet familiar with basic models and parameters of semiconductor lasers. By exploring simple analytical models, this chapter tries to develop an intuitive understanding of internal laser physics that will help to digest the more complicated theory outlined in subsequent chapters. While these analytical models are quite popular, they often have limits beyond which numerical simulations are required, as shown below.

Section 26.2 examines basic formulas for optical gain and optical losses in semiconductor lasers, followed by the introduction of key performance parameters, threshold current, and slope efficiency. As practical applications often suffer from undesired self-heating of the laser, temperature effects are discussed in Section 26.4 and the resulting changes in laser efficiency in Section 26.5. The rate equation model for dynamic lasing processes is introduced in Section 26.6, including formulas for small signal analysis. Section 26.7 briefly reviews basic laser cavity designs.

26.2 Optical Gain and Optical Loss

Traveling through a semiconductor, a single photon with an energy close to the band gap is able to generate an identical second photon by stimulating the recombination of an electron–hole pair. This is the basic physical mechanism of lasing. The second photon exhibits the same wavelength and the same phase as of the first photon, doubling the amplitude of their monochromatic wave. Subsequent repetition of this process leads to strong light amplification. However, the competing process is the absorption of photons by the generation of new electron–hole pairs (see Figure 26.1). Stimulated emission prevails when more electrons are present at the higher energy level (conduction band) than at the lower energy level (valence band).

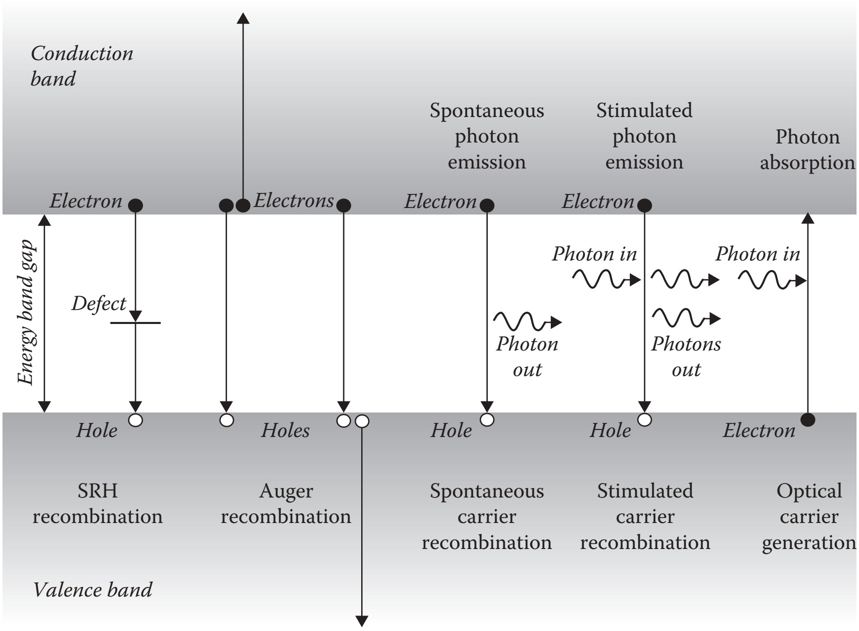

Conduction band
Spontaneous photon emission
Stimulated photon emission
Photon absorption
Electron
Electrons
Electron
Electron
Energy band gap
Defect
Photon in
Photon in
Photon out
Photons out
Hole
Holes
Hole
Hole
Electron
SRH recombination
Auger recombination
Spontaneous carrier recombination
Stimulated carrier recombination
Optical carrier generation
Valence band

FIGURE 26.1 Electron–hole recombination and generation mechanisms in semiconductors. (SRH, Shockley–Read–Hall.)

This population inversion is one of the key requirements for lasing. Semiconductor lasers typically employ *pin* junctions with a thin active layer of lower band gap (Figure 26.2). At forward bias, electrons and holes are collected in the active layer to achieve inversion. Continuous current injection into the device leads to a continuous stimulated emission of photons, but only if enough photons are constantly present in the device to trigger this process. Thus, only part of all photons can be allowed to leave the laser diode as a lasing beam, the rest must be reflected to remain inside the diode and to generate new photons (Figure 26.3). This optical feedback and confinement of photons in an optical resonator is the second basic requirement of lasing.

The light amplification in the active layer is described by the optical gain $g(n, p, T, \lambda, S)$ as a function of the density of electrons n and holes p inside the active layer, the optical wavelength λ (or photon energy), the photon density S, and the temperature T. This gain function is the heart of laser physics, and a realistic calculation may require sophisticated models (see Chapter 3 for details). We here briefly discuss some popular analytical approximations. The linear gain approximation $g(N) = a(N - N_{\mathrm{tr}})$ is often used, employing the transparency density N_{tr} and assuming a fixed differential gain $a = \mathrm{d}g/\mathrm{d}N$ as well as identical densities of electrons and holes ($N = n = p$). For $N = N_{\mathrm{tr}}$, the absorption and gain are the same and the material is transparent. The differential gain $\mathrm{d}g/\mathrm{d}N$ is a key parameter for laser light modulation (which is discussed Section 26.6). It can only be considered constant for small variations of the carrier density since $\mathrm{d}g/\mathrm{d}N$ is known to decline with increasing carrier densities (Figure 26.4). This is described by the more general logarithmic function (Coldren and Corzine 1995)

$$g(N, S) = \frac{g_{\mathrm{o}}}{1 + \varepsilon S} \ln\left(\frac{N + N_{\mathrm{s}}}{N_{\mathrm{tr}} + N_{\mathrm{s}}}\right) \tag{26.1}$$

The gain compression factor ε describes the gain saturation at high photon densities, e.g., due to carrier depletion (g_{o} and N_{s} are fit parameters). All parameters in Equation 26.1 need to be extracted from more fundamental gain models, as measurements are difficult (Shtengel et al. 1998). However, the lack of

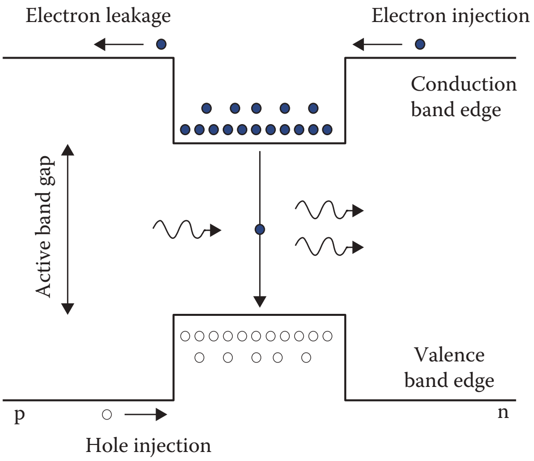

Electron leakage
Electron injection
Conduction band edge
Active band gap
Valence band edge
p
n
Hole injection

FIGURE 26.2 Electron and hole injection into the light-emitting active layer that is sandwiched between a p-doped and an n-doped material of higher energy bandgap. (Adapted from Piprek, J., In S. S. Sun and L.R. Dalton [eds.], *Introduction to Organic Electronic and Optoelectronic Materials and Devices*, Boca Raton, FL: CRC Press, 2008.)

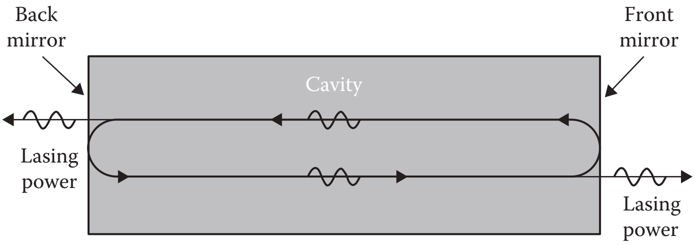

Back mirror
Front mirror
Cavity
Lasing power
Lasing power

FIGURE 26.3 Optical wave propagation in an Fabry–Pérot (FP) laser cavity (resonator) formed between two reflecting facets. (Adapted from Piprek, J., In S. S. Sun and L. R. Dalton [eds.], *Introduction to Organic Electronic and Optoelectronic Materials and Devices*, Boca Raton, FL: CRC Press, 2008.)

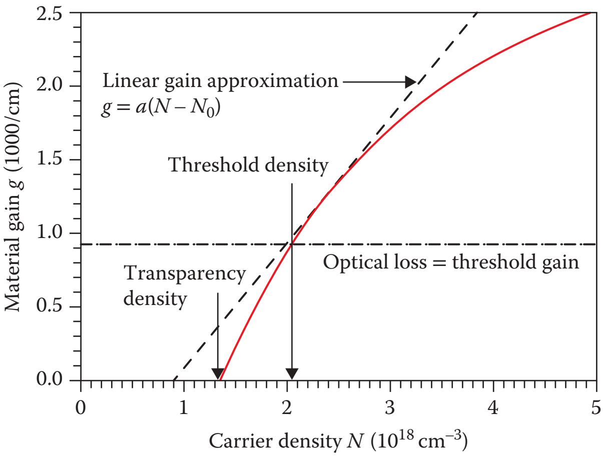


FIGURE 26.4 Optical gain provided by the active layer material as function of the carrier density.

experimental validation of the gain model creates a major uncertainty in any laser simulation. The analysis of laser measurements typically only delivers the relationship $g_m(j) = G_o \ln(j/j_{tr})$ between the modal gain $g_m = \Gamma_a g$ and the current density j, with G_o and the transparency current density j_{tr} as fit parameters (see Chapter 27). The optical confinement factor Γ_a gives the overlap of the active layer and the lasing mode, which is usually extracted from waveguide simulations (see Chapter 4).

Figure 26.4 indicates optical losses that are caused by photon emission from the laser as well as by the internal absorption and photon scattering. For the simple Fabry–Pérot (FP) laser structure shown in Figure 26.3, the optical loss at the two cavity mirrors is given by

$$\alpha_m = \frac{1}{2L} \ln\left(\frac{1}{R_f R_b}\right) \tag{26.2}$$

with the cavity length L and the reflectances R_f and R_b of front and back mirrors, respectively. The internal optical loss α_i can be extracted from laser measurements (see Chapter 27) but the microscopic mechanisms causing this loss are often hard to identify. They are typically approximated as

$$\alpha_i = \alpha_b + \sum_i \Gamma_i \left(k_{n,i} n + k_{p,i} p\right) \tag{26.3}$$

with the background loss α_b (e.g., due to photon scattering) and the sum over all free-carrier-related loss in any individual layer i of the laser structure (where k is a free-carrier absorption parameter). The confinement factor Γ_i is the ratio of the layer volume to the volume of the optical lasing mode. Photon absorption by free carriers depends on the energy band structure of the conduction and valence bands. It is typically stronger for holes due to intervalence band absorption (Piprek et al. 2000).

To reach the lasing threshold, the optical gain must compensate for the internal optical loss (α_i) and for photon emission from the device (α_m). Both loss parameters apply to the whole lasing mode so that the threshold gain g_{th} is defined by

$$\Gamma_a g_{th}\left(N_{th}\right) = \alpha_m + \alpha_i \tag{26.4}$$

with the active layer confinement factor Γ_a and the threshold carrier density N_{th} (see Figure 26.4).

26.3 Threshold Current and Slope Efficiency

The threshold current I_{th} provides the threshold carrier density N_{th} and compensates for various carrier loss mechanisms, some of which are illustrated in Figure 26.1. A spontaneous electron–hole recombination is needed to provide initial photons for stimulated recombination (lasing), but most spontaneously emitted photons are lost. In a common analytical approach, the spontaneous emission rate $R_{spon} = BN^2$ is proportional to the square of the carrier density. Nonradiative recombination mechanisms are either defect-related Shockley–Read–Hall (SRH) recombinations ($R_{SRH} = AN$) or Auger recombinations ($R_{Aug} = CN^3$). The former transfers the recombination energy to lattice vibrations (phonons) and the latter to other free carriers. This simple but very popular ABC recombination model leads to the threshold current

$$I_{th} = \frac{eV_a}{\eta_a} R_{ABC}(N_{th}) = \frac{eV_a}{\eta_a}\left(AN_{th} + BN_{th}^2 + CN_{th}^3\right) = I_{SRH} + I_{spon} + I_{Aug} + I_{leak} \tag{26.5}$$

with the injection efficiency η_a giving the fraction of electrons that recombines within the active layer of volume V_a, thereby accounting for carriers that recombine outside the active layer, e.g., due to electron leakage (see Figure 26.2). Leakage can occur by various mechanisms (Piprek et al. 2000). It is hard to asses

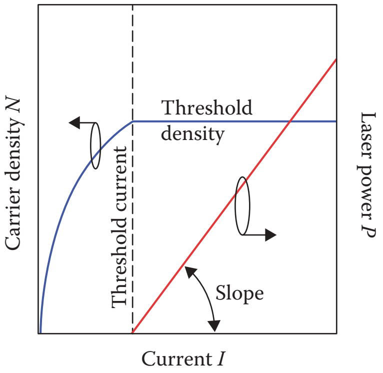


FIGURE 26.5 Illustration of carrier density and lasing power as function of the injected current without self-heating.

with analytical models and typically requires full solutions of the semiconductor transport equations (see Chapter 2). Simpler models often neglect leakage or use a fictitious parameter η_a.

With a stronger current injection $I > I_{th}$, the carrier density remains constant at N_{th} as additional electron–hole pairs are consumed by stimulated recombination (Figure 26.5). The stimulated recombination rate is $R_{stim} = v_g g S$ (where v_g is photon group velocity). Under ideal conditions without self-heating, the laser power rises proportional to $I_{stim} = I - I_{th}$ as

$$P = \eta_d \frac{h\nu}{e}\left(I - I_{th}\right) = \eta_i \frac{\alpha_m}{\alpha_m + \alpha_i} \frac{h\nu}{e}\left(I - I_{th}\right) \tag{26.6}$$

with the photon energy $h\nu$. Note that P gives the total emission from both facets. The differential quantum efficiency η_d is the fraction of carriers injected above threshold that contributes photons to the laser beams. It can be separated into internal differential efficiency η_i and optical efficiency η_{opt}. The latter is equal to $\alpha_m(\alpha_m + \alpha_i)^{-1}$ and gives the fraction of stimulated photons that leaves the laser. The internal differential efficiency η_i is often close to unity above threshold as there are no further recombination losses with constant carrier density N_{th} in the active layer. However, the leakage current may rise above the threshold, especially at higher temperatures (Piprek et al. 2000).

26.4 Temperature Effects

The current flow through the laser diode as well as nonradiative recombination processes generate heat inside the laser and elevate the internal temperature distribution $T(x, y, z)$. This temperature rise more or less affects all material parameters. One of the most fundamental changes occurs with the semiconductor bandgap $E_g(T)$, which is commonly modeled by the Varshni formula

$$E_g(T) = E_g(0) - \frac{AT^2}{B + T} \tag{26.7}$$

using the phenomenological parameters A and B (Piprek 2003). As the band gap shrinks, the lasing wavelength increases (redshifts). In addition, the Fermi distribution of carriers inside the energy bands broadens with higher temperature. Both these effects change the gain spectrum. Figure 26.6 illustrates this change for a fixed carrier density. The gain peak determines the emission wavelength of FP lasers, and it redshifts with higher temperature. But the gain peak also declines, so that more carriers are needed to maintain the lasing threshold (cf. Figure 26.4). Consequently, the threshold current rises with higher temperature, which is often described by the characteristic temperature $T_0 = (T_2 - T_1)/\ln(I_2/I_1)$ that can be extracted

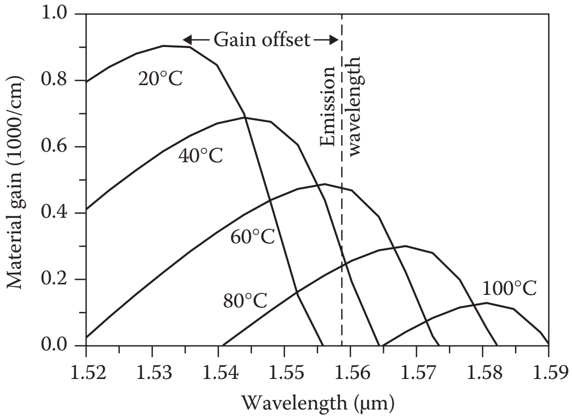


FIGURE 26.6 Optical gain spectrum shift with increasing temperature and constant carrier density.

from measurements of $I_{\mathrm{th}}(T)$. The decline of the slope efficiency is calculated in a similar way. These phenomenological parameters are useful in characterizing the temperature sensitivity of a given laser diode (see Chapter 27). However, they don't reveal the physical mechanism behind temperature effects and are not suitable for predictive simulations. For instance, if the lasing wavelength is fixed by the optical cavity design, the impact of the gain shift very much depends on the initial difference between emission wavelength and gain peak wavelength (gain offset). An example is illustrated in Figure 26.6 (Piprek et al. 1998). Here, the rising temperature first increases the gain available at the emission wavelength. The threshold current initially declines and reaches a minimum near 60°C when the gain offset is zero. With further heating, the threshold current rises rapidly. Thus, T_0 depends on the temperature in this case and fails even as a descriptive parameter.

More advanced laser models don't employ fit parameters such as T_0 and describe the underlying physical mechanisms instead, starting with the shift of the gain spectrum (Piprek et al. 2000). An increasing carrier density in the active layer goes hand in hand with increasing carrier losses. For instance, the Auger recombination rate is not affected only by the increasing carrier density, but also by the temperature sensitivity of the Auger process, which depends on its activation energy E_a: $C(T) \propto \exp(-E_a/kT)$. Carrier leakage is also sensitive to temperature changes, and it may raise the threshold current and reduce the slope efficiency. The slope efficiency also depends on the free-carrier absorption, which increases with the carrier density. All these interdependencies require advanced numerical laser models for a more reliable analysis of the temperature sensitivity (see Chapter 27).

26.5 Efficiency Analysis

As an illustrative example for this section, Figure 26.7 shows the simulated power–current and bias–current characteristics of a GaN-based laser diode in continuous-wave (CW) operation (Piprek 2016). Self-heating apparently causes a decline of the slope efficiency, which limits the maximum power achievable. Another key performance parameter is the power conversion efficiency η_{PCE}. It is defined as the ratio of light output power P to electrical input power IV (where V is bias). It can be separated into electrical efficiency $\eta_{\mathrm{ele}} = h\nu/eV$ and external quantum efficiency η_{EQE}. The latter is the ratio of emitted photon number to injected number of electron–hole pairs. Equation 26.6 leads to the following power conversion efficiency formula:

$$\eta_{\mathrm{PCE}}(I) = \frac{h\nu}{eV}\eta_{\mathrm{i}}\frac{\alpha_{\mathrm{m}}}{\alpha_{\mathrm{m}}+\alpha_{\mathrm{i}}}\frac{I-I_{\mathrm{th}}}{I} = \eta_{\mathrm{ele}}\eta_{\mathrm{EQE}} = \eta_{\mathrm{ele}}\eta_{\mathrm{s}}\eta_{\mathrm{th}} \tag{26.8}$$

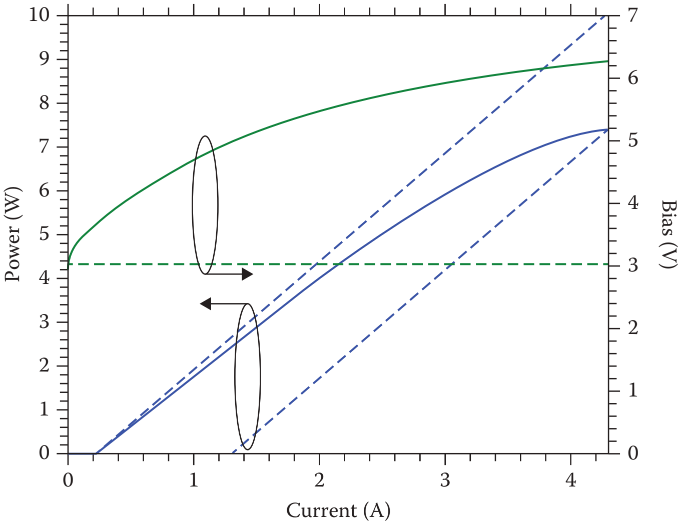


FIGURE 26.7 Lasing power and bias versus current. The dashed bias line indicates the photon energy. The dashed power lines indicate the change in threshold current for constant slope efficiency.

including the unitless slope efficiency η_s and the threshold current efficiency $\eta_{th} = (I - I_{th})/I$. Note that η_s is different from the differential slope efficiency dP/dI and from the averaged slope efficiency $P/(I - I_{th})$, which are both given in W/A and not used here. However, this popular analytical model is somewhat ambiguous when the laser experiences relevant self-heating, which causes a sublinear $P(I)$ characteristic as shown in Figure 26.7. Most parameters in Equation 26.8 change as the internal laser temperature rises with increasing current. The threshold current rises together with the threshold carrier density due to declining material gain. The slope efficiency declines due to increasing carrier leakage and/or rising internal absorption.

Figure 26.8 illustrates these efficiencies as simulated for an InGaN/GaN laser diode (Piprek 2016). At low current, η_{PCE} is mainly limited by the threshold current efficiency (which is zero for $I < I_{th}$). At high current, the strongest efficiency limitation is caused by the decline of the electrical efficiency due to the increasing excess bias above the minimum required bias $h\nu/e$ (dashed bias line shown in Figure 26.7). This excess bias is relatively high in GaN-based lasers, partially due to the low hole conductivity. However, the influence of slope efficiency and threshold efficiency depends on the assumption made in the analysis. The assumption of a constant, temperature-insensitive threshold current leads to the solid curves shown in Figure 26.8, so that temperature effects are mainly reflected by the slope efficiency. This is a convenient and common approach to extract efficiency plots directly from the measured *PI* and *VI* characteristics, without any simulation (Crump et al. 2013). However, with strong self-heating, the dependence on $I_{th}(T)$ needs to be considered, which is hard to extract directly from experimental results. Numerical simulations are indicated by the dashed lines in Figure 26.8, which reveal that the threshold efficiency is more temperature sensitive than the slope efficiency (Piprek 2016). For constant, temperature-insensitive slope efficiency, an approximate method of extracting $I_{th}(I)$ directly from measurements is illustrated by the dashed power lines shown in Figure 26.7. Starting with the measured slope dP/dI at threshold, the dashed line is shifted parallel to the current axis so that each power $P(I)$ is connected to a threshold current $I_{th}(I)$.

26.6 Rate Equation Analysis

Time-dependent effects are often analyzed in terms of rate equations considering all physical processes that change the densities of photons and carriers. We here discuss a set of two rate equations for the

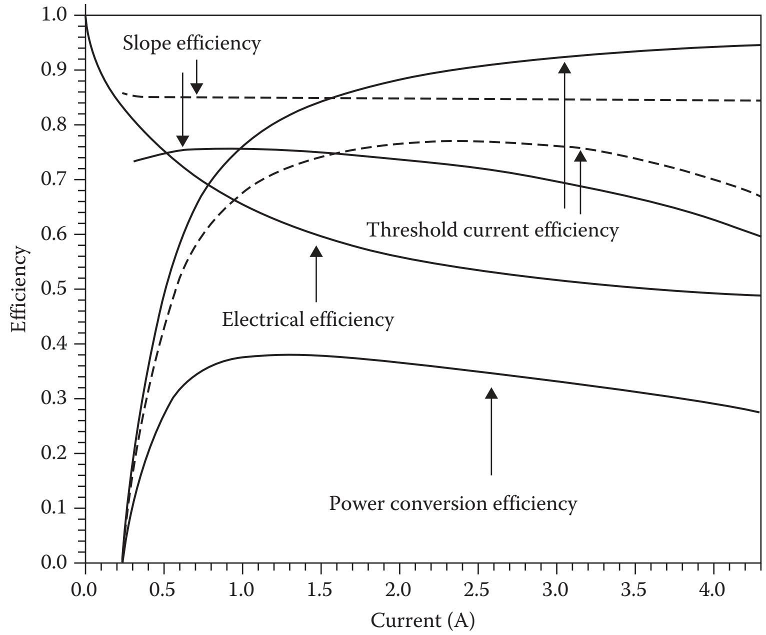


FIGURE 26.8 Changes in efficiency with increasing injection current. The solid lines indicate constant threshold current and the dashed lines indicate constant slope efficiency.

single-mode photon density S averaged over the modal volume V_m and the carrier density N averaged over the active volume V_a:

$$\frac{dN}{dt} = \frac{\eta_a I}{eV_a} - \left[AN + BN^2 + CN^3\right] - v_g g(N, S)\,S \tag{26.9}$$

$$\frac{dS}{dt} = \Gamma_a v_g g(N, S)\,S + \beta\Gamma_a BN^2 - v_g\left[\alpha_i + \alpha_m\right] S \tag{26.10}$$

The active layer carrier density $N(t)$ in Equation 26.9 is increased by current injection (first term, including the injection efficiency η_a) and is reduced by all four recombination processes (see Figure 26.1) that limit the carrier lifetime. The photon density $S(t)$ in Equation 26.10 is increased by stimulated emission (first term) and by the small fraction β of spontaneously emitted photons that enters the lasing mode (second term). $S(t)$ is reduced by photon emission and internal photon losses (third term in Equation 26.10), which limit the photon lifetime τ_p defined by $\tau_p^{-1} = v_g(\alpha_i + \alpha_m)$. The dynamic response of both densities can be understood from these rate equations. For instance, when $N(t)$ increases, $S(t)$ increases due to the rising gain $g(N)$. But that decreases the carrier density according to the last term in Equation 26.9. $N(t)$ is also reduced by ABC recombination. Consequently, the photon density $S(t)$ drops again, also due to photon losses. Thus, the dynamic behavior of both densities is strongly influenced by loss mechanisms.

However, this approach neglects the nonuniform distribution of carriers and photons (Carroll et al. 1998). A nonuniform distribution of photons can result in spatial hole burning into the carrier distribution in regions with high photon density, reduce the gain, and increase the refractive index (see Chapter 27). Multiple optical modes would require multiple rate equations for calculating each photon density (Petermann 1988). Lateral diffusion of carriers out of the active layer is partially considered by the injection efficiency η_a and is often minimized by lateral carrier confinement.

Under steady-state conditions ($dS/dt = 0$) with vanishing β, Equation 26.10 gives the relation $\Gamma_a g(N, S) = \alpha_i(N) + \alpha_m$ describing the balance of gain and losses required for lasing (see Figure 26.4).

The first equation $\mathrm{d}N/\mathrm{d}t = 0$ then yields the steady-state photon density $S_o = \eta_i \tau_p (I - I_{th})(eV_m)^{-1}$. The steady-state optical power emitted through both mirrors is given by $P_o = v_g \alpha_m h\nu V_m S_o$ in agreement with Equation 26.6.

With analog modulation, sinusoidal variations are added to the steady-state input current I_o. In the simple case of just one angular frequency $\omega = 2\pi f$ and constant amplitude ΔI, the injection current in Equation 26.9 becomes $I(t) = I_o + \Delta I \times \sin(\omega t)$. If the period of the modulation is much larger than any time constant, the output power still follows the steady-state solution (Equation 26.6). But in the general dynamic case, analytical solutions of the rate equations cannot be found and numerical methods need to be applied.

Current modulations ΔI well below $(I_o - I_{th})$ lead to variations ΔN, ΔS, and ΔP, which are much smaller than the steady-state values N_{th}, S_o, and P_o, respectively. This small signal case allows the rate equations to be solved analytically (Coldren and Corzine 1995) using the linear gain approximation illustrated in Figure 26.4. Assuming $\beta = 0$, the small signal solution to the rate equations is

$$\Delta P(\omega) = M(\omega) \times \Delta P = \frac{\omega_r^2}{\omega_r^2 - \omega^2 + i\omega\gamma} \times \eta_i \frac{\alpha_m}{\alpha_i + \alpha_m} \frac{h\nu}{e} \Delta I \tag{26.11}$$

with the angular electron–photon resonance frequency $\omega_r = 2\pi f_r$ given by

$$\omega_r^2 = \frac{a v_g S_o}{\tau_p \left(1 + \varepsilon S_o\right)} \left(1 + \frac{\varepsilon}{v_g a \tau_c}\right) \tag{26.12}$$

including the photon lifetime τ_p and the gain compression factor ε from Equation 26.1. At low photon densities ($\varepsilon = 0$), Equation 26.12 is reduced to $\omega_r^2 = (\tau_{st}\tau_p)^{-1}$ with the differential stimulated emission time $\tau_{st} = \mathrm{d}R_{stim}/\mathrm{d}N = (a v_g S_o)^{-1}$. The damping constant is given by $\gamma = \tau_{st}^{-1} + \tau_c^{-1}$ including the differential carrier lifetime τ_c, with $\tau_c^{-1} = \mathrm{d}R_{ABC}/\mathrm{d}N = A + 2BN_{th} + 3CN_{th}^2$.

Figure 26.9 illustrates the normalized modulation response $|M(\omega)|$ as a function of modulation frequency for different gain compression factors ε (Piprek and Bowers 2002). At low frequencies, the photon

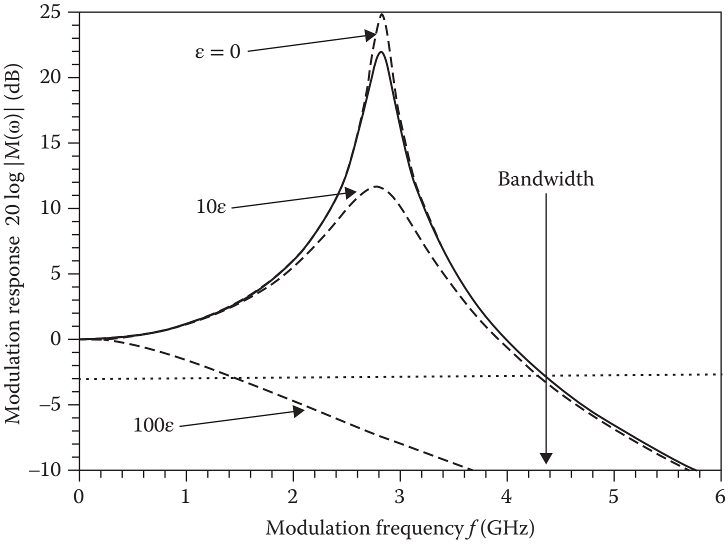


FIGURE 26.9 Modulation response versus modulation frequency. The dashed lines show the effect of changing the gain suppression factor ε.

density can easily follow the current modulation and the response function is quite flat. The response is most intense near the resonance frequency ω_r. The peak frequency is given by $\omega_p^2 = \omega_r^2 - \gamma^2/2$. At even faster current modulation, the photons cannot follow any more and the response function declines. The modulation bandwidth f_b is the frequency at which the response drops to $|M(2\pi f_b)| = 2^{-1/2} = -3dB$ (see Figure 26.9). The bandwidth rises with the steady-state current but saturates at high currents due to increased damping, device heating, gain compression, transport effects, or parasitics (Kjebon et al. 1996).

26.7 Basic Laser Cavity Designs

Chapters 27 through 34 feature specific laser designs and applications in much more detail, we here only give a brief overview of the main types of optical resonators. The most simple cavity design uses the reflection at the two laser facets for optical feedback (FP laser, see Figures 26.3 and 26.10, and Chapter 27). Constructive interference of forward and backward traveling optical waves is restricted to specific wavelengths. Those wavelengths constitute the longitudinal mode spectrum of the laser. The cavity length is typically on the order of several hundred microns, much larger than the lasing wavelength, so that many longitudinal modes may exist. The actual lasing modes are those receiving strong optical gain. Single-mode lasing is hard to achieve in simple FP structures, especially under modulation. Dynamic single-mode operation is required in many applications and is achieved using optical cavities with selective reflection. The distributed feedback (DFB) laser is widely used in single-mode fiber optic applications (Figure 26.10). Typical DFB lasers exhibit a periodic longitudinal variation of the refractive index within one layer of the edge-emitting waveguide structure. This index variation provides continuous (distributed) reflection at a wavelength given by the variation period. Facet reflection is not needed in DFB lasers; however, facet coating may be used to increase the light emission from one end of the cavity. Other laser resonators terminate the optical cavity by two distributed Bragg reflectors (DBRs) with a stepwise alternating index. More details on DFB and DBR lasers are given in Chapters 30 and 31.

A special type of DBR laser is the vertical-cavity surface-emitting laser (VCSEL) which emits light through the bottom and/or top surface of the layered structure (see Figure 26.10 and Chapters 34 and 45). The light travels perpendicular to the active layer and receives optical gain only over a very short travel distance. Thus, many more photon roundtrips and highly reflective VCSEL mirrors with more than 99% reflectivity are needed for lasing.

In transversal directions, the optical wave is typically confined by the refractive index profile, using a ridge (DFB laser) or a pillar (VCSEL) to form the waveguide. Even with restriction to one longitudinal mode, multiple transversal optical modes may occur in all three types of lasers.

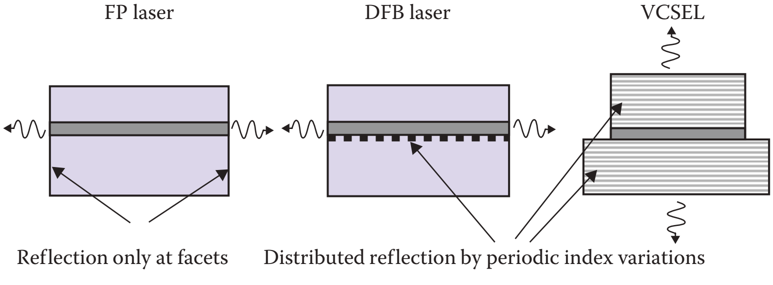

FP laser
DFB laser
VCSEL
Reflection only at facets
Distributed reflection by periodic index variations

FIGURE 26.10 Illustration of basic laser cavity designs. [DFB, distributed feedback; FP, Fabry–Pérot; VCSEL, vertical-cavity surface-emitting laser. (Adapted from Piprek, J., In S. S. Sun and L. R. Dalton [eds.], *Introduction to Organic Electronic and Optoelectronic Materials and Devices*, Boca Raton, FL: CRC Press, 2008.)]

References

Carroll, J., Whiteaway, J., and Plumb, D. *Distributed Feedback Semiconductor Lasers*. London/Washington: IEE/SPIE Press, 1998.

Coldren, L. A., and Corzine, S. W. *Diode Lasers and Photonic Integrated Circuits*. New York, NY: Wiley, 1995.

Crump, P., Erbert, G., Wenzel, H. et al. Efficient high-power laser diodes. *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 19, p. 1501211, 2013.

Kjebon, O., Schatz, R., Lourdudoss, S., Nilsson, S., and Stalnacke, B. Modulation response measurement and evaluation of MQW InGaAsP lasers of various design. *SPIE Proceedings*, vol. 2687, pp. 138–152, 1996.

Petermann, K. *Laser Diode Modulation and Noise*. Dordrecht, the Netherlands: Kluwer Academic Publishers, 1988.

Piprek, J., Akulova, Y. A., Babic, D. I., Coldren, L. A., and Bowers, J. E. Minimum temperature sensitivity of 1.55-micron vertical-cavity lasers at −30 nm gain offset. *Applied Physics Letters*, vol. 72, no. 15, pp. 1814–1816, 1998.

Piprek, J., Abraham, P., and Bowers, J. E. Self-consistent analysis of high-temperature effects on strained-layer multi-quantum well InGaAsP/InP lasers. *IEEE Journal of Quantum Electronics*, vol. 36, no. 3, pp. 366–374, 2000.

Piprek, J. and Bowers, J. E. Analog modulation of semiconductor lasers, Chapter 3. In *RF Photonic Technology in Optical Fiber Links*. Chang, W., ed. Cambridge, UK: Cambridge University Press, 2002.

Piprek, J. *Semiconductor Optoelectronic Devices: Introduction to Physics and Simulation*. San Diego, CA: Academic Press, 2003.

Piprek, J. Introduction to optoelectronic device principles, Chapter 2. In *Introduction to Organic Electronic and Optoelectronic Materials and Devices*. Sun, S. S. and Dalton, L. R, eds. Boca Raton, FL: CRC Press, 2008, pp. 25–46.

Piprek, J. Analysis of efficiency limitations in high-power InGaN/GaN laser diodes, *Optical Quantum Electron*, vol. 48, p. 471, 2016.

Shtengel, G. E., Kazarinov, R. F., Belenky, G. L., Hybertsen, M. S., and Ackerman, D. A. Advances in measurements of physical parameters of semiconductor lasers. *International Journal of High Speed Electronics and Systems*, vol. 9, pp. 901–940, 1998.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

27

High-Power Lasers

27.1 Introduction 15
27.2 Phenomenological Model 17
Summary of the Governing Equations • Determination of the Parameters Entering the Model • Comparison of Simulated and Measured CW Characteristics
27.3 Models for the Optical Power 24
Basic Relations • Rigrod Model • Treat-Power-as-a-Parameter Method • LSH and the Impact of Series Resistance and Internal Loss • Discussion of Further Nonthermal Power Saturation Effects
27.4 Model for the Optical Field 35
Basic Three-Dimensional Equations • Ansatz for Dispersion, Nonlinear Susceptibility, and Spontaneous Emission • Final Field Equation and Balance of Radiative Energy • Cavity Modes, Beam Propagation Method, and Roundtrip Operator
27.5 Models for Nonlinear Modes and Filamentation 42
Longitudinal–Lateral Approximate Projected Equations • Nonlinearities due to Virtual Transitions • Nonlinearity Induced by Lateral Spatial Hole Burning
27.6 Thermodynamic-Based Energy-Transport Model 47
Basic Equations • Spatial Distributions of the Heat Sources • Energy Conservation

Hans Wenzel
and
Anissa Zeghuzi

27.1 Introduction

High-power diode lasers deliver the energy to all high-performance laser systems, either as a pump source or as a source for direct material processing. Comprehensive descriptions of their manufacturing and applications can be found in, for example, References [1–4].

The lasers are constructed like all edge-emitting diode lasers as shown in Figure 27.1. The layer structure grown by metal-organic vapor phase epitaxy (MOVPE) or molecular beam epitaxy (MBE) on a crystalline substrate (e.g., GaAs, InP, GaN) consists basically of n-doped cladding and optical confinement layers, an active region (typically a single quantum well [QW]) and p-doped optical confinement and cladding layers and is completed by a highly p-doped cap (contact) layer. However, there are three peculiarities compared to other lasers: first, the vertical waveguide is weak (large total thickness of confinement layers or small index step between confinement and cladding lasers). Second, the emitting aperture is very broad, ranging from tens to hundreds of micrometers, which results in a nonstationary behavior [5,6]. Third, the cavity between the cleaved facets is very long, reaching values of several millimeters and the two facets are extremely differently coated.

The lateral optical and current confinement can be achieved by several means. In the most simple case, there is no built-in waveguide at all so that the optical field is confined to the region below the contact

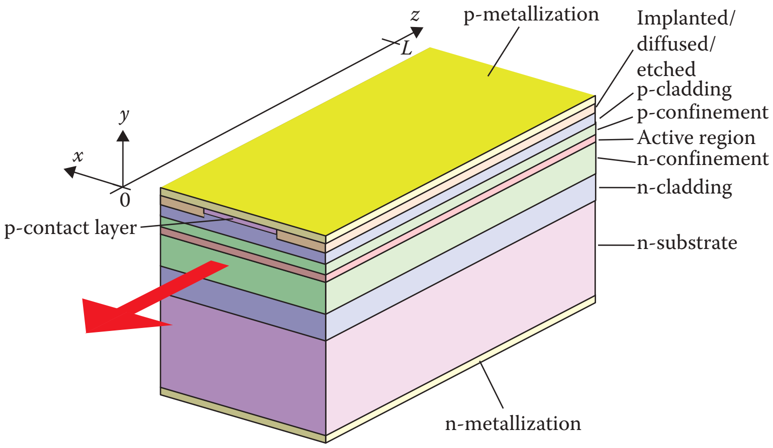


FIGURE 27.1 Schematic view of a high-power broad-area semiconductor laser.

stripe by gain guiding at low power and under pulsed operation, but is strongly influenced by the thermally induced waveguide created under continuous-wave (CW) operation at high power. In order to restrict the current spreading, the electrical conductivity of parts of the p-doped region besides the contact stripe can be reduced by ion implantation, impurity diffusion, or by implementation of a reverse-biased p-n junction. If the semiconductor besides the p-contact is etched away up to a defined depth and subsequently filled with an insulator, the index contrast between the semiconductor below the contact stripe and the insulator results in a lateral built-in index guide, which stabilizes the optical field at low power and under pulsed operation, but increases the far-field divergence.

The cleaved facets located at $z = 0$ and $z = L$ in Figure 27.1 are low reflection coated on the front side ($R_0 \propto 0.01$) and high reflection coated on the rear side ($R_L > 0.9$). If a small optical spectrum and a reduced drift of the wavelength with temperature and injection current are required, Bragg gratings could be implemented into the cavity, resulting in distributed feedback (DFB) or distributed Bragg-reflection (DBR) lasers [7]. However, here we focus on the Fabry–Pérot (FP) type of high-power lasers where the coated facets provide the feedback.

The simulation of broad-area lasers is challenging because of the different temporal and spatial scales involved. The timescales for the variations of the optical field, the carrier densities, and the temperature are ps, ns, and μs, respectively. The spatial scales range from nanometers (active QW) and micrometers (epitaxial layers and lateral waveguides) up to millimeters (cavity). Another difficulty arises from the highly nonlinear behavior because of the coupling of the optical, electronic, and thermal phenomena. Therefore, until now, no simulation tool covering all spatiotemporal scales and physical phenomena has been available. In this chapter, we survey models with different complexity for the simulation of high-power lasers [6].

This chapter is organized as follows. In Section 27.2, we present a model for high-power lasers based on measurable parameters, which can be used to compare different laser structures and to predict the electro-optical characteristics as a function of cavity length and facet reflectivities, for example. In Section 27.3, we investigate the profile of the optical power in the cavity in more detail and derive some of the equations used in Section 27.2. Furthermore, we discuss several effects responsible for the saturation of the output power with increasing injection current. The parabolic paraxial wave equation based on the slowly varying amplitude and rotating wave approximations, taking into account gain dispersion, spontaneous emission, and a third-order nonlinear susceptibility, is derived in Section 27.4. A balance equation for the electromagnetic energy density will be given. In Section 27.5, we present the equations that can be used to calculate the nonlinear lateral optical modes of broad-area lasers and discuss several root causes for the multipeaked and not diffraction-limited lateral field profile of broad-area lasers. Finally, in Section 27.6, we summarize a thermodynamic-based model for the transport of the charged carriers and the temperature flow. Particular attention is paid to a consistent formulation with the model for the optical field presented in Section 27.4.

27.2 Phenomenological Model

In this section, a model for high-power lasers based on measurable parameters is presented. The model can be used to compare different laser structures and to predict the electro-optical characteristics in dependence on cavity length and facet reflectivities, for example.

27.2.1 Summary of the Governing Equations

The total output power P_{out} of a semiconductor laser in dependence on the injection current I is given by

$$P_{\text{out}} = \frac{\hbar\omega}{e}\eta_{\text{ext}}\left(I - I_{\text{thr}}\right) \tag{27.1}$$

where η_{ext} is the external differential efficiency, I_{thr} the threshold current, ω the angular lasing frequency, $\hbar$ the reduced Planck constant, and e the elementary charge. Here and in what follows $I \geq I_{\text{thr}}$ is assumed. The photon energy $\hbar\omega$ can be written as (c vacuum speed of light, h Planck constant)

$$\hbar\omega = \frac{hc}{\lambda} \tag{27.2}$$

where the vacuum lasing wavelength λ is assumed to vary linearly with the temperature T in the cavity as

$$\lambda = \lambda_{\text{ref}} + \frac{\mathrm{d}\lambda}{\mathrm{d}T}\Delta T \qquad \text{with} \qquad \Delta T = T - T_{\text{ref}} \tag{27.3}$$

with T_{ref} being the reference (heat sink or ambient) temperature. The output powers P_0 at the left ($z = 0$) facet and P_L at the right ($z = L$) facet with intensity reflection coefficients R_0 and R_L, respectively, are obtained from the total power as

$$P_0 = \frac{P_{\text{out}}}{1 + (1 - R_L)(1 - R_0)^{-1}\sqrt{R_0 R_L^{-1}}} \quad \text{and} \quad P_L = \frac{P_{\text{out}}}{1 + (1 - R_0)(1 - R_L)^{-1}\sqrt{R_L R_0^{-1}}} \tag{27.4}$$

(cf. Equation 27.40).

To calculate the threshold current, a model for the modal gain is required. First one should be aware that the gain depends on both the wavelength and the current density. For a fixed wavelength λ, the dependence of the modal gain of QW lasers on the current density can be well approximated by a logarithmic relation,

$$g_{\text{m}} = G_0 \ln \frac{j}{j_{\text{tr}}} \tag{27.5}$$

where G_0 is the gain prefactor (proportional to the differential gain) and j_{tr} the transparency current density. At $j = j_{\text{tr}}$ the gain vanishes, i.e., a wave with the wavelength λ propagating along the QW embedded in a nonabsorbing medium is neither absorbed nor amplified (transparency). Due to the fact that FP lasers lase at the maximum of the gain spectrum, we need the dependence of the peak gain on the current density for which the relation Equation 27.5 can be used, too. However, one must keep in mind that the peak gain never becomes negative, but approaches zero for vanishing current. Therefore, $j \gg j_{\text{tr}}$ must be observed and the transparency current density obtained by the analysis of the length dependencies of external differential efficiency and threshold current as described in Section 27.2.2 has no direct physical interpretation. One should mention that of course any other functional dependence $g_{\text{m}}(j)$ that can be solved analytically for j, such as a linear one, could be used. However, then the procedure presented in Section 27.2.2 would yield differing parameters G_0 and j_{tr}.

Assuming the logarithmic dependence of the modal gain on the current density and an exponential dependence of the threshold current on the temperature, the threshold current can thus be written as

$$I_{\mathrm{thr}} = WLj_{\mathrm{tr}}e^{\frac{g_{\mathrm{thr}}}{G_0}}e^{\frac{\Delta T}{T_0}} \tag{27.6}$$

where W is the contact width, L the contact length (assumed equal to the cavity length), and T_0 the characteristic temperature of the threshold current. Similarly, the external differential efficiency can be written as (cf. Equation 27.44)

$$\eta_{\mathrm{ext}} = \frac{\alpha_{\mathrm{out}}}{g_{\mathrm{thr}}}\eta_{\mathrm{i}}e^{-\frac{\Delta T}{T_1}} \tag{27.7}$$

where T_1 is the characteristic temperature of the external differential efficiency. The internal efficiency η_{i} gives the fraction of the total current increment that results in stimulated emission of photons. It describes the effect that not all electron–hole pairs *additionally* injected above threshold are converted into photons by stimulated emission. The threshold gain g_{thr} is

$$g_{\mathrm{thr}} = \alpha_{\mathrm{out}} + \alpha_{\mathrm{i}} \tag{27.8}$$

with the internal losses α_{i} and the outcoupling (mirror) losses

$$\alpha_{\mathrm{out}} = -\frac{1}{2L}\ln(R_0R_L) \tag{27.9}$$

Finally, the temperature rise can be calculated from the dissipated power Q and the thermal resistance r_{th} related to the contact area,

$$\Delta T = \frac{r_{\mathrm{th}}}{WL}Q \tag{27.10}$$

The dissipated power is given by the difference between electrical input power and optical output power,

$$Q = UI - P_{\mathrm{out}} \tag{27.11}$$

with the voltage–current characteristics

$$U = \Delta U + \frac{\hbar\omega}{e} + \frac{r_{\mathrm{s}}}{WL}I \tag{27.12}$$

where ΔU is the so-called defect voltage and r_{s} the series resistance related to the contact area. The main contribution to the defect voltage arises from the spacing of the quasi-Fermi potentials of electrons and holes in the active region (Fermi voltage), which is assumed to be clamped above threshold and is always larger than the photon energy divided by the elementary charge. The last term in Equation 27.12 is due to the Ohmic voltage drop in the bulk semiconductor layers, at the heteroboundaries, and at the semiconductor–metal junctions.

Instead of Equation 27.11 sometimes other expressions are reported. For example, by inserting Equations 27.1 and 27.12 into Equation 27.11 and rearranging the terms,

$$Q = \frac{\hbar\omega}{e}I_{\mathrm{thr}} + \frac{\hbar\omega}{e}(1-\eta_{\mathrm{i}}e^{-\frac{\Delta T}{T_1}})(I-I_{\mathrm{thr}}) + \frac{\hbar\omega}{e}(\eta_{\mathrm{i}}e^{-\frac{\Delta T}{T_1}} - \eta_{\mathrm{ext}})(I-I_{\mathrm{thr}}) + \Delta UI + \frac{r_{\mathrm{s}}}{WL}I^2 \tag{27.13}$$

is obtained. The third term can be also written as

$$\frac{\hbar\omega}{e}(\eta_i e^{-\frac{\Delta T}{T_1}} - \eta_{\text{ext}})(I - I_{\text{thr}}) = \alpha_i L\bar{P} \tag{27.14}$$

where we have used Equations 27.7, 27.8, and 27.43. The averaged internal power $\bar{P}$ is defined in Equation 27.32. The dissipated power is thus the sum of the heat generated at threshold by nonradiative and spontaneous recombination (first term in Equation 27.13), the heat generated above the threshold due to additional nonstimulated recombination and carrier leakage (second term), the heat generated by absorption (third term), the heat caused by the defect voltage (fourth term), and the Joule heat (last term). It is worth to mention that in Equations 27.11 and 27.13, it is assumed that the spontaneously emitted radiation is completely absorbed in the cavity so that P_{out} contains only the stimulated emission.

All equations have been written in such a manner that the geometrical scaling given by W and L is explicitly separated. Note, however, that the dependence of r_{th} on W cannot be neglected because the thermal resistance $R_{\text{th}} = r_{\text{th}}/(WL)$ is not inverse proportional to W. For example, for a rectangular-shaped infinitely thin heat source,

$$r_{\text{th, 2D}} \approx \frac{W}{\kappa\pi} \ln\left(\frac{4h}{W}\right) \tag{27.15}$$

holds under the condition $W \ll h < W_s$, where κ is the thermal conductivity of the material with thickness h and width $2W_s$ separating the heat source of width W from the heat sink [8]. Only for a purely one-dimensional heat flow

$$r_{\text{th, 1D}} = \frac{h}{\kappa} \tag{27.16}$$

is independent of W.

For shallow-edged samples with substantial lateral current spreading, the transparency current density obtained by the procedure presented in Section 27.2.2 depends on the p-contact width, too [9]. In order to determine the transparency current density for an infinite p-contact width $j_{\text{tr},W\to\infty}$ without current spreading, the transparency *current* has to be plotted and linearly fitted in dependence on W. From the slope $\mathrm{d}I_{\text{tr}}/\mathrm{d}W$ one obtains $j_{\text{tr},W\to\infty}$ and from the extrapolation $W \to 0$ the spreading current $I_{\text{tr, spread}}$ according to the relation [9]

$$I_{\text{tr}}(W) = j_{\text{tr},W\to\infty}LW + I_{\text{tr, spread}} \tag{27.17}$$

The basic assumption underlying Equation 27.17 is that $I_{\text{tr, spread}}$ is independent of W. If different cavity lengths are considered, Equation 27.17 has to be divided by L.

For a numerical calculation of the electro-optical characteristics, Equation 27.10 is best suited because $\Delta T(I)$ is a monotonous function (in contrast to $P(I)$). Knowing ΔT, the threshold current and external efficiency can be determined from Equations 27.6 and 27.7, respectively, and finally the output power from Equation 27.1.

27.2.2 Determination of the Parameters Entering the Model

In what follows we consider an asymmetric super-large optical-cavity structure published in Reference [10]. The parameters η_i, α_i, G_0, and j_{tr} are determined by measuring the power–current characteristics of as-cleaved lasers having different cavity lengths L, operated under pulsed conditions to avoid self-heating ($\Delta T = 0$). Thus pulse lengths below 1 μs and duty cycles below 1% are required. Note that the lasing wavelength could substantially vary with L, which must be taken into account in the calculation of the optical power from the measured photovoltage of the detector by adjusting the calibration factor. The measured *P–I* characteristics have to be linearly fitted to extract the slope efficiencies and threshold currents for the different cavity lengths.

If the inverse external differential efficiency is plotted versus cavity length (cf. Figure 27.2a),

$$\eta_{\mathrm{ext}}^{-1}(L) = \eta_{\mathrm{i}}^{-1}\left(1 - \frac{\alpha_{\mathrm{i}}}{\ln(R)}L\right) \tag{27.18}$$

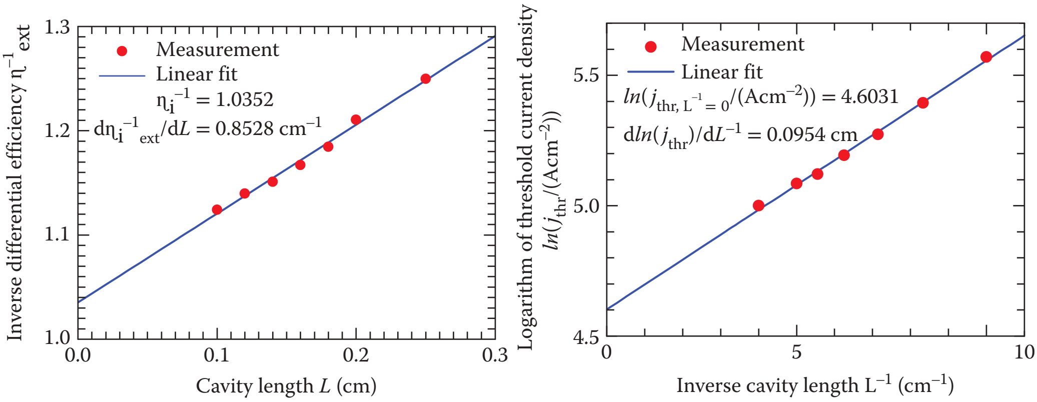
a linear fit delivers $-\eta_{\mathrm{i}}^{-1}\alpha_{\mathrm{i}}/\ln(R)$ from the slope $\eta_{\mathrm{ext}}^{-1}(L)$ and η_{i}^{-1} from the extrapolation of η_{ext}^{-1} to $L \rightarrow 0$. A linear fit of the logarithm of the threshold current density versus the inverse cavity length

$$\ln(j_{\mathrm{thr}})(L^{-1}) = \ln(j_{\mathrm{tr}}) + G_0^{-1}\left(\alpha_{\mathrm{i}} - \ln(R)L^{-1}\right) \tag{27.19}$$

as shown in Figure 27.2b yields $-G_0^{-1}\ln(R)$ from the slope $\ln(j_{\mathrm{thr}})(L^{-1})$ and the threshold current density for an infinite cavity length $\ln(j_{\mathrm{thr},\infty}) = \ln(j_{\mathrm{tr}}) + G_0^{-1}\alpha_{\mathrm{i}}$ from the extrapolation of $\ln(j_{\mathrm{thr}})$ to $L^{-1} \rightarrow 0$. The results of this procedure are contained in Table 27.1, where facet reflectivities $R_0 = R_L = 0.3$ have been assumed in the evaluation. For other cases they must be correspondingly chosen.

The cavity lengths L have to be carefully chosen to ensure linear dependencies $\eta_{\mathrm{ext}}^{-1}(L)$ and $\ln(j_{\mathrm{thr}})(L^{-1})$. If L is too small (large threshold gain), η_{ext}^{-1} could increase with decreasing L due to an increase of α_{i} (enhanced free-carrier absorption) and a decrease of η_{i} (enhanced carrier leakage). If L is too large (small threshold gain), the threshold gain does not depend logarithmically on the threshold current density because $j_{\mathrm{th}} \rightarrow j_{\mathrm{tr}}$ so that the model fails, too.

The characteristic temperatures should be determined by measuring the power–current characteristics at different chip temperatures under pulsed conditions. The outcoupling losses (determined by cavity length and facet reflectivities) of the laser under investigation should coincide, nearly, with the intended values for CW operation, because T_0 and T_1 decrease with increasing threshold gain. The temperature range must be correspondingly chosen, too, because T_0 and T_1 decrease with increasing temperature. The results of the measurements and the linear fits of the logarithms of the threshold current density and external



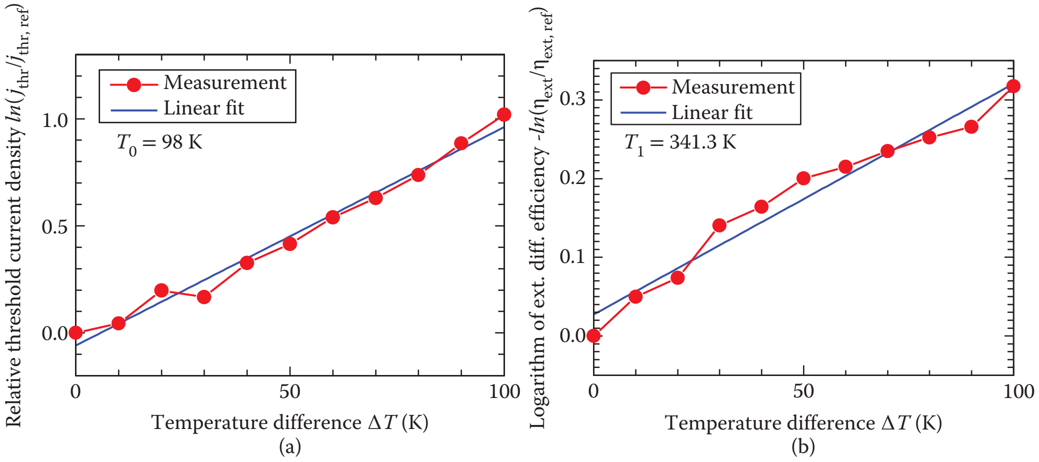
(a)

(b)

FIGURE 27.2 Determination of internal efficiency η_{i}, internal optical losses α_{i}, gain prefactor G_0, and transparency current density j_{tr}. (a) Inverse external differential efficiency versus cavity length. Bullets indicate measurements and solid line indicates linear fit. (b) Logarithm of threshold current density versus inverse cavity length. Bullets indicate measurements and solid line indicate linear fit.

TABLE 27.1 Phenomenological Laser Parameters

Parameter	Symbol	Value	Unit
Front facet reflectivity	R_0	0.012	
Rear facet reflectivity	R_L	0.95	
Width of active region	W	90	μm
Cavity length	L	4000	μm
Internal efficiency	η_i	0.97	
Internal optical losses	α_i	0.99	cm^{-1}
Gain prefactor	G_0	12.6	cm^{-1}
Transparency current density	j_{tr}	92.2	$\mathrm{A/cm}^2$
Characteristic temperature of threshold current	T_0	98	K
Characteristic temperature of external efficiency	T_1	341	K
Reference wavelength	λ_{ref}	899	nm
Change of wavelength with temperature	$\mathrm{d}\lambda/\mathrm{d}T$	0.331	nm/K
Series resistance	r_s	$0.767 \cdot 10^{-4}$	$\Omega \cdot \mathrm{cm}^2$
Thermal resistance	r_{th}	$0.936 \cdot 10^{-2}$	$\mathrm{K} \cdot \mathrm{cm}^2/\mathrm{W}$
Defect voltage	ΔU	0.033	V
Differential threshold current	j'_{thr}	5	cm^{-2}
Saturation current density	j_{sat}	$5 \cdot 10^{-6}$	cm^2/A
Differential series resistance	r'_s	$-5 \cdot 10^{-3}$	$\Omega \cdot \mathrm{cm}^4/A$



(a)

(b)

FIGURE 27.3 Determination of characteristic temperatures. Cavity length and facet reflectivities are given in Table 27.1. (a) Logarithm of threshold current density versus temperature rise. Bullets indicate measurements and solid line indicate linear fit. (b) Logarithm of external differential efficiency versus temperature rise. Bullets indicate measurements and solid line indicate linear fit.

differential efficiency versus temperature rise,

$$\ln\left(\frac{j_{thr}(T)}{j_{thr}(T_{ref})}\right) = \frac{\Delta T}{T_0} \quad \text{and} \quad -\ln\left(\frac{\eta_{ext}(T)}{\eta_{ext}(T_{ref})}\right) = \frac{\Delta T}{T_1} \tag{27.20}$$

are shown in Figure 27.3a and 27.3b.

The series resistance and defect voltage are obtained from a fit of the linear part of the voltage–current characteristics above threshold:

$$U = U_0 + \frac{r_s}{WL} I \qquad \text{and} \qquad \Delta U = U_0 - \frac{hc}{e\lambda_{\text{ref}}} \tag{27.21}$$

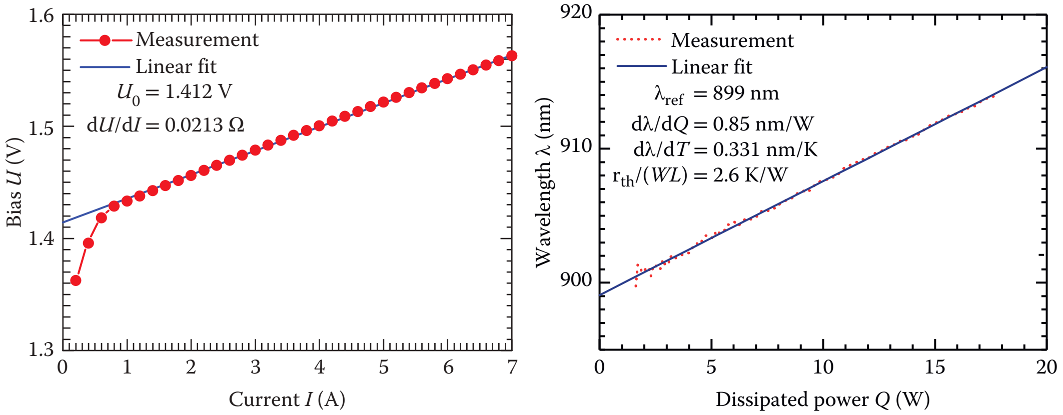
The result is shown in Figure 27.4a. The voltage must be measured with care to ensure that the true series resistance of the chip is obtained. Typically a so-called four-terminal or 4-wire sensing separating the current and voltage electrodes has to be applied. Finally, the thermal resistance can be determined from a linear fit of the lasing wavelength versus the dissipated power,

$$\lambda(Q) = \lambda_{\text{ref}} + \frac{\mathrm{d}\lambda}{\mathrm{d}T} \frac{r_{\text{th}}}{WL} Q, \tag{27.22}$$

where Q is calculated according to Equation 27.11

There are two possibilities to employ Equation 27.22. The first possibility is based on the measurement of the center wavelength λ_{center} of the emission spectrum above threshold versus injection current. This yields a correct result, if a sufficiently large number of modes is lasing and the envelope of the spectrum exhibits a Gaussian-like shape. The shift of λ_{center} is mainly determined by the temperature dependence of the energy gap of the active region, but sometimes also by the temperature dependence of the confinement factor or screening of polarization fields in GaN-based lasers, for example. The method has the advantage, that the requirement on the spectral resolution of the optical spectrometer is modest. A disadvantage is that a unique center wavelength cannot always be determined.

The second possibility is to trace the wavelength of a single longitudinal mode while increasing the current, using a high-resolution optical spectrometer. The shift of the wavelength λ_{mode} of a mode is determined by the temperature dependence of its modal phase index, which results in a much weaker coefficient $\mathrm{d}\lambda_{\text{mode}}/\mathrm{d}T$ than for the center wavelength $\mathrm{d}\lambda_{\text{center}}/\mathrm{d}T$ of the emission spectrum. This method yields more reliable results for the thermal resistance. In any case, $\mathrm{d}\lambda_{\text{center}}/\mathrm{d}T$ or $\mathrm{d}\lambda_{\text{mode}}/\mathrm{d}T$ has to be determined in advance by measuring the functions $\lambda_{\text{center}}(T)$ or $\lambda_{\text{mode}}(T)$, respectively, where T is the chip temperature.



(a) (b)

FIGURE 27.4 Determination of series and thermal resistances. (a) Voltage–current characteristic. Bullets indicate measurements. Solid line indicates linear fit. (b) Center wavelength versus dissipated power. Light grey line indicates measurements. Straight line indicates linear fit. Cavity length and facet reflectivities are given in Table 27.1.

27.2.3 Comparison of Simulated and Measured CW Characteristics

In Figure 27.5 the measured $P - I$, $U - I$ and $\eta_c - I$ characteristics (solid lines) are compared with the results obtained with the phenomenological model (dashed lines) presented above in Section 27.2.1. The conversion efficiency η_c is defined as

$$\eta_c = \frac{P_{\text{out}}}{UI} \tag{27.23}$$

The bending of the power–current characteristics at high bias is caused by the temperature-induced increase of the threshold current and decrease of the slope efficiency (determined by the parameters T_0 and T_1 here) due to the power dissipation. It leads to what is commonly referred to as “thermal roll-over.” However, the measured bending is stronger than the simulated one.

This discrepancy is caused by the fact that there are nonthermal reasons for the bending of the P–I characteristic at high bias (also referred as “power saturation”) which cannot be addressed by the phenomenological laser models based on parameters measured at low bias. For example, bending of the conduction and valence bands and carrier accumulation effects occurring at large bias leading to enhanced nonstimulated recombination and internal absorption [11] result in an increased threshold current and a decreased external differential efficiency. Other effects are longitudinal spatial holeburning (LSH), nonlinear gain compression, and two-photon absorption as discussed in Sections 27.3.2 and 27.3.5, which result in a saturation of the output power.

These effects could be modeled phenomenologically by additional explicit current dependencies $I_{\text{thr}}(I)$ and $\eta_{\text{ext}}(I)$ with parameters j'_{thr} and j_{sat},

$$I_{\text{thr}} = I_{\text{thr},0} + WLj'_{\text{thr}}\left(I - I_{\text{thr},0}\right) \tag{27.24}$$

and

$$\eta_{\text{ext}} = \frac{\eta_{\text{ext},0}}{1 + \frac{I - I_{\text{thr},0}}{WLj_{\text{sat}}}}, \tag{27.25}$$

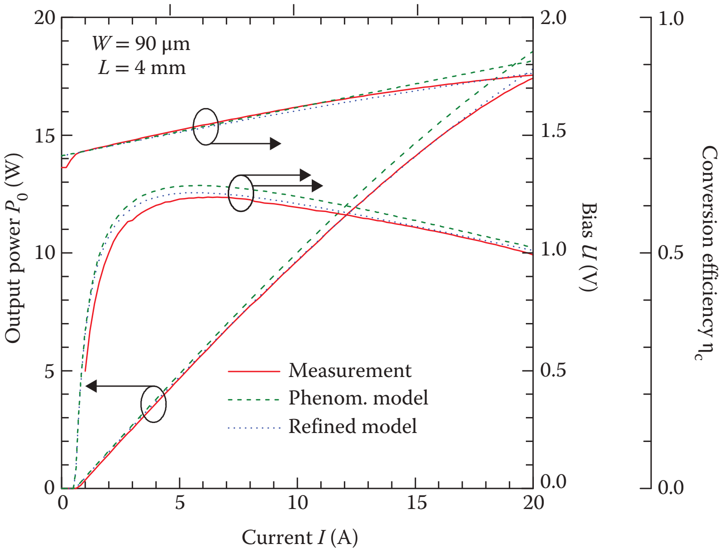


FIGURE 27.5 Output power at $z = 0$ (left axis), applied voltage (right axis), and conversion efficiency (right-most axis) versus injection current. Solid line indicates measurement and dashed line indicates phenomenological laser model. Dotted line indicates refined model.

respectively. The proportionality factor $j'_{\mathrm{thr}} > 0$ models the increase of the threshold current due to nonthermal effects with $I_{\mathrm{thr},0}$ given by Equation 27.6. Similarly, the parameter $j_{\mathrm{sat}} > 0$ causes a reduction of the external differential efficiency where $\eta_{\mathrm{ext},0}$ is given by Equation 27.7.

Furthermore, the series resistance varies also far above threshold, either caused by carrier accumulation (differential resistance $r'_{\mathrm{s}} < 0$) or temperature effects ($r'_{\mathrm{s}} > 0$),

$$r_{\mathrm{s}} = r_{\mathrm{s},0} + \frac{r'_{\mathrm{s}}}{WL}\left(I - I_{\mathrm{thr},0}\right) \tag{27.26}$$

With the refined model (dotted lines), a better match can be achieved as Figure 27.5 reveals. Note that the used parameters j'_{thr}, j_{sat}, and r'_{s} given in Table 27.1 are guess values. An even better agreement between theory and experiment could be achieved by employing a fitting procedure, probably.

27.3 Models for the Optical Power

In this section, we investigate the profile of the optical power in the cavity in more detail and derive some of the equations used in Section 27.2. Furthermore, we discuss the nonthermal effects responsible for the saturation of the output power with increasing injection current.

27.3.1 Basic Relations

The total power P can be always separated into forward and backward propagating parts,

$$P(z,t) = P^{+}(z,t) + P^{-}(z,t) \tag{27.27}$$

In an FP cavity, they fulfil the partial differential equations ($\partial_t = \partial/\partial t$, $\partial_z = \partial/\partial z$)

$$\partial_t P^{\pm} \pm v_{\mathrm{g}} \partial_z P^{\pm} = v_{\mathrm{g}}(g_{\mathrm{m}} - \alpha_{\mathrm{m}})P^{\pm} + \frac{\dot{P}_{\mathrm{sp}}}{2} \tag{27.28}$$

subject to the boundary conditions

$$\begin{aligned} P^{+}(0,t) &= R_0 P^{-}(0,t) \\ P^{-}(L,t) &= R_L P^{+}(L,t) \end{aligned} \tag{27.29}$$

Here $g_{\mathrm{m}} - \alpha_{\mathrm{m}}$ is the modal net gain in the cavity with g_{m} being the modal gain and α_{m} the modal losses, L the cavity length, $v_{\mathrm{g}} = c/n_{\mathrm{g}}$ the group velocity with n_{g} being the modal group index, and $\dot{P}_{\mathrm{sp}}$ the rate of spontaneous emission coupled into the lasing modes. The total output power at the facets is $P_{\mathrm{out}} = P_0 + P_{\mathrm{L}}$ with

$$P_0 = (1 - R_0)P^{-}(0), \quad P_{\mathrm{L}} = (1 - R_{\mathrm{L}})P^{+}(L) \tag{27.30}$$

Integrating Equation 27.28 along z, adding both equations, and taking into account Equation 27.29 we obtain the power balance

$$\frac{1}{v_{\mathrm{g}}}\frac{\mathrm{d}\bar{P}}{\mathrm{d}t} = -\frac{P_{\mathrm{out}}}{L} + \frac{1}{L}\int_0^L (g_{\mathrm{m}} - \alpha_{\mathrm{m}})P\,\mathrm{d}z + \frac{1}{v_{\mathrm{g}}L}\int_0^L \dot{P}_{\mathrm{sp}}\,\mathrm{d}z \tag{27.31}$$

with the average internal power

$$\bar{P} = \frac{1}{L}\int_0^L (P^+ + P^-)\,\mathrm{d}z. \tag{27.32}$$

In what follows, we consider the steady state ($\partial_t = 0$) and two special cases, namely the case around threshold and the case above threshold.

27.3.1.1 Steady State around Threshold

Around threshold both $g_\mathrm{m} - \alpha_\mathrm{m}$ and $\dot{P}_\mathrm{sp}$ can be assumed to be constant (z-independent). Therefore, the solution of Equation 27.28 taking into account Equation 27.29 is

$$P^\pm(z) = \frac{\dot{P}_\mathrm{sp}}{2v_\mathrm{g}} \cdot \frac{C^\pm e^{\pm(g_\mathrm{m}-\alpha_\mathrm{m})z} - 1}{g_\mathrm{m} - \alpha_\mathrm{m}} \tag{27.33}$$

with

$$C^+ = \frac{1 - R_0 + (1 - R_L)R_0 e^{(g_\mathrm{m}-\alpha_\mathrm{m})L}}{1 - R_0 R_L e^{2(g_\mathrm{m}-\alpha_\mathrm{m})L}} \tag{27.34}$$

$$C^- = \frac{1 - R_L + (1 - R_0)R_L e^{(g_\mathrm{m}-\alpha_\mathrm{m})L}}{1 - R_0 R_L e^{2(g_\mathrm{m}-\alpha_\mathrm{m})L}} e^{(g_\mathrm{m}-\alpha_\mathrm{m})L}. \tag{27.35}$$

The ratio between the outcoupled powers at the facets is given by

$$\frac{P_L}{P_0} = \frac{1 - R_L}{1 - R_0} \cdot \frac{1 + R_0 e^{(g_\mathrm{m}-\alpha_\mathrm{m})L}}{1 + R_L e^{(g_\mathrm{m}-\alpha_\mathrm{m})L}} \tag{27.36}$$

If the spontaneous emission approaches zero, $C^\pm$ must go to infinity to obtain a nonzero output power, which yields the so-called threshold condition $g_\mathrm{m} - \alpha_\mathrm{m} = \alpha_\mathrm{out}$.

27.3.1.2 Steady State above Threshold

Above threshold $g_\mathrm{m} - \alpha_\mathrm{m}$ cannot longer be assumed to be constant due to LSH, but the spontaneous emission can be neglected, $\dot{P}_\mathrm{sp} = 0$. Hence, the solution of Equation 27.33 is

$$P^\pm(z) = P^\pm(0) e^{\pm \int_0^z [g_\mathrm{m}(z') - \alpha_\mathrm{m}(z')]\,dz'}. \tag{27.37}$$

In order to obey the boundary conditions Equation 27.29, the threshold condition

$$g_\mathrm{thr} = \frac{1}{L}\int_0^L \alpha_\mathrm{m}\,\mathrm{d}z + \alpha_\mathrm{out} \tag{27.38}$$

must hold, where α_out are the outcoupling losses Equation 27.9 and

$$g_\mathrm{thr} = \frac{1}{L}\int_0^L g_\mathrm{m}\,\mathrm{d}z \tag{27.39}$$

is the threshold gain. The ratio between the outcoupled powers

$$\frac{P_\mathrm{L}}{P_0} = \frac{1-R_\mathrm{L}}{1-R_0}\sqrt{\frac{R_0}{R_\mathrm{L}}} \tag{27.40}$$

follows from the boundary conditions Equation 27.29, the general relation

$$\partial_z(P^+P^-) = 0 \tag{27.41}$$

and from the ratio between the internal and external powers given by Equation 27.30.

27.3.1.3 External Differential Efficiency

From the threshold condition Equation 27.38, the modal threshold gain g_thr can be calculated and from that, depending on the model used, the threshold carrier density N_thr, the threshold voltage U_thr, or the threshold current I_thr. Finally, the external differential efficiency is the ratio between the energy leaving the cavity and the energy generated by stimulated recombination (cf. Equation 27.47),

$$\eta_\mathrm{ext} = \frac{P_\mathrm{out}}{\hbar\omega\int R_\mathrm{st}\,\mathrm{d}V} = \frac{P_\mathrm{out}}{\int g_\mathrm{m}P\,\mathrm{d}z}. \tag{27.42}$$

Assuming z-independent $g_\mathrm{m} - \alpha_\mathrm{m}$, the relation between the total output power and the average internal power follows from Equations 27.31 and 27.38 to

$$P_\mathrm{out} = \alpha_\mathrm{out}L\bar{P} \tag{27.43}$$

and the external differential efficiency is

$$\eta_\mathrm{ext} = \frac{\alpha_\mathrm{out}}{g_\mathrm{thr}} \tag{27.44}$$

27.3.2 Rigrod Model

In order to determine the variation of the modal gain along z above threshold, a model for the carrier densities is required. The simplest model consists of a rate equation for the excess carrier density N in the active region

$$\frac{\mathrm{d}N}{\mathrm{d}t} = \frac{j}{ed} - R(N,P) \tag{27.45}$$

with a constant injection current density j. The recombination rate [12]

$$R = R_\text{non-rad} + R_\mathrm{sp} + R_\mathrm{st} \tag{27.46}$$

consists of nonradiative (Shockley–Read–Hall, Auger) recombination $R_\text{non-rad}$, radiative spontaneous recombination R_sp, and radiative stimulated recombination:

$$R_\mathrm{st} = \frac{g_\mathrm{m}P}{dW\hbar\omega} \tag{27.47}$$

where $P = P^+ + P^-$, g_m is the modal gain, $\hbar\omega$ the photon energy, d the thickness of the active region, and W the width of the active region.[†]

Even for the steady-state above threshold, there exists no analytical solution of Equations 27.28 and 27.45. However, if the nonstimulated recombination rate and the gain are linearized,

$$R_\text{non-rad} + R_\text{sp} = \frac{N}{\tau_N} \tag{27.48}$$

and

$$g_\mathrm{m} = g'_\mathrm{m}(N - N_\mathrm{tr}), \tag{27.49}$$

respectively, and inserted into Equation 27.45, the excess carrier density N can be determined.

Introducing the obtained expression for N again into Equation 27.49 yields

$$g_\mathrm{m} = \frac{g_0}{1 + \frac{P}{P_\mathrm{sat}}} \tag{27.50}$$

with the unsaturated gain

$$g_0 = \frac{g'_\mathrm{m}\tau_N}{ed}(j - j_\mathrm{tr}) \tag{27.51}$$

where g'_m is the differential gain, τ_N the effective carrier lifetime, $j_\mathrm{tr} = edN_\mathrm{tr}/\tau_N$ the transparency current density, N_tr the transparency carrier density, and

$$P_\mathrm{sat} = \frac{dW\hbar\omega}{g'_\mathrm{m}\tau_N} \tag{27.52}$$

is the saturation power. For the typical values of $d = 10$ nm, $W = 100$ µm, $g'_\mathrm{m} = 10 \cdot 10^{-18}\ \mathrm{cm}^2$, $\tau_N = 1$ ns, and $\hbar\omega = 1.24$ eV, we obtain $P_\mathrm{sat} = 0.2$ W, which is much smaller than the internal power of state-of-the-art broad-area lasers operated far above threshold.

The resulting equation for the steady-state

$$\pm\frac{\mathrm{d}P^\pm}{\mathrm{d}z} = \left(\frac{g_0}{1 + \frac{P^+ + P^-}{P_\mathrm{sat}}} - \alpha_\mathrm{m}\right)P^\pm + \frac{\dot{P}_\mathrm{sp}}{2v_\mathrm{g}} \tag{27.53}$$

has been analytically solved by Rigrod [13] neglecting spontaneous emission and modal losses ($\dot{P}_\mathrm{sp} = \alpha_\mathrm{m} = 0$). For the general case a numerical solution has to be employed.

For $P_\mathrm{sat} \ll P$, $P^+ \ll P^-$, $g_\mathrm{m} \approx g_0 P_\mathrm{sat}/P^-$ and the resulting equation

$$-\frac{\mathrm{d}P^-}{\mathrm{d}z} = g_0 P_\mathrm{sat} - \alpha_\mathrm{m}P^- + \frac{\dot{P}_\mathrm{sp}}{2v_\mathrm{g}} \tag{27.54}$$

[†] Equation 27.47 follows from Equation 27.140 by averaging over the active region $1/(\mathrm{d}W)\iint \mathrm{d}x\mathrm{d}y$, taking into account Equation 27.106 and $n_\mathrm{m} = \bar{n}$.

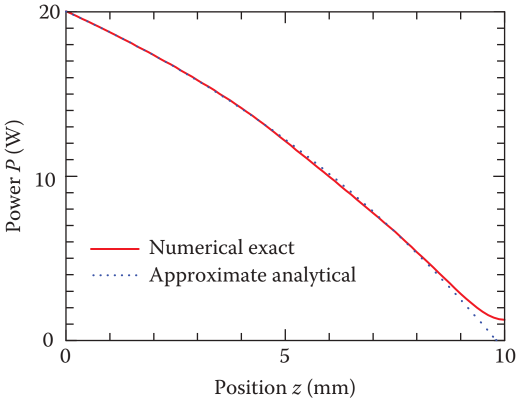


FIGURE 27.6 Longitudinal profiles of the total power for the parameters in Table 27.2 based on the numerical solution of Equation 27.53 (solid) and the analytical solution of Equation 27.55 (dashed).

has an analytical solution, too, namely

$$P(z) \approx P^{-}(z) = P^{-}(0)e^{\alpha_{\mathrm{m}} z} + \left(g_0 P_{\mathrm{sat}} + \frac{\dot{P}_{\mathrm{sp}}}{2v_{\mathrm{g}}} \right) \left(\frac{1 - e^{\alpha_{\mathrm{m}} z}}{\alpha_{\mathrm{m}}} \right) \tag{27.55}$$

In Figure 27.6, the exact and approximate solutions are compared. Except near the right facet located at $z = 10$ mm where the condition $P^{+} \ll P^{-}$ is not fulfilled, a good agreement can be noted. Note that according to Equation 27.55 $P(z)$ varies linearly for vanishing modal absorption $\alpha_{\mathrm{m}} = 0$.

27.3.3 Treat-Power-as-a-Parameter Method

The equation for the steady-state longitudinal power profile above threshold

$$\pm \frac{dP^{\pm}}{dz} = \left[g_{\mathrm{m}}(P) - \alpha_{\mathrm{m}} \right] P^{\pm} \tag{27.56}$$

subject to the boundary conditions Equation 27.29 can be conveniently solved numerically by the "treat-power-as-a-parameter" (TPP) method as introduced for the simulation of DFB lasers in Reference [14] and applied to high-power Fabry–Pérot lasers in Reference [15]. In a first step, g_{m} is calculated as a function of an external parameter (such as injection current I or bias U) and the power P, and stored in a look-up table, together with other quantities of interest (e.g., λ and I). In a second step, the boundary value problem (Equation 27.56) is solved by interpolating g_{m} in the look-up table. For given I or U, one chooses a guess value $P^{-}(0)$ and integrates Equation 27.56 from $z = 0$ to $z = L$ where typically the boundary condition is not fulfilled. Therefore, the function $P^{-}(L) - R_{\mathrm{L}} P^{+}(L)$ has to be nullified by varying $P^{-}(0)$. It is also possible to give $P^{-}(0)$ and to vary I or U to fulfil the boundary condition.

The lasing wavelength can be determined approximately by searching the maximum of integral $\int_0^L (g_{\mathrm{m}} - \alpha_{\mathrm{m}})\, dz$ with respect to λ. LSH is included automatically via the power dependence of g_{m} in Equation 27.56. If g_{m} is evaluated at the average power $\bar{P}$ in the cavity, the usual model neglecting LSH is recovered.

27.3.4 LSH and the Impact of Series Resistance and Internal Loss

The assumption of a constant injection current density in Equation 27.45 results in an overestimation of spatial hole burning. At the ohmic contacts, the quasi-Fermi potentials of electrons and holes, φ_n and

φ_p, respectively, are fixed and given by the applied bias, $\varphi_n = 0$ at the n-contact and $\varphi_p = U$ at the p-contact. The injection current density can be set equal to the hole current density and thus proportional to the gradient of the quasi-Fermi potential of the holes at the boundary between the active region and the p-doped region. Assuming isothermal conditions, one-dimensional current flow, and vanishing recombination outside the active region as well as infinitely high electron conductivity ($\sigma_n \to \infty$) in the n-doped region,

$$j(N) = \frac{U - \varphi_{\mathrm{F}}(N)}{r_{\mathrm{s}}} \tag{27.57}$$

can be derived from Equations 27.138 and 27.149 for the region beneath the p-contact, where $\varphi_{\mathrm{F}}(N) = \varphi_p(N) - \varphi_n(N)$ is the Fermi voltage and

$$r_{\mathrm{s}} = \sum_i \frac{d_i}{\sigma_{\mathrm{p},i}} \tag{27.58}$$

is the area-related total series resistance of the p-doped layers with thicknesses d_i and hole conductivities $\sigma_{\mathrm{p},i}$ between the active region and p-contact stripe.

For parabolic bands the relations between quasi-Fermi potentials φ_n and φ_p and the electron and hole densities n and p, respectively, are given by

$$e\varphi_n = -k_{\mathrm{B}}T\mathcal{F}_i^{\mathrm{inv}}\left(\frac{n}{N_{\mathrm{c}}}\right) - E_{\mathrm{c}} + e\varphi \qquad \text{and} \qquad e\varphi_p = k_{\mathrm{B}}T\mathcal{F}_i^{\mathrm{inv}}\left(\frac{p}{N_{\mathrm{v}}}\right) - E_{\mathrm{v}} + e\varphi, \tag{27.59}$$

where E_{c}, E_{v} are the (effective) conduction and valence band edges, N_c and N_v the conduction and valence band edge density of states, $\mathcal{F}_i^{\mathrm{inv}}$ the inverse Fermi integrals with $i = 1/2$ for bulk and $i = 0$ for QW active regions and φ the electrostatic potential. Hence, the Fermi voltage is given by

$$\varphi_{\mathrm{F}} = \frac{k_{\mathrm{B}}T}{e}\left[\mathcal{F}_i^{\mathrm{inv}}\left(\frac{p}{N_{\mathrm{v}}}\right) + \mathcal{F}_i^{\mathrm{inv}}\left(\frac{n}{N_{\mathrm{c}}}\right)\right] + \frac{E_{\mathrm{g}}}{e} \tag{27.60}$$

where $E_{\mathrm{g}} = E_{\mathrm{c}} - E_{\mathrm{v}}$ is the energy gap. In Equation 27.57, local charge neutrality $p - n + p_{\mathrm{D}} - n_{\mathrm{A}} = 0$ is assumed so that the electron and hole densities are given by $n = n_0 + N$ and $p = p_0 + N$, respectively, with n_0, p_0 being the equilibrium densities and N the excess carrier density.

For $r_{\mathrm{s}} \to \infty$, the usual model of a constant current density is recovered, whereas for $r_{\mathrm{s}} \to 0$, the carrier density becomes constant. This can be more readily seen by expanding $\varphi_{\mathrm{F}}(N)$ around the average carrier density $\bar{N}$ which yields

$$j(N) = \bar{j} - \frac{\varphi_{\mathrm{F}}'}{r_{\mathrm{s}}}(N - \bar{N}) \tag{27.61}$$

where the transition between both models (constant current density and constant carrier density, respectively) is governed by the parameter $\varphi_{\mathrm{F}}'/r_{\mathrm{s}}$.

Equations 27.56 and 27.45 with Equations 27.46, 27.47, and 27.57 assuming linear recombination and gain models, Equations 27.48 and 27.49, respectively, have been solved using the TPP method. The parameters used are given in Table 27.2. In Figure 27.7a, the influence of the series resistance on the power–current characteristics is clearly visible. For $r_{\mathrm{s}} = 10^{-6}\ \Omega \cdot \mathrm{cm}^2$ almost the same characteristics as without LSH is obtained.

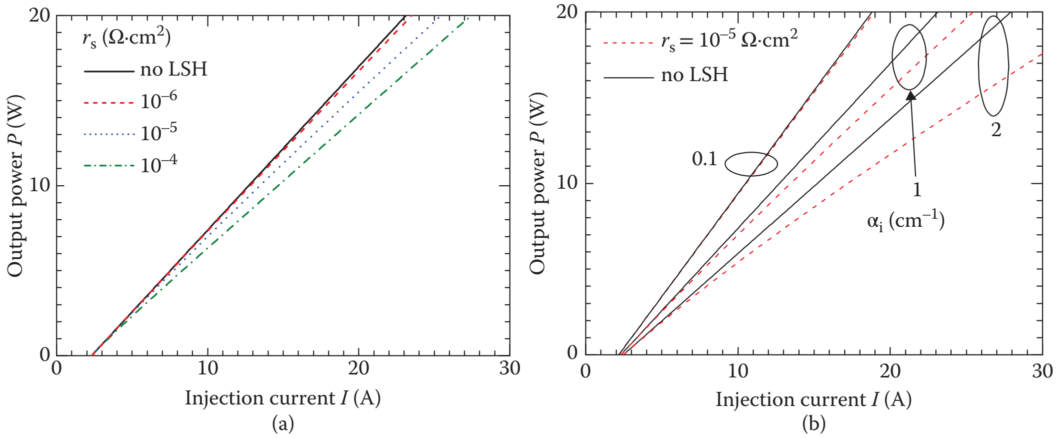


FIGURE 27.7 Power–current characteristics without (black solid) and with longitudinal spatial holeburning for different series resistances r_s and modal losses α_m. (a) Influence of the series resistance for a modal loss of $\alpha_m = 1\ \text{cm}^{-1}$ and (b) Influence of modal losses for a series resistance of $r_s = 10^{-5}\Omega \cdot \text{cm}^2$.

TABLE 27.2 Laser Parameters Used to Investigate Power Saturation Effects

Parameter	Symbol	Value	Unit
Front facet reflectivity	R_0	0.001	
Rear facet reflectivity	R_L	0.95	
Width of active region	W	100	μm
Thickness of active region	d	10	nm
Cavity length	L	1	cm
Modal losses	α_m	1	cm^{-1}
Rate of spontaneous emission	$\dot{P}_{sp}$	0	
Differential modal gain	g'_m	$10 \cdot 10^{-18}$	cm^2
Transparency carrier density	N_{tr}	10^{18}	cm^{-3}
Carrier lifetime	τ_N	1	ns
Wavelength	λ	1000	nm
Series resistance	r_s	10^{-5}	Ω · cm^2
Temperature	T	300	K
Conduction band density of states	N_c	$0.4 \cdot 10^{18}$	cm^{-3}
Valence band density of states	N_v	$13 \cdot 10^{18}$	cm^{-3}

If there are no internal losses, LSH has no impact on the slope efficiency at all, because from Equation 27.31

$$P_{\text{out}} = \int_0^L g_m P\,dz \tag{27.62}$$

follows and Equation 27.42 gives $\eta_{\text{ext}} = 1$ independent of the spatial profiles of g_m and P. As Figure 27.7b reveals, the difference between the power–current characteristics calculated with and without spatial hole burning increases with increasing modal losses.

The longitudinal profiles of forward, backward, and total power; modal gain; and injected current density are shown in Figure 27.8 for the extreme values of r_s. For a high value of r_s (Figure 27.8a), the injected current density is almost constant, so that the gain varies strongly. In contrast, for a low value of r_s (Figure 27.8b) the injected current density varies strongly so that the gain is almost constant. This results in quite different power profiles. In particular, the averaged internal power differs by almost a factor of 2 ($\bar{P} = 11$ W versus $\bar{P} = 6$ W). Besides by a reduction of the internal losses, the series resistance or the asymmetry in the facet coating ("unfolding the cavity" [16]) LSH could be mitigated by a tapered lateral waveguide design where the contact width W increases from the rear facet toward the front facet [17,18].

27.3.5 Discussion of Further Nonthermal Power Saturation Effects

27.3.5.1 Nonlinear Gain Compression

There are a couple of effects that can be described by an effective decrease of the optical gain with increasing photon density $|u|^2$ or power P, described by the functions

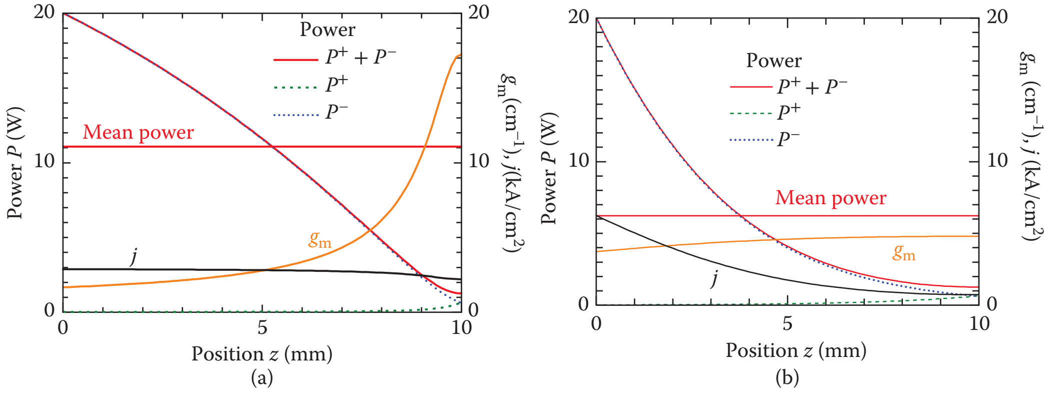
$$g = \frac{g_0}{1 + \epsilon_s |u|^2} \quad \text{or} \quad g = \frac{g_0}{1 + \frac{P}{P_s}} \tag{27.63}$$

similar to Equation 27.50. The relation between the gain compression factor ϵ_s and the saturation power P_s is given by

$$P_s = \frac{dWv_g \hbar\omega}{\Gamma \epsilon_s} \tag{27.64}$$

where Γ is the optical confinement factor of the active region. The gain compression factor is of the order $\epsilon_s \propto 10^{-17}$ cm^3, which results in a saturation power of $P_s \propto 150$ W for typical values $d = 10$ nm, $W = 100$ µm, $\Gamma = 0.01$, $n_g = 4$, $\lambda_0 = 1$ µm. Hence gain compression effects are of minor importance for CW lasers, but might result in power saturation of pulsed lasers where $P = 100$ W from a stripe width of $W = 100$ µm can be achieved [19].

The first effect contributing to nonlinear gain compression is spectral hole burning, which arises due to the interplay of the depletion of the carriers at the lasing wavelength due to stimulated emission and



(a)

(b)

FIGURE 27.8 Longitudinal profiles of forward (P^+), backward (P^-), and total ($P = P^+ + P^-$), power, modal gain g_m, and injected current density j for two series resistances r_s at an output power of $P_{out} = 20$ W. (a) Series resistance $r_s = 10^{-4}$ cm^2 and (b) series resistance $r_s = 10^{-6}$ cm^2. The horizontal line is the averaged internal (mean) power $\bar{P}$.

the relaxation of injected carriers into the depleted spectral region. At high power densities, the relaxation processes due to intraband carrier–carrier scattering with a time scale of 50–100 fs are not sufficiently fast to fill the spectral hole formed by the optical transitions. This results in a reduction of the optical gain around the lasing wavelength.

A second effect is the increase of the temperatures of the carrier distributions beyond the lattice temperature by the removal of "cold" carriers near the band-edges due to stimulated emission or by the transfer of carriers to high energies within the bands by free-carrier, inter-valence band and two-photon absorption. The temperature increase results in a reduction of the optical gain, too.

Nonlinear gain compression also can be caused by lateral spatial hole burning [20] and the Bragg grating induced by the standing waves in Fabry–Pérot lasers [21]. The impact of both the gain compression and two-photon absorption on the power–current characteristics is discussed in the next section (cf. Figure 27.9).

27.3.5.2 Two-Photon Absorption

The two-photon absorption coefficient β is described in more detail in Section 27.5.2. The resulting modal absorption for a broad-area laser with a lateral top-hat intensity profile is given by (cf. Equation 27.121)

$$\alpha_{2,\mathrm{m}} = \frac{\int n_{\mathrm{r}}(x,y)\beta(x,y)|\Phi(x,y)|^4\,\mathrm{d}x\mathrm{d}y}{n_{\mathrm{m}}(\int |\Phi(x,y)|^2\,\mathrm{d}x\mathrm{d}y)^2}P \approx \frac{\int n_{\mathrm{r}}(y)\beta(y)|\phi(y)|^4\,\mathrm{d}y}{n_{\mathrm{m}}(\int |\phi(y)|^2\,\mathrm{d}y)^2}\frac{P}{W} = \frac{\bar{\beta}_{\mathrm{m}}}{A_{\mathrm{m}}}P, \tag{27.65}$$

where ϕ is the profile of the vertical mode, W the lateral width, n_{m} the modal index, and A_{m} the mode area given by

$$A_{\mathrm{m}} = \frac{W(\int |\phi(y)|^2\,\mathrm{d}y)^2}{\int |\phi(y)|^4\,\mathrm{d}y} \tag{27.66}$$

and

$$\bar{\beta}_{\mathrm{m}} = \frac{\int n_{\mathrm{r}}(y)\beta(y)|\phi(y)|^4\,\mathrm{d}y}{n_{\mathrm{m}}\int |\phi(y)|^4\,\mathrm{d}y} \tag{27.67}$$

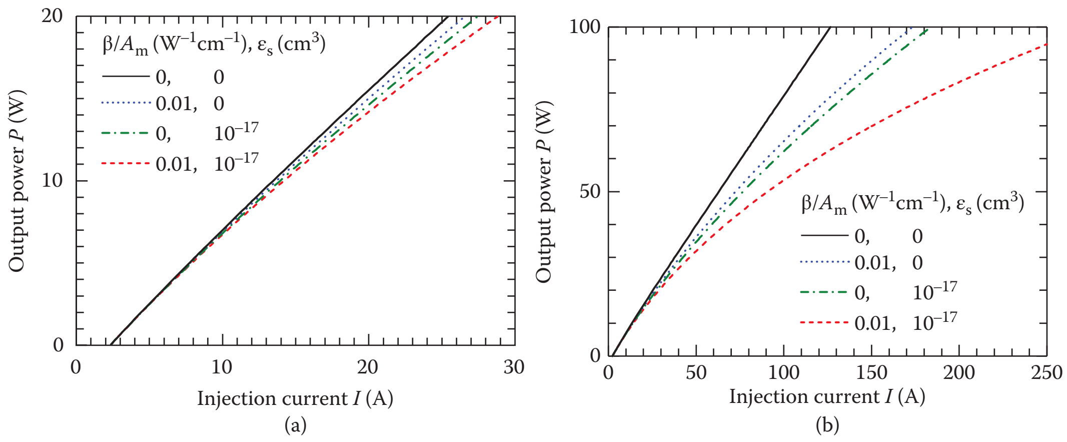


FIGURE 27.9 Impact of nonlinear-gain compression ϵ_{s} and two-photon absorption $\bar{\beta}_{\mathrm{m}}/A_{\mathrm{m}}$ on the power–current characteristics. (a) Power range corresponding to CW operation. (b) Power range corresponding to pulsed operation.

Approximating the intensity profile of the vertical mode by a Gaussian with a full width d_{1/e^2} at $1/e^2$ maximum,

$$\phi^2(y) = e^{-8y^2/d^2_{1/e^2}} \tag{27.68}$$

the mode area is $A_\mathrm{m} = \frac{\sqrt{\pi}}{2} d_{1/e^2} W \approx d_{1/e^2} W$. Assuming an averaged two-photon absorption coefficient of $\bar{\beta}_\mathrm{m} = 15\ \mathrm{cm \cdot GW^{-1}}$ (cf. Figure 27.12a) and $d_{1/e^2} = 1.5\ \mu\mathrm{m}$ as appropriate for a high-power laser we obtain $\bar{\beta}_\mathrm{m}/d_{1/e^2} = 10^{-4}\ \mathrm{W^{-1}}$. For $W = 100\ \mu\mathrm{m}$ and $P = 10$ W this results in a modal absorption coefficient of $\alpha_{2,\mathrm{m}} = 0.1\ \mathrm{cm^{-1}}$. Note that the modal two-photon absorption is maximal at minimum d_{1/e^2}, i.e., for a strong optical confinement. The secondary two-photon loss mechanism due to intraband absorption caused by the free carriers generated by two-photon absorption is strongly dependent on the carrier lifetime and the effective drift-diffusion length. In most cases, the resulting absorption coefficient can be neglected, except in structures where the intensity peak of the vertical mode is not located at the position of the active layer [22].

Figure 27.9 shows the impact of nonlinear gain compression and two-photon absorption on the power–current characteristics using the model presented in Section 27.3.4 with the parameters given in Table 27.2. The following conclusions can be drawn: First, a gain compression factor $\epsilon_\mathrm{s} = 10^{-17}\ \mathrm{cm^3}$ (saturation power $P_\mathrm{s} = 150$ W) or a two-photon absorption coefficient $\bar{\beta}/A_\mathrm{m} = 0.01\ (\mathrm{Wcm})^{-1}$ act very similar. Second, at a current of $I = 15$ A, which is typical for lasers operating in CW mode, the combination of both effects results in a reduction of the optical power by 10% (2 W; cf. Figure 27.9a). However, at a current of $I = 125$ A, which can be reached under pulsed operation [19], the power is reduced from 100 to 60 W, which seems a little bit too large, considering that leakage currents are not accounted for at all. Hence, either the gain compression factor or the two-photon absorption coefficient or both are smaller in reality than assumed here.

27.3.5.3 Leakage Currents

Leakage currents caused by the transport of carriers into regions without stimulated recombination can be divided into vertical and lateral ones. Vertical leakage currents are minority currents caused by the accumulation and subsequent recombination of electrons and holes in the p-doped and n-doped, respectively, optical confinement layers. This effect results in an increased free-carrier absorption, too. An analytical investigation of the carrier accumulation in the confinement layers and its impact on the power–current characteristics can be found in Reference [23] and an exact treatment based on the numerical solution of the drift-diffusion equations has been given in Reference [11,24].

The electron current flowing into the p-doped region can be obtained by integrating the continuity equation for the electron current density (Equation 27.137) over the p-doped region between the upper boundary of the active region $y = y_p$ and the p-metallization for the steady state,

$$\int_{\partial(\text{p-region})} \boldsymbol{j}_n \cdot \boldsymbol{n} dS = e \int_{\text{p-region}} R\, dV \tag{27.69}$$

where ∂(p-region) denotes the surface of the p-doped region with $\boldsymbol{n}$ being the normal vector. Assuming that the normal components of the electron current density vanish at the outer boundaries including the p-contact, we obtain the electron leakage current

$$I_\mathrm{n,leak} = \iint j_{n,y}|_{y=y_p}\, \mathrm{d}x\mathrm{d}z. \tag{27.70}$$

Hence, the electron leakage current can be calculated either from the electron current density at the boundary between active region and p-doped region or by integrating the recombination rate over the corresponding volume.

Figure 27.10a and b shows the vertical profiles of the current densities and the recombination rate of the asymmetric super-large optical-cavity structure published in [10] at a current of $I = 15$ A. Figure 27.10a reveals that for a properly designed structure the hole current flowing into the n-doped region can be neglected. The leakage current calculated from the electron current density at $y = 1.9$ μm is $I_{\mathrm{n,leak}} = 0.44$ A, which is also obtained from the integration of the recombination rate. The relative fraction of the leakage current is $(I - I_{\mathrm{n,leak}})/I = 0.97$, which coincides with the internal efficiency $\eta_{\mathrm{i}} = 0.97$, obtained from the length-dependent measurement of the external differential efficiency as described in 27.2.

With increasing bias the leakage current increases due to the bending of the band edges as shown in Reference [11], which contributes to the roll-over of the power–current characteristics. The electron leakage current can be reduced by an increase of the energy gap and doping level or a decrease of the thickness of the p-doped optical confinement layer as it is the case in extreme-asymmetric super large optical cavities. However, one should keep in mind that an increase of the energy gap results in a decrease of the electrical conductivity in the AlGaAs material system, an increase of the doping results in an increase of the free-carrier absorption and an increase of the waveguide asymmetry results in a decrease of the field intensity at the position of the active region.

The lateral leakage current $I_{\mathrm{p,spead}}$ caused by current spreading in the p-doped region can be reduced by an insulation of the highly doped p-contact and p-cladding layers beyond the p-metallization by implantation, diffusion of impurities, or etching as indicated in Figure 27.1 or by implementing a reverse-biased p-n junction. The lateral leakage current due to the diffusion of the carriers along the active region is more difficult to mitigate.

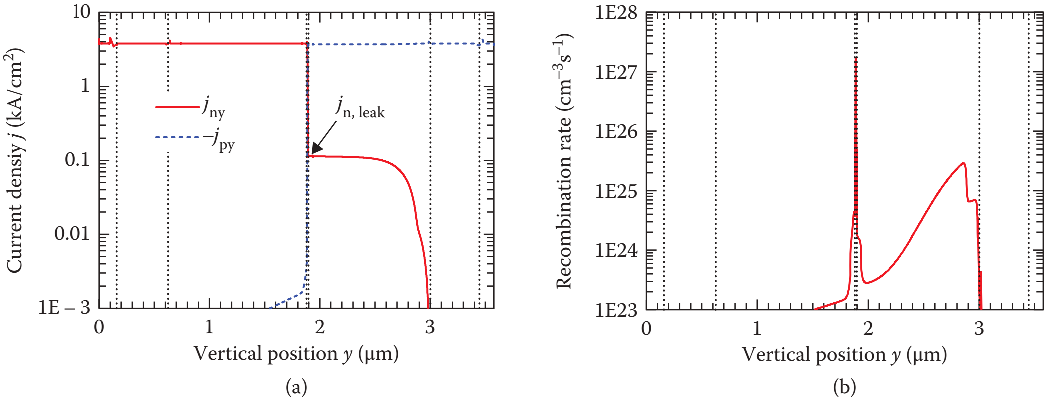
An approximate expression for $I_{\mathrm{p,spead}}$ has been derived in Reference [25],

$$\frac{I_{\mathrm{p,spead}}}{L} \approx -2\frac{\partial \varphi_{\mathrm{F}}}{\partial x}\Big|_{x=W/2} \sum_i \sigma_{\mathrm{p},i} d_i \tag{27.71}$$

where φ_{F} is the Fermi voltage given in Equation 27.60, $\sigma_{\mathrm{p},i}$ the hole conductivity of layer i and d_i the corresponding thickness. The derivative has to be taken at the edge of the p-contact stripe, and the summation includes all p-doped continuous semiconductor layers above the active region.

27.3.5.4 Carrier Capture

As it is well known [26–30], the carriers belonging to bound states in the QW(s) of the active region are not necessarily in thermal equilibrium with the carriers resulting from the continuum states located



(a) (b)

FIGURE 27.10 Vertical profiles of (a) current densities and (b) recombination rate for p-contact width $W = 100$ μm, cavity length $L = 4000$ μm, and output power $P = 15$ W. Heteroboundaries are indicated by dotted vertical lines. Note the logarithmic scales of the ordinates.

energetically above the QWs. It can be assumed that both types of carriers coexisting in the QW region are in quasi-equilibrium among themselves described by Fermi–Dirac statistics with independent quasi-Fermi potentials and transferred into each other by capture-escape processes. The transport of both types of carriers in the in-plane directions (x, z) can be described by classical drift diffusion. Most conveniently the bound carriers are described by sheet densities (unit $1/\mathrm{m}^2$) and appear in the rhs of the Poisson equation (27.141) as interface states.

An expression for the capture–escape rate that fulfils universal conditions such as the balance between capture and escape in case of equal quasi-Fermi potentials and no capture but finite escape in case of a fully occupied bound states has been given in [30]. For electrons (similarly for holes) it reads

$$Q_n = \frac{dn_\mathrm{f}}{\tau_\mathrm{c}}\left[1 - \frac{n_\mathrm{b}}{N_\mathrm{b}}\right]\left[1 - e^{\frac{e(\varphi_\mathrm{f}-\varphi_\mathrm{b})}{kT}}\right] \tag{27.72}$$

where n_f and n_b are the free and bound, electron densities, respectively; φ_f and φ_b are the corresponding quasi-Fermi potentials; d is the thickness of the QW(s); τ_c is the capture time; and N_b is the maximum bound electron density ($\varphi_\mathrm{b} \rightarrow -\infty$). Whereas the quasi-Fermi potentials of the free carriers depend directly on the bias applied to the contacts, the quasi-chemical potentials of the bound carriers are determined by the capture-escape rates.

The impact of the nonequilibrium between confined and unconfined carriers in high-power lasers on the power–current characteristics as a function of the hole capture time has been investigated in Reference [28], using, however, a capture-escape rate that differs from the one given in Equation 27.72.

27.4 Model for the Optical Field

We derive the parabolic paraxial wave equation based on the slowly varying amplitude and rotating wave approximations, taking into account gain dispersion, spontaneous emission, and a third-order nonlinear susceptibility. A balance equation for the electromagnetic energy density will be given.

27.4.1 Basic Three-Dimensional Equations

We start with the homogeneous Maxwell equation for the electric field $\boldsymbol{E}$ and the magnetic field $\boldsymbol{H}$:

$$\nabla \cdot \boldsymbol{D} = 0 \tag{27.73}$$

$$\nabla \cdot \boldsymbol{H} = 0 \tag{27.74}$$

$$\nabla \times \boldsymbol{E} + \mu_0 \partial_t \boldsymbol{H} = 0 \tag{27.75}$$

$$\nabla \times \boldsymbol{H} - \partial_t \boldsymbol{D} = 0 \tag{27.76}$$

where $\boldsymbol{D} = \varepsilon_0 \boldsymbol{E} + \boldsymbol{P}$ denotes the electric displacement and $\boldsymbol{P}$ the macroscopic polarization density of the material. The fields $\boldsymbol{E}$, $\boldsymbol{H}$, $\boldsymbol{D}$ depend on three spatial variables $\boldsymbol{r} = (x, y, z)$ and the time t. By applying $\nabla\times$ to Equation 27.75, ∂_t to Equation 27.76, and using Equation 27.73 we get the wave equation

$$\frac{1}{\varepsilon_0}\nabla(\nabla \cdot \boldsymbol{P}) + \nabla^2 \boldsymbol{E} = \frac{1}{c^2}\partial_{tt}\boldsymbol{E} + \mu_0 \partial_{tt}\boldsymbol{P} \tag{27.77}$$

The polarization density $\boldsymbol{P}$ contains a linear convolution of the electric field history $\boldsymbol{E}(t - \tau)$, $\tau \geq 0$, with a susceptibility function $\chi(r, \tau)$ and a second-order part which models dispersion ($\boldsymbol{P}_\mathrm{disp}$), nonlinear effects

($\boldsymbol{P}_{\mathrm{NL}}$), and spontaneous emission ($\boldsymbol{P}_{\mathrm{sp}}$):

$$\boldsymbol{P}(\boldsymbol{r},t)=\varepsilon_0\int_0^{\infty}\chi(r,\tau)\boldsymbol{E}(\boldsymbol{r},t-\tau)\,\mathrm{d}\tau+\boldsymbol{P}_{\mathrm{disp}}(\boldsymbol{r},t)+\boldsymbol{P}_{\mathrm{NL}}(\boldsymbol{r},t)+\boldsymbol{P}_{\mathrm{sp}}(\boldsymbol{r},t) \tag{27.78}$$

The spontaneous emission is modeled by introducing a spontaneous current density,

$$\partial_t\boldsymbol{P}_{\mathrm{sp}}(\boldsymbol{r},t)=\boldsymbol{J}_{\mathrm{sp}}(\boldsymbol{r},t) \tag{27.79}$$

We assume that $\boldsymbol{P}$ either is slowly varying in space, so that $\nabla\cdot\boldsymbol{P}\approx 0$ holds, or has discontinuities at boundaries between different materials. At these boundaries $\boldsymbol{E}$ and $\boldsymbol{H}$ have to fulfill certain continuity rules that can be derived from Maxwell equations. Due to the nearly planar geometry of edge-emitting lasers defined by the epitaxial layer structure, the electromagnetic field is either mainly transverse electric (TE) or transverse magnetic (TM) polarized. By assuming TE polarization and choosing a reference frequency ω_0, we have

$$\begin{aligned}\boldsymbol{E}(\boldsymbol{r},t)&=\boldsymbol{e}_x\frac{1}{2}E(\boldsymbol{r},t)\mathrm{e}^{i\omega_0 t}+\text{c.c.}\\ \boldsymbol{P}_{\mathrm{disp}}(\boldsymbol{r},t)&=\boldsymbol{e}_x\frac{1}{2}P_{\mathrm{disp}}(\boldsymbol{r},t)\mathrm{e}^{i\omega_0 t}+\text{c.c.}\\ \boldsymbol{P}_{\mathrm{NL}}(\boldsymbol{r},t)&=\boldsymbol{e}_x\frac{1}{2}P_{\mathrm{NL}}(\boldsymbol{r},t)\mathrm{e}^{i\omega_0 t}+\text{c.c.}\\ \boldsymbol{J}_{\mathrm{sp}}(\boldsymbol{r},t)&=\boldsymbol{e}_x\frac{1}{2}j_{\mathrm{sp}}(\boldsymbol{r},t)\mathrm{e}^{i\omega_0 t}+\text{c.c.}\end{aligned} \tag{27.80}$$

where $\boldsymbol{e}_x$ denotes the unit vector in x direction, c.c. stands for complex conjugation, and $E, P_{\mathrm{disp}}, P_{\mathrm{NL}}, j_{\mathrm{sp}}$ are scalar complex-valued functions. In the slowly varying amplitude approximation the corresponding derivatives are given by

$$\begin{aligned}\partial_{tt}\boldsymbol{E}&\approx-\boldsymbol{e}_x\frac{1}{2}\left(\omega_0^2E-2i\omega_0\partial_tE\right)\mathrm{e}^{i\omega_0 t}+\text{c.c.}\\ \partial_{tt}\boldsymbol{P}_{\mathrm{disp}}&\approx-\boldsymbol{e}_x\frac{1}{2}\omega_0^2P_{\mathrm{disp}}\mathrm{e}^{i\omega_0 t}+\text{c.c.}\\ \partial_{tt}\boldsymbol{P}_{\mathrm{NL}}&\approx-\boldsymbol{e}_x\frac{1}{2}\omega_0^2P_{\mathrm{NL}}\mathrm{e}^{i\omega_0 t}+\text{c.c.}\\ \partial_{tt}\boldsymbol{P}_{\mathrm{sp}}&\approx\boldsymbol{e}_x\frac{1}{2}i\omega_0j_{\mathrm{sp}}\mathrm{e}^{i\omega_0 t}+\text{c.c.}\end{aligned} \tag{27.81}$$

Let $\chi(\boldsymbol{r},\omega)$ denote the Fourier transform of $\chi(\boldsymbol{r},t)$ evaluated at frequency ω. First we insert Equation 27.80 into Equation 27.78 and replace $E(\boldsymbol{r},t-\tau)$ by its first-order approximation $E(\boldsymbol{r},t)-\partial_tE(\boldsymbol{r},t)\tau$. Then we insert the result and Equation 27.80 into Equation 27.77 taking into account Equation 27.81. Finally, we multiply by $e^{-i\omega_0 t}$ and neglect the rapidly varying terms containing $e^{-i2\omega_0 t}$ (rotating wave approximation) and get

$$\begin{aligned}\Delta E=&-\frac{\omega_0^2}{c^2}E+\frac{2i\omega_0}{c^2}\partial_tE-\frac{\omega_0^2}{c^2}\chi(\boldsymbol{r},\omega_0)E+\frac{i\omega_0}{c^2}\left(2\chi(\boldsymbol{r},\omega_0)+\omega_0\partial_\omega\chi(\boldsymbol{r},\omega)_{|\omega=\omega_0}\right)\partial_tE\\ &-\mu_0\omega_0^2\left(P_{\mathrm{disp}}+P_{\mathrm{NL}}\right)+i\mu_0\omega_0j_{\mathrm{sp}}\end{aligned} \tag{27.82}$$

Consistent with the neglection of $\partial_{tt}E(\boldsymbol{r},t)$ we approximate in front of $\partial_t E(\boldsymbol{r},t)$:

$$2\chi(\boldsymbol{r},\omega_0) + \omega_0\partial_\omega\chi(\boldsymbol{r},\omega)_{|\omega=\omega_0} \approx 2\bar{\chi}(\omega_0) + \omega_0\partial_\omega\bar{\chi}(\omega)_{|\omega=\omega_0} \tag{27.83}$$

where $\bar{\chi}$ is an averaged real-valued susceptibility of the medium. This means that the spatial variation and the imaginary part of χ is considered to be a first-order correction, which can be neglected in front of the time derivative.

Introducing the complex refractive index $n^2(\boldsymbol{r},\omega) = 1 + \chi(\boldsymbol{r},\omega)$, the real-valued reference index $\bar{n}^2 = 1 + \bar{\chi}$, the real-valued group index $n_{\rm g} = \bar{n} + \omega_0\partial_\omega\bar{n}(\omega)_{|\omega=\omega_0}$ and $k_0 = \omega_0/c$, we get

$$\Delta E = 2ik_0\frac{\bar{n}n_{\rm g}}{c}\partial_t E - k_0^2 n^2 E - \mu_0\omega_0^2\left(P_{\rm disp} + P_{\rm NL}\right) + i\mu_0\omega_0 j_{\rm sp} \tag{27.84}$$

Next we remove the rapid oscillations along the longitudinal z direction with the Ansatz

$$E(\boldsymbol{r},t) = E^+(\boldsymbol{r},t){\rm e}^{-i\bar{n}k_0 z} + E^-(\boldsymbol{r},t){\rm e}^{i\bar{n}k_0 z} \tag{27.85}$$

Inserting Equation 27.85 into Equation 27.84, neglecting $\partial_{zz}E^\pm$, multiplying the result with ${\rm e}^{\pm i\bar{n}k_0 z}$, dropping rapidly varying terms containing ${\rm e}^{\pm 2i\bar{n}k_0 z}$, and dividing by $2i\bar{n}k_0$, we get

$$\begin{aligned}\frac{1}{v_{\rm g}}\partial_t E^\pm(\boldsymbol{r},t) \pm \partial_z E^\pm(\boldsymbol{r},t) &= \frac{1}{2i\bar{n}k_0}\left[(\partial_x)^2 + (\partial_y)^2\right]E^\pm(\boldsymbol{r},t) - ik_0\frac{n^2(\boldsymbol{r},\omega_0) - \bar{n}^2}{2\bar{n}}E^\pm(\boldsymbol{r},t)\\ &+ \frac{\mu_0\omega_0^2}{2i\bar{n}k_0}\left[P_{\rm disp}(\boldsymbol{r},t) + P_{\rm NL}(\boldsymbol{r},t)\right]{\rm e}^{\pm i\bar{n}k_0 z} - \frac{\mu_0\omega_0}{2\bar{n}k_0}j_{\rm sp}(\boldsymbol{r},t){\rm e}^{\pm i\bar{n}k_0 z}\end{aligned} \tag{27.86}$$

where $v_{\rm g} = c/n_g$.

Equation 27.86 must be supplemented by appropriate boundary conditions. At the plane facets of the laser located at $z = 0$ and $z = L$

$$\begin{aligned}E^+(x,y,0,t) - r_0 E^-(x,y,0,t) &= 0\\ E^-(x,y,L,t) - r_L e^{-i2\bar{n}k_0 L}E^+(x,y,L,t) &= 0\end{aligned} \tag{27.87}$$

hold. We should mention that in the paraxial approximation, the amplitude reflection coefficients r_0 and r_L are input parameters, which have to be calculated in advance.

At the transverse boundary denoted by Γ one can assume, for example, decaying fields or a perfect electric wall,

$$\lim_{|r_{\rm t}|\to\infty} E^\pm = 0 \quad \text{or} \quad E^\pm|_{r_{\rm t}\in\Gamma} = 0, \tag{27.88}$$

respectively. If only a part of the cross section of the cavity is simulated, a nonreflecting or transparent boundary condition has to be used, which models the fact that only outgoing waves should be present. A very popular method to implement a boundary condition of this type is the introduction of a so-called perfectly matched layer (PML) [31]. However, in the frequency domain, it results in a large number of spurious modes [32]. Within the slowly varying approximation discontinuities of n can only be treated by employing corresponding transition conditions, except for resonant Bragg waveguide gratings, which can be described by extra terms coupling the forward and backward propagating waves.

27.4.2 Ansatz for Dispersion, Nonlinear Susceptibility, and Spontaneous Emission

27.4.2.1 Dispersion

We use the following Ansatz for $P_{\mathrm{disp}}(\boldsymbol{r}, t)$:

$$P_{\mathrm{disp}}(\boldsymbol{r}, t) = -i\frac{n_{\mathrm{r}}(\boldsymbol{r}, t)g_{\mathrm{r}}(\boldsymbol{r}, t)k_0}{\omega_0^2\mu_0}\left[\mathcal{P}(\boldsymbol{r}, t) - E(\boldsymbol{r}, t)\right] \tag{27.89}$$

where the electric field E is subtracted to ensure that the contribution to the dispersion vanishes at the gain peak. We model dispersion via the following response function in the frequency domain [33–35]

$$\mathcal{P}(\boldsymbol{r}, \omega) = \mathfrak{L}(\boldsymbol{r}, \omega)E(\boldsymbol{r}, \omega) \tag{27.90}$$

with

$$\mathfrak{L}(\boldsymbol{r}, \omega) = \frac{\gamma(\boldsymbol{r})}{i[\omega - \omega_0 - (\omega_{\mathrm{p}}(\boldsymbol{r}) - \omega_0)] + \gamma(\boldsymbol{r})} \tag{27.91}$$

The Lorentzian $\mathfrak{L}(\boldsymbol{r}, \omega)$ achieves its maximum value 1 at the frequency ω_{p}. The imaginary part $\mathrm{Im}\,\mathfrak{L}(r, \omega)$ has a half width γ at half of the maximum. The approximation is valid only within a small frequency region around ω_0 corresponding to the frequency range of optical transitions in the active material. By multiplying Equation 27.90 with the denominator and taking the inverse Fourier transform with respect to $\omega - \omega_0$ we get the ordinary differential equation

$$\partial_t\mathcal{P}(\boldsymbol{r}, t) = \left[i(\omega_{\mathrm{p}}(\boldsymbol{r}) - \omega_0) - \gamma(\boldsymbol{r})\right]\mathcal{P}(\boldsymbol{r}, t) + \gamma(\boldsymbol{r})E(\boldsymbol{r}, t) \tag{27.92}$$

We use similar decompositions for the polarization $\mathcal{P}(\boldsymbol{r}, t)$ as for the electric field $E(\boldsymbol{r}, t)$,

$$\mathcal{P}(\boldsymbol{r}, t) = \mathcal{P}^{+}(\boldsymbol{r}, t)\mathrm{e}^{-i\tilde{n}k_0 z} + \mathcal{P}^{-}(\boldsymbol{r}, t)\mathrm{e}^{i\tilde{n}k_0 z} \tag{27.93}$$

Inserting Equations 27.93 and 27.85 into Equation 27.92, multiplying with $\mathrm{e}^{\pm i\tilde{n}k_0 z}$, and again neglecting $\mathrm{e}^{\pm 2i\tilde{n}k_0 z}$ we get

$$\partial_t\mathcal{P}^{\pm}(\boldsymbol{r}, t) = i(\omega_{\mathrm{p}}(\boldsymbol{r}) - \omega_0)\mathcal{P}^{\pm}(\boldsymbol{r}, t) + \gamma(\boldsymbol{r})\left[E^{\pm}(\boldsymbol{r}, t) - \mathcal{P}^{\pm}(\boldsymbol{r}, t)\right] \tag{27.94}$$

Dispersion can also be taken into account with higher order time derivatives [36], on a microscopic level [37–39], by means of a digital filter [40], or by a convolution integral [41].

27.4.2.2 Nonlinear Susceptibility

Considering a third-order nonlinearity and assuming an isotropic medium the nonlinear polarization density reads [42–44]

$$P_{\mathrm{NL}}(\boldsymbol{r}, t) = \varepsilon_0\frac{3}{4}\chi^{(3)}_{xxxx}(\boldsymbol{r}, t)|E(\boldsymbol{r}, t)|^2E(\boldsymbol{r}, t) \tag{27.95}$$

with the third-order susceptibility $\chi^{(3)}_{xxxx}(\omega_0; -\omega_0, -\omega_0, \omega_0)$. The optical Kerr coefficient $\tilde{n}_2$ and the two-photon absorption coefficient $\tilde{\beta}$ are defined as

$$\tilde{n}_2 - i\frac{\tilde{\beta}}{2k_0} = \frac{3}{8n_\mathrm{r}}\chi^{(3)}_{xxxx} \equiv \Delta\tilde{n}. \tag{27.96}$$

The connection with the commonly used coefficients n_2 and β related to the time-averaged intensity $I = \varepsilon_0\bar{n}c|E|^2/2$ and discussed in Section 27.5.2 is given by

$$n_2 - i\frac{\beta}{2k_0} = \frac{2\Delta\tilde{n}_2}{\varepsilon_0\bar{n}c} \equiv \Delta n_2 \tag{27.97}$$

where n_2 has the unit m^2/W and β the unit m/W.

27.4.2.3 Spontaneous Emission

The stochastic forces

$$F^{\pm}_{\mathrm{sp}}(\boldsymbol{r}, t) = -\frac{\mu_0\omega_0}{2\bar{n}k_0}j_{\mathrm{sp}}(\boldsymbol{r}, t)\mathrm{e}^{\pm i\bar{n}k_0 z} \tag{27.98}$$

in Equation 27.86 have the properties

$$\langle F^{+}_{\mathrm{sp}}(\boldsymbol{r}, t)\rangle = \langle F^{-}_{\mathrm{sp}}(\boldsymbol{r}, t)\rangle = 0 \tag{27.99}$$

$$\langle F^{+}_{\mathrm{sp}}(\boldsymbol{r}, t)F^{+*}_{\mathrm{sp}}(\boldsymbol{r}, t')\rangle = \langle F^{-}_{\mathrm{sp}}(\boldsymbol{r}, t)F^{-*}_{\mathrm{sp}}(\boldsymbol{r}', t')\rangle = \frac{2\hbar\omega_0 n_\mathrm{r}(\boldsymbol{r}, t)g(\boldsymbol{r}, t)n_{\mathrm{sp}}(\boldsymbol{r}, t)}{\bar{n}^2\varepsilon_0 c}\delta(\boldsymbol{r} - \boldsymbol{r}')\delta(t - t') \tag{27.100}$$

where $\langle\rangle$ denotes ensemble average and $\delta(x)$ is the Dirac delta function. Equation 27.100 is not suited for a numerical evaluation, because in dependence on the carrier densities the inversion (or spontaneous emission) factor n_{sp} has a singularity when the gain g changes its sign (transparency) so that the product $g \cdot n_{\mathrm{sp}}$ is undefined. One possibility to circumvent this is to take the second moment (Equation 27.100) proportional to $\beta_{\mathrm{sp}}R_{\mathrm{sp}}$ where R_{sp} given in Equation 27.139 is the rate of spontaneous emission into all modes and the dimensionless factor β_{sp} is chosen such that at threshold the correct values of the second moments are obtained [8].

27.4.3 Final Field Equation and Balance of Radiative Energy

Summarizing the results of the previous subchapter, the paraxial wave equation can be written as

$$\begin{aligned}\frac{1}{v_\mathrm{g}}\partial_t E^{\pm}(\boldsymbol{r}, t) \pm \partial_z E^{\pm}(\boldsymbol{r}, t) = &-\frac{i}{2\bar{n}k_0}\left(\partial_{xx} + \partial_{yy}\right)E^{\pm}(\boldsymbol{r}, t)\\ &- \frac{ik_0}{2\bar{n}}\Delta n^2(\boldsymbol{r}, t)E^{\pm}(\boldsymbol{r}, t) - \frac{n_\mathrm{r}(\boldsymbol{r}, t)g_\mathrm{r}(\boldsymbol{r}, t)}{2\bar{n}}\left[E^{\pm}(\boldsymbol{r}, t) - \mathcal{P}^{\pm}(\boldsymbol{r}, t)\right]\\ &- \frac{ik_0 n_\mathrm{r}(\boldsymbol{r}, t)}{\bar{n}}\Delta\tilde{n}_2(\boldsymbol{r}, t)\left[|E^{+}|^2 + |E^{-}|^2\right]E^{\pm}(\boldsymbol{r}, t) + F^{\pm}_{\mathrm{sp}}(\boldsymbol{r}, t)\end{aligned} \tag{27.101}$$

where

$$\Delta n^2(\boldsymbol{r},t) \equiv n^2(\boldsymbol{r},t,\omega_0) - \bar{n}^2 = \Delta n_{\mathrm{r}}^2 + i\frac{n_{\mathrm{r}}\,(g-\alpha)}{k_0} \tag{27.102}$$

Here Δn_{r}^2 is the real part of Δn^2, n_{r} the real part of n, g the coefficient of optical gain originating from the emission or absorption of photons due to transitions between the conduction and valence bands, and α the coefficient of absorption of photons due to transitions within the conduction and valence bands (such as free-carrier absorption and intervalence band absorption). Optical losses due to scattering on waveguide imperfections could be included in α, too.

Multiplying Equation 27.101 with the complex conjugate $E^{\pm *}$, multiplying the complex conjugate of Equation 27.101 with $E^{\pm}$ and adding all equations results in

$$\begin{aligned}\frac{1}{v_{\mathrm{g}}}\frac{d\|E\|^2}{dt} = {}& \frac{1}{\bar{n}k_0}\mathrm{Im}\left[\partial_x(E^{+*}\partial_x E^{+} + E^{-*}\partial_x E^{-}) + \partial_y(E^{+*}\partial_y E^{+} + E^{-*}\partial_y E^{-})\right] \\ & + \partial_z\left[|E^{-}|^2 - |E^{+}|^2\right] \\ & + \frac{n_{\mathrm{r}}(g-\alpha)}{\bar{n}}\|E\|^2 + \frac{n_{\mathrm{r}}g_{\mathrm{r}}}{\bar{n}}\left[\mathrm{Re}(E^{+*}\mathcal{P}^{+} + E^{-*}\mathcal{P}^{-}) - \|E\|^2\right] \\ & - \frac{n_{\mathrm{r}}\tilde{\beta}}{\bar{n}}\|E\|^4 + 2\mathrm{Re}(E^{+*}F_{\mathrm{sp}}^{+} + E^{-*}F_{\mathrm{sp}}^{-})\end{aligned} \tag{27.103}$$

with $\|E\|^2 = |E^{+}|^2 + |E^{-}|^2$. This equation can be interpreted as a balance equation for the radiative energy density

$$u_{\mathrm{rad}}(\boldsymbol{r},t) = \frac{\varepsilon_0\bar{n}n_{\mathrm{g}}}{2}\|E(\boldsymbol{r},t)\|^2 \tag{27.104}$$

Integrating Equation 27.103 over the device volume V and using the boundary conditions Equations 27.88 and 27.87 leads to

$$\begin{aligned}\frac{\mathrm{d}\int_V u_{\mathrm{rad}}\,\mathrm{d}V}{\mathrm{d}t} = {}& -P_{\mathrm{out}} + \frac{\varepsilon_0 c}{2}\int_V n_{\mathrm{r}}(g - g_{\mathrm{r}} - \alpha)\|E\|^2\,\mathrm{d}V \\ & + \frac{\varepsilon_0 c}{2}\int_V n_{\mathrm{r}}g_{\mathrm{r}}\left[\mathrm{Re}(E^{+*}\mathcal{P}^{+} + E^{-*}\mathcal{P}^{-})\right]\mathrm{d}V - \frac{\varepsilon_0 c}{2}\int_V n_{\mathrm{r}}\tilde{\beta}\|E\|^4\,\mathrm{d}V \\ & + \frac{\varepsilon_0 c}{\bar{n}}\int_V \mathrm{Re}(E^{+*}F_{\mathrm{sp}}^{+} + E^{-*}F_{\mathrm{sp}}^{-})\,\mathrm{d}V\end{aligned} \tag{27.105}$$

which is a generalization of Equation 27.31. The output power $P_{\mathrm{out}} = P_0 + P_L$ is given in Equation 27.30 where $R_0 = |r_0|^2$ and $R_L = |r_L|^2$. The optical power is

$$P^{\pm}(z) = \frac{\varepsilon_0\bar{n}c}{2}\iint |E^{\pm}(x,y,z)|^2\,\mathrm{d}x\mathrm{d}y. \tag{27.106}$$

Thus the first term in Equation 27.103 is the divergence of the transverse energy flux density, which is assumed to vansish on the transverse surface of the device. The second term gives the radiation leaving the cavity. The third and fourth terms describe the increase or decrease of the energy due to stimulated emission or absorption, respectively, and the last two terms due to two-photon absorption and spontaneous emission.

27.4.4 Cavity Modes, Beam Propagation Method, and Roundtrip Operator

The cavity modes are time-periodic solutions of the form $E^{\pm}_{\mathrm{m}}\exp(i\Omega_{\mathrm{m}}t)$, $\mathcal{P}^{\pm}\exp(i\Omega_{\mathrm{m}}t)$ and obey the linear equations

$$\pm\, i\partial_z E^{\pm}_{\mathrm{m}}(\boldsymbol{r}) = \left[\frac{\Omega_{\mathrm{m}}}{v_{\mathrm{g}}} + \mathbf{H}(\boldsymbol{r}, z)\right] E^{\pm}_{\mathrm{m}}(\boldsymbol{r}) \tag{27.107}$$

subject to the boundary conditions (Equations 27.87 and 27.88). The operator **H** is given by

$$\mathbf{H}(\boldsymbol{r}, z) = \frac{1}{2\bar{n}k_0}\left(\partial_{xx} + \partial_{yy}\right) + \frac{k_0}{2\bar{n}}\Delta n^2(\boldsymbol{r}) + i\frac{n_{\mathrm{r}}(\boldsymbol{r})g_{\mathrm{r}}(\boldsymbol{r})}{2\bar{n}}\left[\frac{\gamma(\boldsymbol{r})}{i\Omega_{\mathrm{m}} - i(\omega_{\mathrm{p}}(\boldsymbol{r}) - \omega_0) + \gamma(\boldsymbol{r})} - 1\right] \tag{27.108}$$

The nontrivial solutions of (Equation 27.107) may depend on time via the dependence of the complex-valued refractive index on temporally varying carrier densities and temperature. The complex-valued relative mode frequencies Ω_{m} are the eigenvalues and the mode profiles $E^{\pm}_{\mathrm{m}}$ are the eigenfunctions of Equation 27.107. The real parts of Ω_{m} give the wavelengths relative to the reference wavelength λ_0,

$$\Delta\lambda_m = \left.\frac{\mathrm{d}\lambda}{\mathrm{d}\omega}\right|_{\lambda_0} \mathrm{Re}(\Omega_{\mathrm{m}}) \tag{27.109}$$

and the imaginary parts describe the damping of the modes. For a passive cavity, $\mathrm{Im}(\Omega_{\mathrm{m}}) > 0$ must hold. Lasing modes of an active cavity are distinguished by vanishing damping, $\mathrm{Im}(\Omega_{\mathrm{m}}) = 0$, due to the balance of the outcoupling and internal losses and the gain.

It can be shown that the cavity modes fulfil an orthogonality relation which does not define a scalar product because of the non-Hermitian character of Equation 27.107. Due to the dispersion term in Equation 27.107 the orthogonality relation differs from that given in [6], but can be derived in the same manner.

If the operator **H** depends only on the transverse coordinates $\boldsymbol{r}_{\mathrm{t}}$, the solution of Equation 27.107 can be formally written as

$$E^{\pm}_{\mathrm{m}}(\boldsymbol{r}_{\mathrm{t}}, z') = \mathrm{e}^{\mp i(\frac{\Omega_{\mathrm{m}}}{v_{\mathrm{g}}} + \mathbf{H})(z'-z)} E^{\pm}_{\mathrm{m}}(\boldsymbol{r}_{\mathrm{t}}, z) \tag{27.110}$$

The numerical evaluation of Equation 27.110 is the basis of what is known as the beam propagation method (BPM) [45–47]. For the case of a spatially and temporally constant index, $n = \text{const.}$, Equation 27.110 can be evaluated exactly to yield

$$E^{\pm}_{\mathrm{m}}(\boldsymbol{r}'_{\mathrm{t}}, z') = \mathrm{e}^{\mp i(\frac{\Omega_{\mathrm{m}}}{v_{\mathrm{g}}} + \frac{k_0}{2\bar{n}}\Delta n^2)(z'-z)} \int G^{\pm}(\boldsymbol{r}'_{\mathrm{t}} - \boldsymbol{r}_{\mathrm{t}}, z' - z) E^{\pm}_{\mathrm{m}}(\boldsymbol{r}_{\mathrm{t}}, z)\, \mathrm{d}x\mathrm{d}y \tag{27.111}$$

with

$$G^{\pm}(\boldsymbol{r}'_{\mathrm{t}} - \boldsymbol{r}_{\mathrm{t}}, z' - z) = \Theta\big(\pm (z' - z)\big)\left[\sqrt{\pm\frac{i\bar{n}k_0}{2\pi(z' - z)}}\right]^2 \mathrm{e}^{\mp\frac{i\bar{n}k_0|\boldsymbol{r}'_{\mathrm{t}} - \boldsymbol{r}_{\mathrm{t}}|^2}{2(z'-z)}} \tag{27.112}$$

where Θ denotes the Heaviside step function. The integral equation (Equation 27.111) together with the propagator Equation 27.112 is known as Huygen's integral in the Fresnel approximation.

Based on Equation 27.110 and the boundary conditions (Equation 27.87), it is possible to construct roundtrip operators $\mathbf{M}^{\pm}$. One starts at some position $z = z_0$ within the cavity and performs a full roundtrip. Depending on whether we start into forward (+) or backward (−) directions, the eigenvalue problems

$$\mathbf{M}^{\pm}(\Omega_{\mathrm{m}})E^{\pm}_{\mathrm{m}}(\boldsymbol{r}_{\mathrm{t}}, z_0) = \mathrm{e}^{i\frac{2\Omega_{\mathrm{m}}}{v_{\mathrm{g}}}L} E^{\pm}_{\mathrm{m}}(\boldsymbol{r}_{\mathrm{t}}, z_0) \tag{27.113}$$

are obtained. Note that the roundtrip operators $\mathbf{M}^{\pm}$ depend on the eigenvalues Ω_{m} if dispersion is taken into account. The eigenfunctions of $\mathbf{M}^{\pm}$ are the mode distributions $E^{\pm}_{\mathrm{m}}(\boldsymbol{r}_{\mathrm{t}}, z_0)$ at the position z_0. A very popular method for solving Equation 27.113 is based on the Fox–Li approach, cf. [48] and the references therein. The idea is to choose a normalized, more or less arbitrary start distribution $E^{\pm}_{\mathrm{m}}(\boldsymbol{r}_{\mathrm{t}}, z_0)$ and to apply the roundtrip operator $\boldsymbol{M}^{\pm}$ recurrently until one arrives (hopefully) at a steady state. It is known that the algorithm fails if there are cavity modes having identical or nearly identical damping $\mathrm{Im}(\Omega_{\mathrm{m}})$, as it is the case in broad-area lasers. The association of the nonconvergence of the Fox–Li iteration with an dynamically unstable laser behavior should be done with care. Instead, for multimode high-power lasers, a time-dependent approach based on Equation 27.101 should be preferred, although it is numerically more challenging.

27.5 Models for Nonlinear Modes and Filamentation

The multipeaked and not diffraction-limited lateral field profile of wide-aperture semiconductor lasers has been a long-standing problem and has been investigated in the past by numerous authors [49–55]. Although the broadening of the far field of CW operating lasers with increasing power (also called far-field blooming) can be at least partially attributed to the thermal lensing effect [56–58], a complete picture of the origin and mechanism has not revealed yet.

One mechanism is believed to be due to carrier-induced antiguiding, i.e., the reduction of the refractive index with increasing carrier density. This leads to a self-focusing mechanism because the index increases in regions of high intensity due to a depletion of the injected carrier density and can result in the formation of what is sometimes called lasing filaments.

Another mechanism that could explain the multipeaked structure and the broadening of the farfield is the simultaneous lasing of a large number of waveguide modes, originating from a built-in or thermally induced waveguide. Indeed, recent experiments reveal, that even at currents several times above threshold the lateral modes can be clearly identified by spectrally resolved near- and far-field measurements [58,59]. However, with increasing current, a broadening of the individual modes and the appearance of new modes with broad near and far fields can be observed [60].

Here, we derive the equation for the description of the lateral field profile and discuss two basic nonlinearities due to the virtual transitions and due to lateral spatial hole burning.

27.5.1 Longitudinal–Lateral Approximate Projected Equations

First we project the basic equations onto the dominant vertical mode using the Ansatz

$$E^{\pm}(\boldsymbol{r}, t) = \sqrt{\frac{2d\hbar\omega_0}{\varepsilon_0 \bar{n} n_{\mathrm{g}}}}\,\phi(y, x, z)u^{\pm}(x, z, t) \tag{27.114}$$

where the mode profile $\phi(y, x, z)$ is normalized according to $\int |\phi|^2\, dy = 1$. The scaling in Equation 27.114 is chosen such that $|u^{\pm}|^2$ is a photon density (unit m^{-3}). The power is given by

$$P^{\pm} = \hbar\omega_0 v_{\mathrm{g}} d \int |u^{\pm}|^2\, dx \tag{27.115}$$

where d is the thickness of the active region.

For each (x, z), $\phi(y, x, z)$ is a solution of a vertical waveguide equation with a real-valued index profile n_0 not dependent on carrier densities and temperature,

$$\partial_{yy}\phi(x, y, z) + k_0^2 n_0^2(x, y, z)\phi(x, z, y) = k_0^2 n_{\mathrm{eff},0}^2(x, z)\phi(x, y, z) \tag{27.116}$$

where the real-valued effective index $n_{\mathrm{eff},0}$ is the eigenvalue of Equation 27.116. Inserting Equation 27.114 into Equation 27.101, multiplying with ϕ, and integrating along y yields the equations

$$\begin{aligned} \frac{1}{v_{\mathrm{g}}}\partial_t u^{\pm}(x, z, t) = {} & -\frac{i}{2\bar{n}k_0}\partial_{xx}u^{\pm}(x, z, t) \mp \partial_z u^{\pm}(x, z, t) - \frac{ik_0}{2\bar{n}}\left[n_{\mathrm{eff}}^2(x, z, t) - \bar{n}^2\right] u^{\pm}(x, z, t) \\ & - ik_0 \Delta n_{2,\mathrm{eff}}(x, z, t)\hbar\omega_0 v_{\mathrm{g}} d \|u(x, z, t)\|^2 u^{\pm}(x, z, t) \end{aligned} \tag{27.117}$$

where we omitted the dispersion and spontaneous emission terms for simplicity and

$$n_{\mathrm{eff}}^2 = n_{\mathrm{eff,r}}^2 + i\frac{\bar{n}\left(g_{\mathrm{eff}} - \alpha_{\mathrm{eff}}\right)}{k_0}, \tag{27.118}$$

$$n_{\mathrm{eff,r}}^2 = n_{\mathrm{eff},0}^2 + \int (n_{\mathrm{r}}^2 - n_0^2)|\phi|^2\, \mathrm{d}y, \tag{27.119}$$

$$g_{\mathrm{eff}} - \alpha_{\mathrm{eff}} = \frac{\int n_{\mathrm{r}}(g - \alpha)|\phi|^2\, \mathrm{d}y}{\bar{n}} \tag{27.120}$$

and

$$\Delta n_{2,\mathrm{eff}} = \frac{\int n_{\mathrm{r}} \Delta n_2 |\phi|^4\, dy}{\bar{n}} \approx \frac{\Delta \bar{n}_2}{d_{1/e^2}} \tag{27.121}$$

The second term in Equation 27.119 treats effects, which are not included in n_0 and thus $n_{\mathrm{eff},0}$, for example, the dependence of the real and imaginary parts of n_{r} on the carrier densities or the temperature. The vertical mode size d_{1/e^2} has been introduced in Equation 27.68. The approach sketched is called the "effective index method" in the semiconductor laser community. Inserting Equation 27.114 into the rate of stimulated recombination Equation 27.140 and averaging over the active region $1/d \int_0^d\, dy$ yields

$$R_{\mathrm{st}} = v_{\mathrm{g}}\, g_{\mathrm{eff}} \|u\|^2 \tag{27.122}$$

with $\|u\|^2 = |u^+|^2 + |u^-|^2$.

Equation 27.117 together with correspondingly projected equations for the polarization (Equation 27.94) and a lateral diffusion equation for the excess carriers have been successfully applied to the simulation of a large variety of high-power laser structures [61–66]. A discussion on numerical issues can be found in [67].

27.5.2 Nonlinearities due to Virtual Transitions

The energy gaps of the confinement and cladding layers of a laser structure are larger than the energy of the photons generated by stimulated emission. In this situation, the photon energy is too small to allow absorption of single-photons due to transitions of electrons between the valence and conduction bands, as intended. However, virtual transitions involving two photons are still possible, which result in a third-order susceptibility defined in Equation 27.96. An approximate expression for the dispersion of the two-photon absorption coefficient β for direct transitions has been given in [68]:

$$\beta(\omega) = K\frac{\sqrt{E_{\mathrm{p}}}}{n_{\mathrm{r}}^2 E_{\mathrm{g}}^3} F_2\left(\frac{\hbar\omega}{E_{\mathrm{g}}}\right) \tag{27.123}$$

where E_{p} is the energy equivalent of the momentum matrix element for direct transitions between the valence and conduction bands, E_{g} is the (direct) energy gap and

$$F_2(x) = \frac{(2x-1)^{1.5}}{(2x)^5} \quad \text{for} \quad 2x > 1 \tag{27.124}$$

The factor K can be considered to be a free parameter. The two-photon absorption coefficient $\beta = 260$ m/TW experimentally determined for GaAs at a wavelength of $\lambda = 1064$ nm [44] is obtained with $K = 41,200\ \mathrm{m/TW} \times \mathrm{eV}^3/\sqrt{\mathrm{eV}}$, using $E_{\mathrm{g}} = 1.42$ eV, $E_{\mathrm{p}} = 26.1$ eV, and $n_{\mathrm{r}} = 3.48$. Note the E_{g}^{-3} dependence of β.

Two-photon absorption as well as Raman and Stark effects result also in an intensity-dependent contribution to the real part of the refractive index expressed as the optical Kerr coefficient n_2,

$$n_2(\omega) = \tilde{K}\frac{\hbar c\sqrt{E_{\mathrm{p}}}}{2n_{\mathrm{r}}^2 E_{\mathrm{g}}^4} G_2\left(\frac{\hbar\omega}{E_{\mathrm{g}}}\right) \tag{27.125}$$

where the function $G_2(\hbar\omega/E_{\mathrm{g}})$ given in [68] is shown in Figure 27.11 together with the function $F_2(\hbar\omega/E_{\mathrm{g}})$. It can be seen, that G_2 is maximal around $\hbar\omega = 0.5E_{\mathrm{g}}$ and changes its sign at $\hbar\omega = 0.7E_{\mathrm{g}}$, where F_2 is

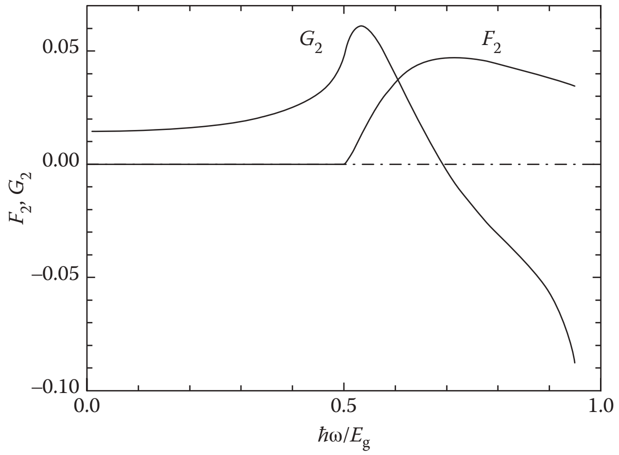
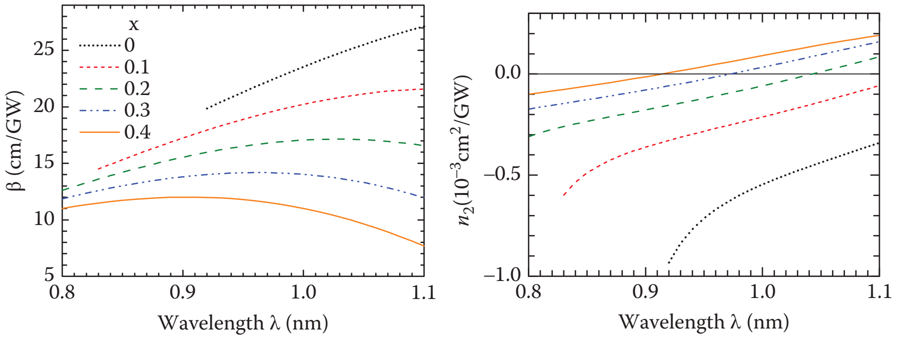


FIGURE 27.11 Functions F_2 and G_2 used in Equations 27.123 and 27.125.



(a)

(b)

FIGURE 27.12 Nonlinear coefficients calculated by Equations 27.123 and 27.125 versus wavelength for different Al compositions x in $\text{Al}_x\text{Ga}_{1-x}\text{As}$. (a) Two-photon absorption coefficient and (b) Optical Kerr coefficient.

maximal[†]. The factor $\tilde{K}$ can again be considered as a free parameter. The optical Kerr coefficient $n_2 = -4.1 \cdot 10^{-17}$ m/W experimentally determined for GaAs at a wavelength of $\lambda = 1064$ nm [44] is obtained with $\tilde{K} = 7.212 \cdot 10^{11}\ \text{m/W} \times 1/(\text{Ws m}) \times \text{eV}^4/\sqrt{\text{eV}}$.

The dependencies of β and n_2 of $\text{Al}_x\text{Ga}_{1-x}\text{As}$ on λ are shown in Figure 27.12a and 27.12b, respectively. The two-photon absorption increases from $\beta \propto 10$ cm/GW to $\beta \propto 25$ cm/GW if the Al composition is decreased from $x = 0.4$ to $x = 0$ due to the decrease of E_g. It shows a nonmonotonous dependence on the wavelength given by $F_2(\lambda)$. The optical Kerr coefficient decreases with increasing x and changes its sign in dependence on λ at higher Al compositions within the wavelength range investigated. For smaller Al compositions and wavelengths n_2 is negative, but for larger Al compositions and wavelengths n_2 is positive. For $\lambda = 980$ nm and relevant compositions $|n_2| < 2 \cdot 10^{-4}\ \text{cm}^2/\text{GW}$. Under CW operation, the maximum power density of broad-area lasers, which are state of the art during the writing of the book is of the order $10^{-2}\ \text{GW/cm}^2$. Hence the resulting index change $|\Delta n| < 4 \cdot 10^{-6}$ should not have a big impact on the optical field.

It should be noted that in nonisotropic media the relation between the nonlinear polarization and the electric field is more complicated than given in Equation 27.95. Hence the two-photon absorption and the optical Kerr effect depend on the crystallographic orientation and on the polarization direction of the optical field.

27.5.3 Nonlinearity Induced by Lateral Spatial Hole Burning

We now derive equations for the right and left traveling fields alone, eliminating the carrier density. This is only possible for a (hypothetical) steady state. To this end, the real and imaginary parts of Δn_{eff}^2 are linearized around a reference carrier density N_{ref}, which is typically set to the transparency carrier density N_{tr} were $g_{\text{eff}}(N_{\text{tr}}) = 0$. Other possible choices could be the carrier density for waveguide transparency (solution of $g_{\text{eff}}(N_{\text{ref}}) = \alpha_{\text{eff}}$) or the threshold carrier density of a laterally infinite laser (solution of $g_{\text{eff}}(N_{\text{thr}}) = \alpha_{\text{eff}} + \alpha_{\text{out}}$).

For the steady state, linearizing the recombination rate as in Section 27.3.2, neglecting drift-diffusion, and assuming a constant injection current density the rate equation can be solved for the excess carrier density,

---

[†] The functions F_2 and G_2 due to indirect transitions behave differently [69]: The maximum of F_2 occurs slightly above the indirect bandgap, and G_2 is positive throughout the transparent wavelength range and crosses zero only above the indirect gap. The two-photon absorption coefficient is dominated by direct transitions for $\hbar\omega > E_{\text{g,direct}}/2$ if $E_{\text{g,indirect}} > E_{\text{g,direct}}/2$.

$$N - N_{\mathrm{tr}} = \frac{\frac{j\tau_N}{ed} - N_{\mathrm{tr}}}{1 + v_{\mathrm{g}} g'_{\mathrm{eff}} \tau_N \|u_{\mathrm{s}}\|^2}. \tag{27.126}$$

From Equation 27.126 it can be seen that $N \to N_{\mathrm{tr}}$ for $\|u_{\mathrm{s}}\|^2 \to \infty$, i.e., the carrier density cannot be depleted below the transparency carrier density. Inserting Equation 27.126 into $\Delta n^2_{\mathrm{eff,r}}(N)$, $g_{\mathrm{eff}}(N)$, and $\alpha_{\mathrm{eff}}(N)$ yields

$$g_{\mathrm{eff}}(j) = \frac{g'_{\mathrm{eff}} \tau_N (j - j_{\mathrm{tr}})}{ed(1 + \epsilon_{\mathrm{sat}} \|u_{\mathrm{s}}\|^2)}, \tag{27.127}$$

$$\alpha_{\mathrm{eff}}(j) = \alpha_{\mathrm{eff}}(N_{\mathrm{tr}}) + \frac{\alpha'_{\mathrm{eff}} \tau_N (j - j_{\mathrm{tr}})}{ed(1 + \epsilon_{\mathrm{sat}} \|u_{\mathrm{s}}\|^2)}, \tag{27.128}$$

$$n^2_{\mathrm{eff,r}}(j) = n^2_{\mathrm{eff,r}}(N_{\mathrm{tr}}) + \frac{\alpha_{\mathrm{H}} \bar{n} g'_{\mathrm{eff}} \tau_N (j - j_{\mathrm{tr}})}{edk_0(1 + \epsilon_{\mathrm{sat}} \|u_{\mathrm{s}}\|^2)} \tag{27.129}$$

with the saturation parameter

$$\epsilon_{\mathrm{sat}} = v_{\mathrm{g}} g'_{\mathrm{eff}} \tau_N \tag{27.130}$$

and Henry's α-factor

$$\alpha_{\mathrm{H}} = 2k_0 n'_{\mathrm{eff}} / g'_{\mathrm{eff}}. \tag{27.131}$$

For typical values $n_{\mathrm{g}} = 4$, $g'_{\mathrm{eff}} = 10 \cdot 10^{-18}\ \mathrm{cm}^2$, and $\tau_N = 1$ ns we obtain $\varepsilon_{\mathrm{sat}} = 7.5 \cdot 10^{-17}\ \mathrm{cm}^3$. In general, $g'_{\mathrm{eff}} > 0$ and $n'_{\mathrm{eff}} < 0$ and hence $\alpha_{\mathrm{H}} < 0$ for frequencies around the gain peak. Typical values of α_{H} range from -1 to -10.

The stationary states correspond to time-harmonic solutions

$$u^{\pm}(x, z, t) = u^{\pm}_{\mathrm{s}}(x, z) e^{i\Omega t} \tag{27.132}$$

of Equation 27.117 with $\mathrm{Im}(\Omega) = 0$. Inserting Equations 27.132, 27.127, and 27.129 into Equation 27.117, we obtain

$$\begin{aligned} \pm i\partial_z u^{\pm}_{\mathrm{s}} = {} & \frac{1}{2\bar{n}k_0} \partial_{xx} u^{\pm}_{\mathrm{s}}(x, z, t) + \frac{\Omega}{v_{\mathrm{g}}} u^{\pm}_{\mathrm{s}} + \frac{k_0}{2\bar{n}} \left[n^2_{\mathrm{eff,r}}(N_{\mathrm{tr}}) - \bar{n}^2 \right] u^{\pm}_{\mathrm{s}} - \frac{i}{2} \alpha_{\mathrm{eff}}(N_{\mathrm{tr}}) u^{\pm}_{\mathrm{s}} \\ & + \frac{\tau_N \left[(i + \alpha_{\mathrm{H}}) g'_{\mathrm{eff}} - i\alpha'_{\mathrm{eff}} \right] \left(j - j_{\mathrm{tr}} \right)}{2ed(1 + \epsilon_{\mathrm{sat}} \|u_{\mathrm{s}}\|^2)} u^{\pm}_{\mathrm{s}} + k_0 \Delta n_{2,\mathrm{eff}} \hbar\omega_0 v_{\mathrm{g}} d \|u_{\mathrm{s}}\|^2 u^{\pm}_{\mathrm{s}} \end{aligned} \tag{27.133}$$

For $\varepsilon_{\mathrm{sat}} \|u\|^2 \ll 1$ we can expand

$$(1 + \epsilon_{\mathrm{sat}} \|u_{\mathrm{s}}\|^2)^{-1} \approx 1 - \epsilon_{\mathrm{sat}} \|u_{\mathrm{s}}\|^2 \tag{27.134}$$

so that the second-last term in Equation 27.133 resembles the last term with

$$\Delta n_{2,\mathrm{SHB}} = - \frac{\tau_N \left[(\alpha_{\mathrm{H}} + i) g'_{\mathrm{eff}} - i\alpha'_{\mathrm{eff}} \right] \left(j - j_{\mathrm{tr}} \right) \epsilon_{\mathrm{sat}}}{2ek_0 \hbar\omega_0 v_{\mathrm{g}} d^2} \tag{27.135}$$

For typical values $\alpha_H = -2$, $g'_{\text{eff}} = 10 \cdot 10^{-18}\ \text{cm}^2$, $j = 10\ \text{kA/cm}^2$, $d = 10\ \text{nm}$, $\tau_N = 1\ \text{ns}$, $j_{\text{tr}} = 0.16\ \text{kA/cm}^2$ (corresponding to $N_{\text{tr}} = 10^{18}\ \text{cm}^{-3}$), $\lambda_0 = 1\ \mu\text{m}$, $n_g = 4$, $\varepsilon_{\text{sat}} = 75 \cdot 10^{-18}\ \text{cm}^3$ we obtain $\text{Real}(\Delta n_{2,\text{SHB}}) = 5 \cdot 10^{-4}\ \text{cm/W}$, which is several orders of magnitudes larger than the absolute value of the effective optical Kerr coefficient due to non-resonant virtual transitions $\text{Real}(\Delta n_{2,\text{eff}}) \approx -7 \cdot 10^{-10}\ \text{cm/W}$ for $\bar{n}_2 = -10^{-13}\ \text{cm}^2/\text{W}$ (cf. Figure 27.12b) and $d_{1/e^2} = 1.5\ \mu\text{m}$ (cf. Equation 27.121).

Due to the fact that $\text{Real}(\Delta n_{2,\text{SHB}})$ is positive, two basic instabilities sometimes termed “filamentation” arise [44]. First, a local intensity maximum induces a variation of the real part of the effective index with a larger value at the position of the intensity peak than outside. Thus a local index waveguide is created (self-focusing) [49]. However, at the position of the intensity peak the effective gain is decreased which results in an reduced amplification in the local-waveguide core and an enhanced amplification outside (self-defocusing).

The other instability, first described by Bespalov and Talanov [70], is the breakup of a laser beam with a homogeneous intensity distribution into a beam with a random intensity distribution as a consequence of the growth of irregularities initially present on the laser wavefront. For a mathematical description of the effect, the forward and backward propagating waves have to be expressed as a sum of three plane-wave components each [44,71].

In [51,72] the mean-field approximation, $\bar{u}_s^{\pm} = 1/L \int_0^L u_s^{\pm}\, dz$ has been applied to Equation 27.133 and the resulting nonlinear second-order ordinary differential equation has been solved numerically. In both references, the case of a purely gain-guided laser is considered, i.e., any impact of a built-in or thermally induced index guide has not been investigated. In [72], a zoo of solutions of different types, including asymmetric ones, has been found. The basic stationary states are the linear modes with the wavelengths

$$\Delta\lambda_{\nu k} = -\frac{\lambda_0}{Lk_0 n_g}\left[\frac{\varphi_0 + \varphi_L}{2} + \pi k - Lk_0\bar{n} - \frac{Lk_0}{2\bar{n}}\text{Re}(n_\nu^2 - \bar{n}^2)\right] \tag{27.136}$$

where k denotes the longitudinal mode, n_ν is the modal index of the νth lateral mode, φ_0 and φ_L are the phases of the reflectivities. The other types of modes exist only due to the nonlinearity in the complex effective index and arise above the thresholds of the linear-guided modes.

There are a number of shortcomings of the theory presented so far. The assumption of a constant injection current density, i.e., an infinite series resistance as in Section 27.3.2 discussed, and the neglect of drift-diffusion result in an overestimation of spatial hole burning. A local variation of the complex effective index is not only caused by its dependence on the carrier density but also on the temperature ($\partial n/\partial T > 0$). Furthermore, in all analytical investigations, we are aware of no built-in or thermally induced index waveguides have been taken into account. These waveguides stabilize the linear guided modes, which are thus observable even far above threshold [60].

Based on an expansion of the optical field on the linear-guided modes and a numerical solution of the drift-diffusion equations in the mean-field approximation the far-field blooming of an index-guided broad-area laser has been investigated in [73]. The simulations revealed that a substantial part of the far-field blooming is not caused by self-heating but by increasing gain nonuniformity due to lateral spatial holeburning and laterally varying hole injection into the QWs. A discussion of filamentation effects based on a numerical solution of Equation 27.117 can be found in [53–55].

27.6 Thermodynamic-Based Energy-Transport Model

In high-power lasers operated in CW mode the transport of the charged carriers (electrons and holes) and the photons must be consistently formulated with the temperature flow in order to describe self-heating effects such as thermal roll-over and thermal lensing properly. A derivation of such a self-consistent energy-transport model by applying fundamental thermodynamic principles has been given in [74] using Boltzmann statistics. The model fulfils the first and second laws of thermodynamics as well as Onsager's reciprocity relations for the current densities. We note, that in this model electron, hole, and lattice temperatures are assumed to be equal, opposed to other energy-transport models [29,75]. A more general energy-transport model for semiconductor devices has been derived in [76] taking into account Fermi statistics and different temperatures for the charged carriers and the lattice, but disregarding optical fields. Previous formulations have been also given in [77,78], for example. In what follows we will summarize the energy-transport model paying particular attention to a consistent formulation with the model for the optical field presented in Section 27.4.

27.6.1 Basic Equations

The particle current flow is governed by the continuity equations for electrons and holes,

$$\nabla \boldsymbol{j}_n = +e\left(R + \partial_t n\right) \tag{27.137}$$

$$\nabla \boldsymbol{j}_p = -e\left(R + \partial_t p\right) \tag{27.138}$$

were $\boldsymbol{j}_n$ and $\boldsymbol{j}_p$ are the current densities for electrons and holes, respectively. The recombination rate R given in Equation 27.46 consists of nonradiative recombination $R_{\text{non-rad}}$, radiative spontaneous recombination R_{sp}, and radiative stimulated recombination R_{st}. The rate of radiative spontaneous recombination is often written as

$$R_{\text{sp}} = B(np - n_0 p_0) \tag{27.139}$$

with the equilibrium electron and hole densities n_0 and p_0, respectively. In the case of Boltzmann statistics, the coefficient B is constant whereas in the general case of Fermi statistics B decreases with increasing carrier densities.

The rate of stimulated recombination is the rate by which the energy density of the optical field changes by stimulated emission or absorption of a photon due to transitions between the conduction and valence bands, divided by the energy of the emitted or absorbed photon and follows from Equations 27.104 and 27.103 to

$$R_{\text{st}} = \frac{\varepsilon_0 c n_{\text{r}}}{2\hbar\omega_0}\left[g\|E\|^2 + g_{\text{r}} \sum_{\nu=+,-} \text{Re}\left(E^{\nu *}\mathcal{P}^{\nu} - E^{\nu *}E^{\nu}\right)\right] \tag{27.140}$$

The electrostatic field itself is affected by the charge distribution of mobile (n and p) and fixed (n_{A} and p_{D}) carrier densities. The corresponding electrostatic potential φ solves the Poisson equation

$$-\nabla(\varepsilon_0\varepsilon_{\text{s}}\nabla\varphi) = e(p - n + p_{\text{D}} - n_{\text{A}}) \tag{27.141}$$

with the relative static dielectric constant ε_{s}. For parabolic bands the relations between quasi-Fermi potentials φ_n and φ_p, electrostatic potential φ and carrier densities n and p are given by Equation 27.59.

In what follows we introduce the entropies per particle[†]

$$s_n = k_B \left[1 + R_n - \frac{\partial_T E_c}{k_B} + \frac{e\varphi_n + E_c - e\varphi}{k_B T}\right] \tag{27.142}$$

$$s_p = k_B \left[1 + R_p + \frac{\partial_T E_v}{k_B} - \frac{e\varphi_p + E_v - e\varphi}{k_B T}\right] \tag{27.143}$$

and the energies per particle

$$u_n = k_B T R_n - T\partial_T E_c + E_c = T(s_n - k_B) + e(\varphi - \varphi_n) \tag{27.144}$$

$$u_p = k_B T R_p + T\partial_T E_v - E_v = T(s_p - k_B) + e(\varphi_p - \varphi) \tag{27.145}$$

The functions R_n and R_p are given by the temperature derivatives of the inverse Fermi integrals,

$$R_n = -k_B T \partial_T \mathcal{F}_i^{\text{inv}}\left(\frac{n}{N_c}\right) \tag{27.146}$$

$$R_p = -k_B T \partial_T \mathcal{F}_i^{\text{inv}}\left(\frac{p}{N_v}\right) \tag{27.147}$$

For the bulk case ($i = 1/2$), Boltzmann statistics, parabolic bands, and temperature-independent electron and holes masses R_n and R_p are equal to $3/2$.

The electron and hole current densites are given by

$$\boldsymbol{j}_n = -\sigma_n(\nabla\varphi_n - P_n \nabla T) \tag{27.148}$$

$$\boldsymbol{j}_p = -\sigma_p(\nabla\varphi_p + P_p \nabla T) \tag{27.149}$$

where σ_n and σ_p are the electrical conductivities and P_n and P_p are the Seebeck coefficients or thermoelectric powers being the entropies per particle divided by the elementary charge e,

$$P_n = \frac{s_n}{e} \quad \text{and} \quad P_p = \frac{s_p}{e} \tag{27.150}$$

If the coefficients in front of the temperature derivatives in Equations 27.148 and 27.149 are derived from the Boltzmann equation in relaxation time approximation, then the factorization $\sigma_n \cdot P_n$ and $\sigma_p \cdot P_p$ is only possible for parabolic bands and Boltzmann statistics. The same holds for the factorization of the electrical conductivities into products of carrier-density independent mobilities and carrier densities, $\sigma_n = e\mu_n \cdot n$ and $\sigma_p = e\mu_p \cdot p$. Furthermore, the magnitudes of P_n, P_p similarly as μ_n, μ_p depend on the scattering processes involved [79]. For example, if the dependence of the relaxation time on the energy is given by $\tau_0[E/(k_B T)]^r$ where r ranges typically between $-3/2$ and $+3/2$, then $R_n = R_p = 3/2+r$ in Equations 27.142 and 27.143. Thus $R_n = R_p = 3/2$ holds only for an energy-independent relaxation time. The temperature derivatives $\partial_T E_c$ and $\partial_T E_v$ of the conduction and valence band edges are often not included in the definition of the Seebeck coefficients.

According to Reference [74] the heat flow equation reads

$$c_h \partial_t T - \nabla \kappa_L \nabla T = h \tag{27.151}$$

[†] The entropies can be also written as $s_n = e\partial_T(\varphi_n)_{(n,p)} + k_B$, $s_p = -e\partial_T(\varphi_p)_{(n,p)} + k_B$, where n, p are kept constant in the differentiation.

with the thermal conductivity of the crystal lattice κ_{L}, the heat capacity

$$c_{\mathrm{h}} = c_{\mathrm{L}} + n\partial_T u_n + p\partial_T u_{\mathrm{p}} \tag{27.152}$$

with c_{L} being the heat capacity of the lattice and the heat source density

$$\begin{aligned} h = {} & \frac{k_{\mathrm{B}}T}{e}\boldsymbol{\nabla}(\boldsymbol{j}_n - \boldsymbol{j}_p) + T(\boldsymbol{j}_n\boldsymbol{\nabla}P_n - \boldsymbol{j}_p\boldsymbol{\nabla}P_p) \\ & + \frac{1}{\sigma_n}\boldsymbol{j}_n^2 + \frac{1}{\sigma_p}\boldsymbol{j}_p^2 + (u_n + u_p)R - \partial_t u_{\mathrm{rad}} - \gamma_{\mathrm{rad}} \end{aligned} \tag{27.153}$$

The first term in Equation 27.153 is related to thermodiffusion and can be written as

$$\frac{k_{\mathrm{B}}T}{e}\boldsymbol{\nabla}\cdot(\boldsymbol{j}_n - \boldsymbol{j}_p) = 2k_{\mathrm{B}}TR + k_{\mathrm{B}}T(\partial_t n + \partial_t p) \tag{27.154}$$

taking into account Equations 27.137 and 27.138. The second term describes Thomson–Peltier heat,

$$h_{\mathrm{TP}} = T(\boldsymbol{j}_n\boldsymbol{\nabla}P_n - \boldsymbol{j}_p\boldsymbol{\nabla}P_p) = h_{\mathrm{Thomson}} + h_{\mathrm{Peltier}} \tag{27.155}$$

which is generated by a current flow along the gradients of the Seebeck coefficients ∇P_n and ∇P_p. By applying the chain rule to the gradients, the contributions due to Thompson heat and Peltier heat can be separated,

$$h_{\mathrm{Thomson}} = T\boldsymbol{\nabla}T\left(\boldsymbol{j}_n\partial_T P_n|_{(n,p)} - \boldsymbol{j}_p\partial_T P_p|_{(n,p)}\right) \tag{27.156}$$

$$h_{\mathrm{Peltier}} = T\left(\boldsymbol{j}_n\partial_n P_n|_T\boldsymbol{\nabla}n - \boldsymbol{j}_p\partial_p P_p|_T\boldsymbol{\nabla}p\right) \tag{27.157}$$

The third and forth terms correspond to Joule heat:

$$h_{\mathrm{J}} = \frac{1}{\sigma_n}\boldsymbol{j}_n^2 + \frac{1}{\sigma_p}\boldsymbol{j}_p^2 \tag{27.158}$$

generated by scattering of the carriers on phonons resulting in a energy loss to the lattice. The last term is due to contributions of the recombination of electron–hole pairs which sets free the energy $u_n + u_p$ that is either transferred to the lattice as heat or transferred to the radiative field. The latter part is described by the term $\partial_t u_{\mathrm{rad}}$, which has to be subtracted from the source term, like the term γ_{rad}, which denotes the energy loss from the cavity. Inserting Equations 27.154, 27.155, and 27.158 into Equation 27.153 gives

$$h = k_{\mathrm{B}}T(\partial_t n + \partial_t p) + h_{\mathrm{TP}} + h_{\mathrm{J}} + h_{\mathrm{rec+abs}} \tag{27.159}$$

with the recombination and absorption heat

$$h_{\mathrm{rec+abs}} = e(TP_n + TP_p + \varphi_p - \varphi_n)R - \partial_t u_{\mathrm{rad}} - \gamma_{\mathrm{rad}} \tag{27.160}$$

where $R = R_{\mathrm{non\text{-}rad}} + R_{\mathrm{sp}} + R_{\mathrm{st}}$. Likewise the rate of radiative recombination, the radiative energy density u_{rad} and the radiative energy loss γ_{rad} have contributions from spontaneous and stimulated emission, $u_{\mathrm{rad}} = u_{\mathrm{sp}} + u_{\mathrm{st}}$ and $\gamma_{\mathrm{rad}} = \gamma_{\mathrm{sp}} + \gamma_{\mathrm{st}}$. Here, u_{sp} is the energy density and γ_{sp} the corresponding cavity loss of the radiation generated by spontaneous emission into all modes (i.e., spatial directions, polarization directions of the field and frequencies) not included in $E^{\pm}$ given by Equation 27.101.

The balance equation for the energy density of the radiation generated by stimulated emission is given by Equations 27.103 and 27.104. As it has been stated before the first term on the rhs of Equation 27.103 gives the divergence of the transverse energy flux density and the second term the radiation leaving the cavity in propagation direction. Both terms combined give the total stimulated-energy loss from the cavity $-\gamma_{\mathrm{st}}$,

$$\begin{aligned}-\gamma_{\mathrm{st}} = {} & \frac{\varepsilon_0 c}{2k_0}\mathrm{Im}\left[\partial_x(E^{+*}\partial_x E^{+} + E^{-*}\partial_x E^{-}) + \partial_y(E^{+*}\partial_y E^{+} + E^{-*}\partial_y E^{-})\right] \\ & + \frac{\varepsilon_0 \bar{n} c}{2}\partial_z\left[|E^{-}(r,t)|^2 - |E^{+}(r,t)|^2\right] \end{aligned} \tag{27.161}$$

and hence

$$\partial_t u_{\mathrm{st}} = \frac{\varepsilon_0 c n_{\mathrm{r}}}{2}\left((g - g_r - \alpha)\|E\|^2 + g_r \mathrm{Re}(E^{+*}\mathcal{P}^{+} + E^{-*}\mathcal{P}^{-})\right) - \frac{\varepsilon_0 c n_{\mathrm{r}}\tilde{\beta}}{2}\|E\|^4 - \gamma_{\mathrm{st}}. \tag{27.162}$$

Inserting Equations 27.162 and 27.140 into Equation 27.160 gives

$$\begin{aligned} h_{\mathrm{rec+abs}} = {} & e(\mathrm{TP}_n + \mathrm{TP}_p + \varphi_p - \varphi_n)R_{\mathrm{non\text{-}rad}} + \frac{\varepsilon_0 c n_{\mathrm{r}}\alpha}{2}\|E\|^2 + \frac{\varepsilon_0 c n_{\mathrm{r}}\tilde{\beta}}{2}\|E\|^4 \\ & + \left[e(\mathrm{TP}_n + \mathrm{TP}_p + \varphi_p - \varphi_n) - \hbar\omega_0\right] R_{\mathrm{st}} \\ & + e(\mathrm{TP}_n + \mathrm{TP}_p + \varphi_p - \varphi_n)R_{\mathrm{sp}} - \partial_t u_{\mathrm{sp}} - \gamma_{\mathrm{sp}} \end{aligned} \tag{27.163}$$

with the rate of stimulated recombination Equation 27.140. Equation 27.163 is the net heat source caused by recombination and absorption. The first term is the heat generated by nonradiative recombination. The second and third terms describe the heat due to absorption of the stimulated radiation. The fourth term is caused by a possible incomplete energy transfer from the carrier ensemble to the radiation field during stimulated emission, also referred to as quantum defect energy. The last terms describe the heat due to the absorption of the spontaneous radiation, which could be treated similarly as the stimulated radiation but approximations have to be employed because the field generated by spontaneous emission is more challenging to calculate.

27.6.2 Spatial Distributions of the Heat Sources

The profiles of the heat sources of an asymmetric super-large optical-cavity structure similar to that published in [10] are shown in Figure 27.13a through 27.13d for the steady state for an output power of $P_{\mathrm{out}} = 18$ W and an averaged internal power of $\bar{P} = 8$ W. Some of the parameters are given in Table 27.1. The one-dimensional simulation (along y) has been performed with the simulator WIAS-TeSCA [80]. The absorption heat

$$h_{\mathrm{abs}} = (f_{\mathrm{c},n} n + f_{\mathrm{c},p} p)\bar{P}\frac{|\phi|^2}{W} \tag{27.164}$$

with $\phi(y)$ being the vertical normalized mode profile is shown in Figure 27.13. The cross sections for free-carrier absorption are $f_{\mathrm{c},n} = 4 \cdot 10^{-18}$ cm^2 and $f_{\mathrm{c},p} = 12 \cdot 10^{-18}$ cm^2. The main contributions arise in the active layer located at $y \approx 1.9$ μm due to the high nonequilibrium electron and hole densities and in the adjacent p- and n-doped confinement layers where the optical mode resides. The recombination heat

$$h_{\mathrm{rec}} = (5k_{\mathrm{B}}T + E_{\mathrm{g}})\left(R_{\mathrm{non\text{-}rad}} + R_{\mathrm{sp}}\right) \tag{27.165}$$

in Boltzmann approximation is maximal in the active layer as expected. Note that the temperature dependencies of E_c and E_v were not taken into account. The recombination of leaky electrons and holes in the p-confinement layer (cf. Figure 27.10b) generates heat, too. The quantum defect heat

$$h_{\text{defect}} = (5k_B T + E_g - \hbar\omega_0)\frac{g\bar{P}|\phi|^2}{\hbar\omega_0 W} \tag{27.166}$$

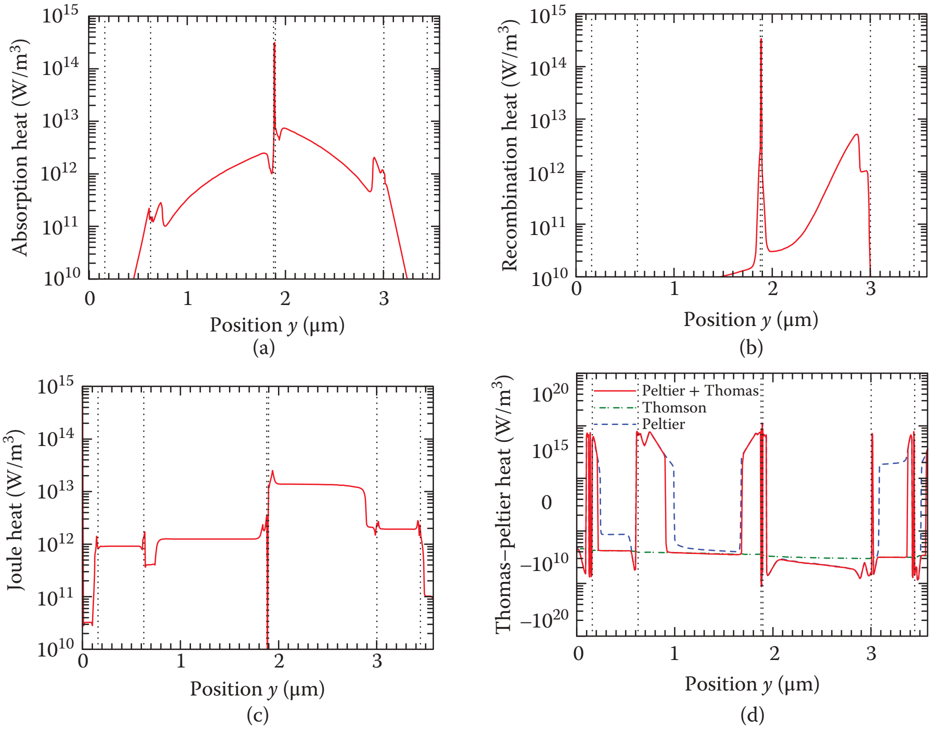
in a Boltzmann approximation, not shown here, is nonvanishing only in the active region and has the same order of magnitude like the recombination heat there ($h_{\text{defect}} \approx 8 \cdot 10^{14}$ W/m^3). The Joule heat (Equation 27.158) shown in Figure 27.13c is mainly generated in the p-confinement layer due to its low doping and the small mobility of the holes. Finally, Thomson heat

$$h_{\text{Thompson}} = \frac{3}{2}\frac{k_B}{e}\frac{dT}{dy}\left(j_{n,y} - j_{p,y}\right) \tag{27.167}$$

and Peltier heat

$$h_{\text{Peltier}} = -\frac{k_B T}{e}\left[\frac{d\ln(n)}{dy} j_{n,y} - \frac{d\ln(p)}{dy} j_{p,y}\right] \tag{27.168}$$

in Boltzmann approximations are shown in Figure 27.13d. The Thompson heat is negative due to the temperature gradient. The Peltier heat is positive or negative corresponding to the signs of the gradients of the carrier densities. Except in the p-cladding layer ($y > 3$ μm) it dominates over the Peltier heat. Note that the Peltier heat results in a cooling of the p-confinement layer. The integrated heat powers are $Q_{\text{Joule}} = 10.2$ W, $Q_{\text{abs}} = 2.7$ W, $Q_{\text{rec}} = 1.3$ W, and $Q_{\text{defect}} = 1.9$ W. Hence Joule heat amounts to more than 50% of the total heat.



(a) (b) (c) (d)

FIGURE 27.13 Vertical profiles of the heat sources of a high-power laser. (a) Absorption heat, (b) recombination heat, (c) Joule heat, and (d) Thomson–Peltier heat. Heteroboundaries are indicated by dotted vertical lines.

27.6.3 Energy Conservation

Inserting Equations 27.149 and 27.148 into the heat sources (Equations 27.155, 27.158 through 27.160) and integrating over the device volume V leads to

$$\begin{aligned}\int_V c_{\mathrm{h}}\partial_t T\, dV - \int_V \boldsymbol{\nabla}\kappa_{\mathrm{L}}\boldsymbol{\nabla}T\, dV = k_{\mathrm{B}}\int_V T(\partial_t n + \partial_t p)\,\mathrm{d}V \\ + \int_V \boldsymbol{j}_n\boldsymbol{\nabla}(TP_n - \varphi_n)\, dV - \int_V \boldsymbol{j}_p\boldsymbol{\nabla}(TP_p + \varphi_p)\,\mathrm{d}V + \int_V eR(TP_n - \varphi_n)\,\mathrm{d}V \\ + \int_V eR(TP_p + \varphi_p)\mathrm{d}V - \partial_t\int_V u_{\mathrm{rad}}\,\mathrm{d}V - \int_V \gamma_{\mathrm{rad}}\,\mathrm{d}V \end{aligned} \tag{27.169}$$

Using a Green's identity we can convert

$$\int_V \boldsymbol{j}_n\boldsymbol{\nabla}(\mathrm{TP}_n - \varphi_n)\,\mathrm{d}V = \int_{\partial V}(\mathrm{TP}_n - \varphi_n)\boldsymbol{j}_n\cdot\boldsymbol{n}dS - \int_V(\mathrm{TP}_n - \varphi_n)\underbrace{\boldsymbol{\nabla j}_n}_{e(R+\partial_t n)}\,\mathrm{d}V \tag{27.170}$$

and

$$-\int_V \boldsymbol{j}_p\boldsymbol{\nabla}(\mathrm{TP}_p + \varphi_p)\,\mathrm{d}V = -\int_{\partial V}(\mathrm{TP}_p + \varphi_p)\boldsymbol{j}_p\cdot\boldsymbol{n}\mathrm{d}S + \int_V(\mathrm{TP}_p + \varphi_p)\underbrace{\boldsymbol{\nabla j}_p}_{-e(R+\partial_t p)}\,\mathrm{d}V \tag{27.171}$$

where ∂V denotes the surface of the device with $\boldsymbol{n}$ being the normal vector. Hence the terms containing the recombination rate cancel, and after rearranging we obtain

$$\begin{aligned}&\int_V c_{\mathrm{h}}\partial_t T\,\mathrm{d}V - \int_{\partial V}\kappa_{\mathrm{L}}\boldsymbol{\nabla}T\cdot\boldsymbol{n}dS \\ &= \int_V \underbrace{(Tk_{\mathrm{B}} - e\mathrm{TP}_n + e\varphi_n)}_{-u_n+e\varphi}\partial_t n\,\mathrm{d}V + \int_V \underbrace{(Tk_{\mathrm{B}} - e\mathrm{T}P_p - e\varphi_p)}_{-u_p-e\varphi}\partial_t p\,\mathrm{d}V \\ &+ \int_{\partial V}(\mathrm{TP}_n - \varphi_n)\boldsymbol{j}_n\cdot\boldsymbol{n}\mathrm{d}S - \int_{\partial V}(\mathrm{T}P_p + \varphi_p)\boldsymbol{j}_p\cdot\boldsymbol{n}\mathrm{d}S \\ &- \partial_t\int_V u_{\mathrm{rad}}\,dV - \int_V \gamma_{\mathrm{rad}}\,\mathrm{d}V\end{aligned} \tag{27.172}$$

From the Poisson equation (Equation 27.141) we can derive the relation

$$e\int_V \varphi(\partial_t n - \partial_t p)\,\mathrm{d}V = \varepsilon_0\int_{\partial V}\varphi\varepsilon_{\mathrm{s}}\partial_t\boldsymbol{\nabla}\varphi\cdot\boldsymbol{n}\mathrm{d}S - \frac{\varepsilon_0}{2}\int_V \varepsilon_{\mathrm{s}}\partial_t|\boldsymbol{\nabla}\varphi|^2\,\mathrm{d}V \tag{27.173}$$

where we applied again a Green's identity. Using this relation and the expression for the heat capacity Equation 27.152 we obtain the energy balance equation

$$\frac{d\int_V u\,\mathrm{d}V}{\mathrm{d}t} = -\int_{\partial V}\boldsymbol{j}_u\cdot\boldsymbol{n}dS - \int_V \gamma_{\mathrm{rad}}\,\mathrm{d}V \tag{27.174}$$

where u is the total energy density (sum of electro-static, internal, and radiative energy)

$$u = \frac{\varepsilon_0\varepsilon_s}{2}|\nabla\varphi|^2 + c_L T + u_n n + u_p p + u_{\mathrm{rad}}, \tag{27.175}$$

and $\boldsymbol{j}_u$ is the energy current density,

$$\boldsymbol{j}_u = -\varepsilon_0\varepsilon_s\varphi\partial_t\nabla\varphi - \kappa_L\nabla T - (TP_n - \varphi_n)\boldsymbol{j}_n + (TP_p + \varphi_p)\boldsymbol{j}_p \tag{27.176}$$

both already introduced in [74]. The first term in Equation 27.176 is related to the displacement current density. The other terms can be reduced to the entropy current density multiplied by T and the particle flux multiplied by the respective quasi-Fermi potentials.

In what follows, we consider the steady state ($d/dt = \partial_t = 0$) and evaluate the surface integral in Equation 27.174. We assume no flow of electrical current through the surface, except at the n-contact located at $y = 0$ and the p-contact at $y = H$:

$$\left.\begin{aligned} \boldsymbol{j}_n \cdot \boldsymbol{n} = 0 \\ \boldsymbol{j}_p \cdot \boldsymbol{n} = 0 \end{aligned}\right\} \quad \text{for} \quad y \neq 0 \quad \text{and} \quad y \neq H \tag{27.177}$$

Between the contacts a forward bias U is applied, so that

$$\varphi_n|_{y=H} = \varphi_p|_{y=H} = U \qquad \text{and} \qquad \varphi_n|_{y=0} = \varphi_p|_{y=0} = 0 \tag{27.178}$$

hold. The normal components of the electron current density at the p-contact $y = H$ and the hole current density at the n-contact $y = 0$ are assumed to vanish,

$$j_{n,y}|_{y=H} = j_{p,y}|_{y=0} = 0 \tag{27.179}$$

Similarly, we assume no heat flow through the surface, except at the surface, located at $y = H$, attached to the heatsink where

$$\kappa_L\partial_y T|_{y=H} = \frac{T_{\mathrm{ref}} - T|_{y=H}}{r_{\mathrm{th}}} \tag{27.180}$$

with r_{th} being the thermal transmission resistance (unit Km^2/W). By inserting the boundary conditions into Equation 27.174 we get for the steady state

$$\iint \frac{T|_{y=H} - T_{\mathrm{ref}}}{r_{\mathrm{th}}}\,\mathrm{d}x\mathrm{d}z = UI - \int_V \gamma_{\mathrm{rad}}\,\mathrm{d}V + \iint [(TP_p)|_{y=H} - (TP_n)|_{y=0}]j\,\mathrm{d}x\mathrm{d}z \tag{27.181}$$

with $j = j_{n,y}|_{y=0} = -j_{p,y}|_{y=H}$ and $I = \iint j\,\mathrm{d}x\mathrm{d}z$. The lhs of Equation 27.181 is the heat flow to the heatsink. The first term on the rhs is the electric input power UI and the second term is the optical power that leaves the cavity. The last term can be considered as Peltier power, which is generated between the electric contacts in the presence of a current flow and results into heating or cooling. Assuming that the spontaneous emission is absorbed in the cavity so that $\gamma_{\mathrm{rad}} = \gamma_{\mathrm{st}}$, using Equation 27.161 and the boundary conditions Equations 27.88 and 27.87, the second term can be shown to be (cf. Equation 27.105)

$$\int_V \gamma_{\mathrm{rad}}\,dV = P_{\mathrm{out}} \tag{27.182}$$

Thus in the steady state the total heat generated in the cavity is given by the electrical input power *UI* minus the optical output power, which coincides with Equation 27.11, minus or plus the Peltier heat.

Acknowledgment

The authors are indebted to K. H. Hasler, M. Lichtner, M. Platz, R. Staske, and H. J. Wünsche for their contributions to the chapter.

References

1. D. F. Welch. A brief history of high-power semiconductor lasers. *IEEE J. Sel. Topics Quantum Electron.*, 6(6):1470–1477, 2000.
2. R. Diehl. *High-power Diode Lasers: Fundamentals, Technology, Applications*, volume 78. Berlin, Heidelberg, New York, NY: Springer, 2000.
3. M. Behringer. High-power diode laser technology and characteristics. In F. Bachmann, P. Loosen, and R. Poprawe, editors, *High Power Diode Lasers*, volume 128 of *Springer Series in Optical Sciences*, pp. 5–74. New York, NY: Springer, 2007.
4. X. Liu, W. Zhao, L. Xiong, and H. Liu. *Packaging of High Power Semiconductor Lasers*. Berlin, Heidelberg, New York, NY: Springer, 2014.
5. I. Fischer, O. Hess, W. Elsäßer, and E. Göebel. Complex spatio-temporal dynamics in the near-field of a broad-area semiconductor laser. *Europhys Lett.*, 35(8):579–584, 1996.
6. H. Wenzel. Basic aspects of high-power semiconductor laser simulation. *IEEE J. Sel. Topics Quantum Electron.*, 19(5):1–13, 2013.
7. P. Crump, O. Brox, F. Bugge, J. Fricke, C. Schultz, M. Spreemann, B. Sumpf, H. Wenzel, and G. Erbert. High-power, high-efficiency monolithic edge-emitting GaAs based lasers with narrow spectral widths. In J. J. Colemann, A. C. Bryce, and C. Jagadish, editors, *Advances in Semiconductor Lasers*, volume 86 of *Semiconductor and Semimetals*, chapter 2. Amsterdam: Elsevier, 2012.
8. L. A. Coldren, S. W. Corzine, and M. L. Mashanovitch. *Diode Lasers and Photonic Integrated Circuits*. Hoboken, NJ: John Wiley & Sons, 2012.
9. H. Wenzel, G. Erbert, A. Knauer, A. Oster, K. Vogel, and G. Tränkle. Influence of current spreading on the transparency current density of quantum-well lasers. *Semicond. Sci. Technol.*, 15(6):557, 2000.
10. K. H. Hasler, H. Wenzel, P. Crump, S. Knigge, A. Maaßdorf, R. Platz, R. Staske, and G. Erbert. Comparative theoretical and experimental studies of two designs of high-power diode lasers. *Semicond. Sci. Technol.*, 29(4):045010, 2014.
11. H. Wenzel, P. Crump, A. Pietrzak, X. Wang, G. Erbert, and G. Tränkle. Theoretical and experimental investigations of the limits to the maximum output power of laser diodes. *New J. Phys.*, 12(8):085007, 2010.
12. J. Piprek. *Semiconductor Optoelectronic Devices: Introduction to Physics and Simulation*. London, Oxford, Boston, New York and San Diego: Academic Press, 2003.
13. W. W. Rigrod. Saturation effects in high-gain lasers. *J. Appl. Phys.*, 36(8):2487–2490, 1965.
14. H.-J. Wünsche, U. Bandelow, and H. Wenzel. Calculation of combined lateral and longitudinal spatial hole burning in $\lambda/4$ shifted DFB lasers. *IEEE J. Quantum Electron.*, 17:1751–1760, 1993.
15. H. Wenzel and G. Erbert. Simulation of single-mode high-power semiconductor lasers. *Proc. SPIE.*, 2693:418–429, 1996.
16. A. Demir, M. Peters, R. Duesterberg, V. Rossin, and E. Zucker. Semiconductor laser power enhancement by control of gain and power profiles. *IEEE Photon Techn. Lett.*, 27(20):2178–2181, 2015.
17. A. Guermache, V. Voiriot, D. Locatelli, F. Legrand, R.-M. Capella, P. Gallion, and J. Jacquet. Experimental demonstration of spatial hole burning reduction leading to 1480-nm pump lasers output power improvement. *IEEE Photon. Techn. Lett.*, 17(10):2023, 2005.

18. Z. Chen, L. Bao, J. Bai, M. Grimshaw, R. Martinsen, M. DeVito, J. Haden, and P. Leisher. Performance limitation and mitigation of longitudinal spatial hole burning in high-power diode lasers. In *Proceedings SPIE*, pp. 82771J–82771J, 2012.
19. X. Wang, P. Crump, H. Wenzel, A. Liero, T. Hoffmann, A. Pietrzak, C. M. Schultz, A. Klehr, A. Ginolas, S. Einfeldt, et al. Root-cause analysis of peak power saturation in pulse-pumped 1100 nm broad area single emitter diode lasers. *IEEE J. Quantum Electron.*, 46(5):658–665, 2010.
20. S. F. Yu, R. G. S. Plumb, L. M. Zhang, M. C. Nowell, and J. E. Carroll. Large-signal dynamic behavior of distributed-feedback lasers including lateral effects. *IEEE J. Quantum Electron.*, 30(8):1740–1750, 1994.
21. H. E. Lassen, H. Olesen, and B. Tromborg. Gain compression and asymmetric gain due to the Bragg grating induced by the standing waves in Fabry-Perot lasers. *IEEE Photon Technol. Lett.*, 1(9):261–263, 1989.
22. E. A. Avrutin and B. S. Ryvkin. Theory of direct and indirect effect of two photon absorption on nonlinear optical losses in high power semiconductor laser. *Semicond. Sci. Technol.*, 32(1): 015004, 2016.
23. E. A. Avrutin and B. S. Ryvkin. Theory and modelling of the power conversion efficiency of large optical cavity laser diodes. In *2015 IEEE High Power Diode Lasers and Systems Conference (HPD)*, pp. 9–10, 2015.
24. H. Wenzel, P. Crump, A. Pietrzak, C. Roder, X. Wang, and G. Erbert. The analysis of factors limiting the maximum output power of broad-area laser diodes. *Opt. Quantum Electron.*, 41(9):645–652, 2009.
25. W. B. Joyce. Current-crowded carrier confinement in double-hetero-structure lasers. *J. Appl. Phys.*, 51(5):2394–2401, 1980.
26. M. Grupen and K. Hess. Simulation of carrier transport and nonlinearities in quantum-well laser diodes. *IEEE J. Quantum Electron.*, 34(1):120–140, 1998.
27. B. Witzigmann, A. Witzig, and W. Fichtner. A multidimensional laser simulator for edge-emitters including quantum carrier capture. *IEEE Trans. Electron Devices.*, 47(10):1926–1934, 2000.
28. L. Borruel, J. Arias, B. Romero, and I. Esquivias. Incorporation of carrier capture and escape processes into a self-consistent cw model for quantum well lasers. *Microelectron J.*, 34(5):675–677, 2003.
29. Y. Liu, W. C. Ng, K. D. Choquette, and K. Hess. Numerical investigation of self-heating effects of oxide-confined vertical-cavity surface-emitting lasers. *IEEE J. Quantum Electron.*, 41(1):15–25, 2005.
30. S. Steiger, R. G. Veprek, and B. Witzigmann. Unified simulation of transport and luminescence in optoelectronic nanostructures. *J. Comp. Electron.*, 7(4):509–520, 2008.
31. A. F. Oskooi, L. Zhang, Y. Avniel, and S. G. Johnson. The failure of perfectly matched layers, and towards their redemption by adiabatic absorbers. *Optics Express.*, 16(15):11376–11392, 2008.
32. T. Tischler. *Die Perfectly-Matched-Layer-Randbedingung in der Finite-Differenzen-Methode im Frequenzbereich: Implementierung und Einsatzbereiche*. Innovationen mit Mikrowellen und Licht. Goettingen: Cuvillier, 2004.
33. C. Z. Ning, R. A. Indik, and J. V. Moloney. Effective Bloch equations for semiconductor lasers and amplifiers. *IEEE J. Quantum Electron.*, 33(9):1543–1550, Sep 1997.
34. U. Bandelow, M. Radziunas, J. Sieber, and M. Wolfrum. Impact of gain dispersion on the spatio-temporal dynamics of multisection lasers. *IEEE J. Quantum Electron.*, 37(2):183–188, 2001.
35. M. Lichtner, M. Radziunas, U. Bandelow, M. Spreemann, and H. Wenzel. Dynamic simulation of high brightness semiconductor lasers. *Proc NUSOD '08*, pp. 65–66, 2008.
36. S. Balsamo, F. Sartori, and I. Montrosset. Dynamic beam propagation method for flared semiconductor power amplifiers. *IEEE J. Quantum Electron.*, 2(2):378–384, Jun 1996.
37. C. Z. Ning, R. A. Indik, J. V. Moloney, W. W. Chow, A. Girndt, S. W. Koch, and R. H. Binder. Incorporating many-body effects into modeling of semiconductor lasers and amplifiers. *Proc. SPIE*, 2994:666–677, 1997.

38. E. Gehrig and O. Hess. Spatio-temporal dynamics of light amplification and amplified spontaneous emission in high-power tapered semiconductor laser amplifiers. *IEEE J. Quantum Electron.*, 37(10):1345–1355, Oct 2001.
39. W. W. Chow and H. Amano. Analysis of lateral-mode behavior in broad-area InGaN quantum-well lasers. *IEEE J. Quantum Electron.*, 37(2):265–273, 2001.
40. M. Kolesik and J. V. Moloney. A spatial digital filter method for broad-band simulation of semiconductor lasers. *IEEE J. Quantum Electron.*, 37(7):936–944, Jul 2001.
41. J. Javaloyes and S. Balle. Quasiequilibrium time-domain susceptibility of semiconductor quantum wells. *Phys. Rev. A*, 81(6):062505, 2010.
42. P. N. Butcher and D. Cotter. *The Elements of Nonlinear Optics*, volume 9. Melbourne, Australia: Cambridge University Press, 1991.
43. G. P. Agrawal. *Nonlinear Fiber Optics*. London, Oxford, Boston, New York and San Diego: Academic Press, 2007.
44. R. W. Boyd. *Nonlinear Optics*. London, Oxford, Boston, New York and San Diego: Academic Press, 2008.
45. R. März. *Integrated Optics: Design and Modeling*. Norwood, MA: Artech House, 1995.
46. R. Scarmozzino, A. Gopinath, R. Pregla, and S. Helfert. Numerical techniques for modeling guided-wave photonic devices. *IEEE J. Sel. Topics Quantum Electron.*, 6(1):150–162, Jan–Feb 2000.
47. T. M. Benson, B. B. Hu, A. Vukovic, and P. Sewell. What is the future for beam propagation methods? *Proc. SPIE.*, 5579:351–358, 2004.
48. A. E. Siegman. Laser beams and resonators: The 1960s. *IEEE J. Sel. Topics Quantum Electron.*, 6(6):1380–1388, Nov–Dec 2000.
49. G. H. B. Thompson. A theory for filamentation in semiconductor lasers including the dependence of dielectric constant on injected carrier density. *Opto-electronics.*, 4:257–310, 1972.
50. D. Mehuys, R. Lang, M. Mittelstein, J. Salzman, and A. Yariv. Self-stabilized nonlinear lateral modes of broad area lasers. *IEEE J. Quantum Electron.*, 23(11):1909–1920, 1987.
51. R. J. Lang, A. G. Larsson, and J. G. Cody. Lateral modes of broad area semiconductor lasers: Theory and experiment. *IEEE J. Quantum Electron.*, 27(3):312–320, 1991.
52. J. R. Marciante and G. P. Agrawal. Nonlinear mechanisms of filamentation in broad-area semiconductor lasers. *IEEE J. Quantum Electron.*, 32(4):590–596, 1996.
53. J. V. Moloney. Semiconductor laser device modeling. In B. Krauskopf and D. Lenstra, editors, *Fundamental Issues of Nonlinear Laser Dynamics*, pp. 149–172. College Park, MD: American Institute of Physics, 2000.
54. E. Gehrig and O. Hess. *Spatio-Temporal Dynamics and Quantum Fluctuations in Semiconductor Lasers*, volume 189. Springer Science & Business Media, 2003.
55. K. Böhringer. *Microscopic Spatio-Temporal Dynamics of Semiconductor Quantum Well Lasers and Amplifiers*. PhD thesis, Institut fuer Technische Physik, Deutsches Zentrum fuer Luft- und Raumfahrt, 2007.
56. J. V. Moloney, M. Kolesik, J. Hader, and S. W. Koch. Modeling high-power semiconductor lasers: From microscopic physics to device applications. *Proc. SPIE.*, 3889:120–127, 2000.
57. H. Wenzel, P. Crump, H. Ekhterei, C. Schultz, J. Pomplun, S. Burger, L. Zschiedrich, F. Schmidt, and G. Erbert. Theoretical and experimental analysis of the lateral modes of high-power broad-area lasers. *Proc NUSOD '11*, pp. 143–144, 2011.
58. P. Crump, S. Böldicke, C. M. Schultz, H. Ekhteraei, H. Wenzel, and G. Erbert. Experimental and theoretical analysis of the dominant lateral waveguiding mechanism in 975 nm high power broad area diode lasers. *Semicond. Sci. Technol.*, 27(4):045001, 2012.
59. N. Stelmakh and M. Flowers. Measurement of spatial modes of broad-area diode lasers with 1-GHz resolution grating spectrometer. *IEEE Photon Technol. Lett.*, 18(15):1618–1620, 2006.

60. P. Crump, M. Ekteraei, C. M. Schultz, G. Erbert, and G. Tränkle. Studies of limitations to lateral brightness in high power diode lasers using spectrally-resolved mode profiles. In *2014 IEEE International Semiconductor Laser Conference (ISLC)*, pp. 23–24, 2014.
61. M. Spreemann, M. Lichtner, M. Radziunas, U. Bandelow, and H. Wenzel. Measurement and simulation of distributed-feedback tapered master-oscillator power amplifiers. *IEEE J. Quantum Electron.*, 45(6):609–616, 2009.
62. C. Fiebig, V. Z. Tronciu, M. Lichtner, K. Paschke, and H. Wenzel. Experimental and numerical study of distributed-Bragg-reector tapered lasers. *Appl. Phys. B.*, 99(1):209–214, 2010.
63. M. Spreemann, H. Wenzel, B. Eppich, M. Lichtner, and G. Erbert. Novel approach to finite-aperture tapered unstable resonator lasers. *IEEE J. Quantum Electron.*, 47(1):117–125, Jan 2011.
64. S. Tronciu, V. Z Schwertfeger, M. Radziunas, A. Klehr, U. Bandelow, and H. Wenzel. Numerical simulation of the amplification of picosecond laser pulses in tapered semiconductor amplifiers and comparison with experimental results. *Optics Communications.*, 285(12):2897–2904, 2012.
65. M. Lichtner, V. Z. Tronciu, and A. G. Vladimirov. Theoretical investigation of striped and non-striped broad area lasers with off-axis feedback. *IEEE J. Quantum Electron.*, 48(3):353–360, 2012.
66. M. Radziunas, R. Herrero, M. Botey, and K. Staliunas. Far-field narrowing in spatially modulated broad-area edge-emitting semiconductor amplifiers. *J. Opt. Soc. Am. B.*, 32(5):993–1000, 2015.
67. M. Radziunas and R. Ciegis. Effective numerical algorithm for simulations of beam stabilization in broad area semiconductor lasers and amplifiers. *Math. Model. Anal.*, 19(5):627–646, 2014.
68. M. Sheik-Bahae and E. W. Van Stryland. Optical nonlinearities in the transparency region of bulk semiconductors. In R. K. Willardson, E. R. Weber, E. Garmire, and A. Kost, editors, *Nonlinear Optics in Semiconductors I*, volume 58 of *Semiconductor and Semimetals*, pp. 257–318. Amsterdam: Elsevier, 1998.
69. M. Dinu. Dispersion of phonon-assisted nonresonant third-order nonlinearities. *IEEE J. Quantum Electron.*, 39(11):1498–1503, 2003.
70. V. I. Bespalov and V. I. Talanov. Filamentary structure of light beams in nonlinear liquids. *ZhETF Pisma Redaktsiiu.*, 3:471, 1966.
71. A. P. Bogatov. Lateral field instability and six-wave mixing in a diode laser with broad active area. *J. Russian Laser Res.*, 15(5):417–453, 1994.
72. S. Blaaberg, P. M. Petersen, and B. Tromborg. Structure, stability, and spectra of lateral modes of a broad-area semiconductor laser. *IEEE J. Quantum Electron.*, 43(11):959–973, 2007.
73. J. Piprek and Z. M. S. Li. On the importance of non-thermal far-field blooming in broad-area high-power laser diodes. *Appl. Phys. Lett.*, 102(22):221110, 2013.
74. U. Bandelow, H. Gajewski, and R. Hünlich. Fabry-Perot lasers: Thermodynamics based modeling. In J. Piprek, editor, *Optoelectronic Devices—Advanced Simulation and Analysis*, pp. 63–85. New York, NY: Springer, 2005.
75. T. Grasser, T.-W. Tang, H. Kosina, and S. Selberherr. A review of hydrodynamic and energy-transport models for semiconductor device simulation. *Proc. IEEE.*, 91(2):251–274, 2003.
76. G. Albinus, H. Gajewski, and R. Hünlich. Thermodynamic design of energy models of semiconductor devices. *Nonlinearity.*, 15(2):367, 2002.
77. G. K. Wachutka. Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling. *IEEE Trans Comput-Aided Design Integr. Circuits Syst.*, 9(11):1141–1149, 1990.
78. J. E. Parrott. Thermodynamic theory of transport processes in semiconductors. *IEEE Trans. Electron Devices.*, 43(5):809–826, 1996.
79. K. Seeger. *Semiconductor Physics*. Berlin and Heidelberg: Springer Science & Business Media, 2013.
80. H. Gajewski, B. Heinemann, H. Langmach, R. Nürnberg, G. Telschow, K. Zacharias, H.-Chr. Kaiser, and U. Bandelow. WIAS-TeSCA Two-Dimensional Semi-Conductor Analysis Package, 2012.

28

High-Brightness Tapered Lasers

Ignacio Esquivias

Antonio Pérez-Serrano

and

José-Manuel G. Tijero

28.1 Introduction 59
28.2 Simulation Model 62
Model Overview • Model Options Regarding Symmetry • Model Calibration and Comparison with Experiments
28.3 Simulation Examples 67
Device Geometries and Simulation Parameters • GG Tapered Laser with Beam Spoilers • GG Tapered Laser without Beam Spoilers • Narrow Index-Guided Tapered Laser • Comparison between Devices
28.4 Summary 78

28.1 Introduction

The brightness of an optical source is commonly defined as the emitted power per unit of emitting area and per unit of the solid angle into which the power is emitted (Walpole 1996). Therefore, a high-brightness source requires not only a high value of the emitted power but also a high “beam quality” in terms of a low product of the beam size and the beam divergence. The product of the beam radius at waist and the beam divergence half angle is called “beam parameter product” and based on it, the most widely used figure of merit for beam quality, the beam propagation ratio M^2, is defined as the ratio of the beam parameter product of the beam of interest to the beam parameter product of a diffraction-limited, perfect Gaussian beam (TEM_{00}) of the same wavelength λ (ISO 2005; Siegman et al. 1998). Therefore, a value $M^2 = 1$ represents an ideal diffraction-limited source, while values higher than unity indicate a degradation of the beam quality.

Semiconductor lasers are optical sources with very well-known advantages over other types of optical sources: small size, high conversion efficiency, and low cost. There are many applications of semiconductor lasers demanding high brightness: material processing, optical pumping of solid state and fiber lasers, medical treatments, optical wireless communications, and in general all applications requiring high power launched into an optical fiber. However, the brightness of a semiconductor laser is usually limited due to two counteracting requirements: a large emitting area is required to produce high power with reduced bulk and surface heating, while reduced dimensions are required to maintain a single spatial mode and thus a high-quality beam. High-power semiconductor lasers are based on broad-area (BA) devices, with a poor beam quality along the lateral axis, while devices with reduced lateral dimensions and good beam quality, such as ridge waveguide (RW) lasers, suffer from a limited maximum output power.

As a consequence, an important research effort has been devoted to improve the brightness of semiconductor lasers during the last years and various new approaches have been proposed, including lasers with a tapered gain region (Walpole 1996; Wenzel et al. 2003; Sumpf et al. 2009), the master-oscillator power amplifier configuration (O'Brien et al. 1993; Spreemann et al. 2009), and the angled grating distributed feedback laser (Lang et al. 1998; Paschke et al. 2003). Tapered lasers, also called flared unstable cavity lasers, are possibly the best choice to achieve high brightness at moderate cost, due to the technological simplicity of their fabrication process.

The schematic of a typical tapered laser is shown in Figure 28.1a. It is similar to the tapered semiconductor optical amplifier (SOA) described in Chapter 22 of this book (Tijero et al. 2017). In brief, it is composed of a straight and narrow index-guided (IG) section, usually an RW structure, and a gain or IG tapered section where the beam is amplified. Ideally, the optical beam of a taper laser is a single lateral mode that diffracts at the exit of the RW section and is amplified in the tapered section while preserving its shape. The main difference with the tapered SOA is that the reflectivities of the facets are modified to provide laser oscillation at a reasonable value of the injection current. The output facet is usually antireflection (AR) coated while the back facet is coated to provide a high reflectance (HR) in order to decrease the threshold current and to maximize the output power. In many cases, the designs include beam spoiler elements in the form of trenches located in the neighborhood of the border between the RW and the tapered section (Figure 28.1a) (Kintzer et al. 1993). The main role of these elements is to filter the backward propagating field out of the RW. Most of the tapered lasers are based on a standard Fabry–Pérot cavity and therefore they present multiple longitudinal mode spectra. Some designs include a distributed Bragg reflector (DBR) at the end of the RW in order to provide narrow and stable emission spectra (Hasler et al. 2008).

In comparison with other semiconductor lasers, the specific design of tapered lasers leads to an optical beam with strong astigmatism. The virtual source for the vertical (y) axis (sometimes referred to as fast axis) is located at the output facet of the laser from where the beam diffracts in air. However, in the lateral (x) axis (sometimes referred to as slow axis), the beam diffracts from the exit of the RW section in the semiconductor medium and therefore, at low power, the separation between the two virtual sources (astigmatism) is approximately given by the taper section length L_{TAP} divided by the effective index n_{eff}. This is illustrated in Figure 28.1b. In the vertical axis, the beam can be considered as diffraction limited no matter the power. However, in the lateral axis the beam often degrades losing its diffraction-limited character when the power increases. As mentioned earlier, this degradation is usually characterized by the beam propagation ratio M^2. Referring to Figure 28.1c, the beam parameter product, bpp, and M^2 in the lateral axis are, respectively, given by

$$\text{bpp} = W_{0x}\theta_x \text{ and } M^2 = \frac{\pi}{\lambda}\text{bpp} \tag{28.1}$$

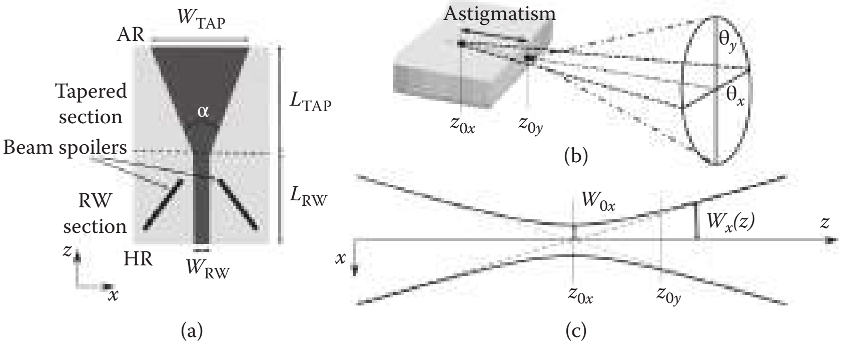


FIGURE 28.1 (a) Schematics of a gain-guided tapered laser. The shaded area is the contact region. Grooves acting as beam spoilers are also depicted. (b) Schematics showing the far field distribution of a tapered laser as a consequence of astigmatism. (c) Evolution of the beam size in the lateral direction along the propagation axis z.

(λ/π being the bpp of a Gaussian beam of wavelength λ). In Equation 28.1, W_{0x} is the beam half width at the virtual source position (beam waist half width), and θ_x is the beam divergence half angle (far-field half width) given by

$$\theta_x = \lim_{z\to\infty} \frac{W_x(z)}{(z - z_{0x})} \tag{28.2}$$

where $W_x(z)$ is the beam half width at the z position and z_{0x} the virtual source location. It is important to point out here that the beam sizes in Equations 28.1 and 28.2 (and therefore the beam divergence) are defined in terms of the second-order moment (variance) of the power density distribution by $W_x(z) = 2\sigma_x(z)$ where

$$\sigma_x^2(z) = \frac{\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} P(x,y,z)(x-\bar{x})^2 \mathrm{d}x\,\mathrm{d}y}{\int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} P(x,y,z)\mathrm{d}x\,\mathrm{d}y} \tag{28.3}$$

and $\bar{x}$ is the first order moment of the power density distribution (x coordinate of the centroid) (ISO 2005). When $W_x(z)$ is defined in this way, and only in this case, the evolution of the beam lateral size along the propagation axis z for any simple astigmatic beam can be described by the hyperbola depicted in Figure 28.1c and given by

$$W_x(z) = W_{0x}\sqrt{1 + \left(\frac{M^2\lambda(z - z_{0x})}{\pi(W_{0x})^2}\right)^2} \tag{28.4}$$

The fitting of the measured $W_x(z)$ to this expression is a method for the experimental determination of M^2, W_{0x}, and z_{0x} (ISO 2005).

Nevertheless, it is a common practice in the research groups having developed tapered lasers in the last years to characterize the beam quality by the parameter $M^2(1/e^2)$ (Krakowski et al. 2002; Sumpf et al. 2002), using the expression

$$M^2(1/e^2) = \frac{\pi}{\lambda}\theta_x(1/e^2)W_{0x}(1/e^2) \tag{28.5}$$

where $\theta_x(1/e^2)$ and $W_{0x}(1/e^2)$ are the half widths of the divergence and virtual source at $1/e^2$, respectively. For Gaussian beams $M^2(1/e^2) = M^2 = 1$ and in general $M^2(1/e^2)$ is a useful parameter to compare different lasers and to estimate the efficiency of the source to couple power into a reduced area device, such as a single mode fiber. However, $M^2(1/e^2)$ for non-Gaussian beams is usually much lower than M^2 and it is not actually a beam propagation ratio in the sense that it is not an invariant of the beam when it propagates in air or across passive, nonaberrating optical elements as required by a real beam propagation ratio. Therefore, the reader should be aware of this when interpreting the real relevance of this parameter. An example of the severe discrepancies between M^2 and $M^2(1/e^2)$ for a non-Gaussian beam will be commented in Section 28.3.5.

Two clearly different types of tapered lasers have been reported to date: gain-guided (GG) and IG lasers. The GG tapered lasers feature a relatively large full taper angle, α_{tap}, electrically defined in the p-contact layer. In these lasers, α_{tap} is designed to match the free diffraction angle (typically 4°–8°, depending on wavelength), as we will describe later. In the IG tapered lasers a small taper angle ($\alpha_{\text{tap}} < 1°$) is defined both, electrically in the p-contact layer and optically by an effective index step created by removing a fraction of the upper epitaxial layers. The beam properties of the two kind of tapered lasers are significantly different (Borruel et al. 2004a). GG tapered lasers with a taper angle close to the free diffraction angle have been fabricated at many different wavelengths from red (Blume et al. 2012) to around 2μm (Pfahler et al. 2006). A review of previous work on these devices can be found in Wenzel et al. (2003) and Sumpf et al. (2009). GG tapered lasers have demonstrated good beam quality and continuous wave (cw) output powers higher than 10 W at 980 nm (Fiebig et al. 2009), 1060 nm (Sumpf et al. 2009), and 1030 nm (Müller et al. 2016). Narrow IG tapered lasers have achieved more than 1 W at 980 nm (Krakowski et al. 2003) and 915 nm (Michel et al. 2005). Narrow IG lasers are usually combined in parallel arrays in high-power laser bars (Auzanneau et al. 2003; Wilson et al. 1999).

It is well known that the interaction between the optical field and the semiconductor gain media promotes a complex spatial-spectral dynamics in semiconductor lasers. In the case of tapered lasers, nonlinear effects, such as spatial hole burning (SHB) and thermal lensing add complexity to the physical phenomena involved and make the high power behavior to significantly deviate from the ideal low-power performance. Efficient and accurate modeling approaches are thus necessary to analyze and predict the beam properties of tapered lasers in order to design new geometries with improved performance. During last 20 years, different approaches have been applied to the modeling and simulation of tapered lasers and amplifiers. For a review see (Tijero et al. 2017) in this book.

Our group at the Universidad Politécnica of Madrid, in collaboration with the University of Nottingham, developed CONAN (Borruel et al. 2002, 2004a; Sujecki et al. 2003), a sophisticated simulator for tapered lasers that solves the electrical, optical, and thermal equations for these devices. Despite the assumptions needed to reduce the model complexity (steady state, single frequency, two-dimensional [2D] propagation of the optical mode), the simulations showed good qualitative and quantitative agreement with experimental results in tapered lasers with different geometries and based on different materials (Sujecki et al. 2003; Borruel et al. 2004a; Odriozola et al. 2009; Esquivias et al. 2010). Furthermore, the simulator demonstrated to be a useful tool to predict the behavior of novel designs prior to their fabrication (Borruel et al. 2005; Michel et al. 2009). Other models in literature (Williams et al. 1999; Mariojouls et al. 2000), based on similar approaches, have also reproduced the main trends observed experimentally.

In this chapter, we present a didactic overview of how the main beam characteristics of high-brightness tapered lasers can be accounted by simulation approaches with specific detail devoted to our simulation tool, and provide illustrative examples representative of some of the most typical behaviors of these devices. After this introduction, Section 28.2 presents a brief description of our simulation model and our procedure to calibrate the model in comparison with experimental results; in Section 28.3, we apply the model to three devices representative of some of the most common geometries and guiding mechanisms and analyze how geometry and guiding determine the beam characteristics. The chapter ends with a summary in Section 28.4.

28.2 Simulation Model

28.2.1 Model Overview

Our quasi-3D (three-dimensional) model (Borruel et al. 2002, 2004a; Sujecki et al. 2003) solves self-consistently the complete steady-state electrical, thermal, and optical equations for the tapered laser, assuming single-frequency operation. The laser simulator includes a 3D electrical solver of the Poisson and continuity equations coupled to a 3D thermal solver of the heat-flow equation with the local heat sources provided by the electrical solution. The optical fields in the tapered laser are solved using a 2D wide-angle finite-difference beam propagation method (WA-FDBPM) making use of the effective index approximation (Hadley 1992). Further details of the model can be found in Chapter 22 of this book, which is devoted to tapered SOAs (Tijero et al. 2017). Here, we briefly describe the model for completeness, in order to emphasize the differences when applying it to tapered lasers instead of tapered amplifiers. In fact, the main difference is that in the simulator for tapered amplifiers the WA-FDBPM is applied for the propagation of the optical solution only in the forward direction, while in the case of tapered lasers, the optical solution has to be propagated in both, the forward and then in the backward direction taking into account the corresponding reflectivities of the front and rear facets as corresponds to an optical resonator (Fox and Li 1961). Table 28.1 summarizes the main physical effects included in the model and how their dependence on temperature, wavelength, or carrier densities has been considered.

The simulator flow is shown in Figure 28.2. The solution procedure is initialized by a one-dimensional (1D) laser simulator (Harold, 3.0) which provides for each bias current the lasing wavelength, the bias voltage, and the average photon density in the laser cavity. The tapered laser is divided into 2D slices perpendicular to the z-axis at positions z_i where $i = 1, \dots, N$. The 1D average photon density is used to define an initial guess optical field at the first slice, at the rear facet of the device. The photon density of this field is used as input for the electrothermal solver to calculate the lateral gain and refractive index perturbation profiles in the first slice. With these inputs, a 2D WA-FDBPM making use of the effective index approximation propagates the optical field through the first slice and provides the electrothermal solver with the photon density profile corresponding to the next slice. This procedure is repeated until arriving to the front facet of the device, $i = N$. After applying the electrothermal solver to the last slice, the output power P_{out} and the excess power P_{exc} are calculated (Tijero et al. 2017) and used as inputs for a 3D thermal solver that is applied to the entire cavity and provides a new temperature profile. The whole process is repeated backward, i.e., propagating the solution from the front to the rear facet. After a number of round trips or iterations, the steady state for all the electrical, thermal, and optical variables is found.

TABLE 28.1 Main Physical Effects Included in the Model

Physical Effect or Parameter	Included	Comments
Temperature dependence of energy gap	YES	Empirical Varshni form
Band-gap renormalization	NOT	
Contribution to current density from thermal gradients	YES	Defined by the electron and hole thermoelectric powers
Carrier capture/escape processes in the quantum well (QW)	YES	Defined by electron and hole capture times
Thermionic emission in heterojunctions	NOT	
Fermi–Dirac statistics in bulk materials	NOT	
Dependence of electron and hole mobilities on dopant concentration	YES	Electric field and temperature dependencies not included
Auger recombination	YES	Temperature-dependent Auger parameters
Shockley–Read–Hall (SRH) nonradiative recombination	YES	Complete SRH formula, dependent on both electron and hole carrier concentrations, and on trap properties (density, energy, and degeneration factor of the trap, considering a temperature-dependent capture cross section)
Wavelength dependence of refractive index	YES	Temperature dependence not included
Free-carrier absorption	YES	Linear with local carrier concentration, defined by electron and hole free-carrier absorption coefficients
Nonconstant linewidth enhancement factor	YES	Calculated in QW region from a carrier-dependent differential refractive index, and carrier and wavelength-dependent differential gain
Gain broadening	YES	Lorentzian function
Coulomb enhancement of the gain	NOT	
Spontaneous emission noise	NOT	
Detailed calculation of local heat sources	YES	Local heat sources: Joule effect, nonradiative recombination, free-carrier absorption
Temperature dependence of thermal conductivities	YES	Included in semiconductor layers

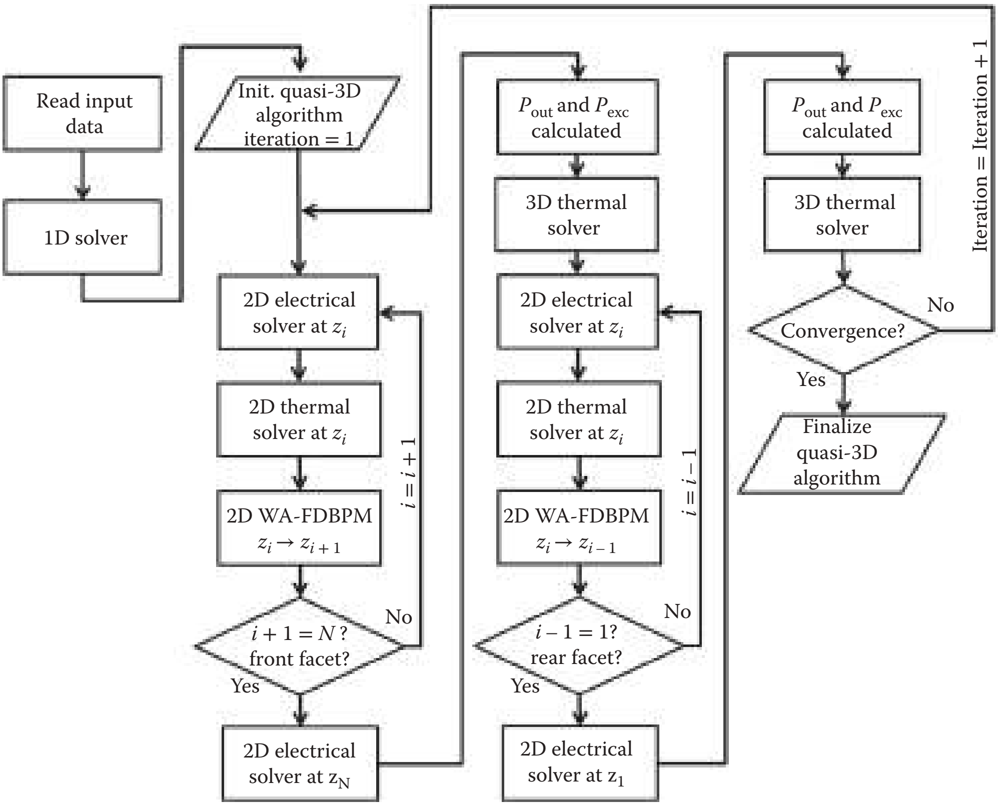

Read input data
1D solver
Init. quasi-3D algorithm iteration = 1
2D electrical solver at z_i
2D thermal solver at z_i
2D WA-FDBPM $z_i \rightarrow z_{i+1}$
$i = i + 1$
$i + 1 = N$? front facet?
No
Yes
2D electrical solver at z_N
P_{out} and P_{exc} calculated
3D thermal solver
2D electrical solver at z_i
2D thermal solver at z_i
2D WA-FDBPM $z_i \rightarrow z_{i-1}$
$i = i - 1$
$i - 1 = 1$? rear facet?
No
Yes
2D electrical solver at z_1
P_{out} and P_{exc} calculated
3D thermal solver
Convergence?
No
Yes
Finalize quasi-3D algorithm
Iteration = Iteration + 1

FIGURE 28.2 Main flow of the quasi-3D simulator for tapered lasers CONAN.

The convergence criterion is based on the stability of the optical field at the rear facet after consecutive iterations, both in shape and power. It can be expressed in terms of a parameter ε as

$$\varepsilon = \frac{\left\langle \left|E_n(x) - E_{n-1}(x)\right|, \left|E_n(x) - E_{n-1}(x)\right| \right\rangle}{\left\langle \left|E_n(x) + E_{n-1}(x)\right|, \left|E_n(x) + E_{n-1}(x)\right| \right\rangle} \tag{28.6}$$

where $E_n(x)$ is the optical field at the rear facet after n iterations and $\langle , \rangle$ denotes the scalar product. The iterative process finishes when ε is lower than a threshold value supplied by the user, typically 5×10^{-5}.

As commented previously, in order to filter the front facet back-reflected optical field reaching the ridge section, some tapered devices include beam spoilers (Walpole 1996). The beam spoilers are modeled by setting the optical field to zero at the beam spoiler locations (Mariojouls et al. 2000).

It should be noticed that our algorithm uses a fixed lasing wavelength provided by the 1D simulator. In a more rigorous approach, the lasing wavelength should be recalculated during the quasi-3D algorithm in order to take into account the spatial effects on it. Nevertheless, the comparison with experimental results has proven the approximate validity of our approach.

In some cases, the tapered lasers are fabricated with separate contacts for each section (Paschke et al. 2005; Odriozola et al. 2009; Michel et al. 2009). The model takes into account this possibility and allows the different sections to have different bias voltages. This is implemented as follows: after the initialization, the 1D simulator provides the laser wavelength, the initial photon density, and a bias voltage V_0. Then, for each section *i*, the applied bias voltage is recalculated as $V_i = V_0 + \Delta V_i$, where ΔV_i can be positive or negative. With these voltage inputs, the simulator proceeds normally and finally, the current at each section is calculated by integration of the current density.

The model can also be applied to the simulation of tapered lasers under patterned injection current (Borruel et al. 2004b). This is an interesting strategy aimed at counteracting the detrimental effects of the SHB. For this, a predefined function that laterally scales the epilayer material conductivities is introduced in the simulator. In this way, the local resistance is modified and therefore the current density profile is also modified according to the conductivity profile.

28.2.2 Model Options Regarding Symmetry

Our simulation procedure is based on launching a trial optical field and solving the forward and backward propagation until convergence (Fox and Li 1961). Depending on the symmetry of the initial trial optical field, the model has three versions:

- Half-cavity (HC) model: This is the basic version of the model. In this version, as in similar models (Williams et al. 1999; Mariojouls et al. 2000), the trial optical field is the fundamental mode of the passive RW section (an even function), although we have checked that the final optical field is independent of the shape of the initial field, provided it is an even function. Since the beam propagation method (BPM) preserves the parity of the field and the device is symmetric, only half of the cavity needs to be explicitly solved, thus reducing the computational effort. However, this version cannot take into consideration the effect of the odd components of the field and fails in reproducing some experimental features of the beam properties of IG tapered lasers, such as the excitation of secondary lateral modes. Therefore, the model was upgraded and two versions including optical fields with both odd and even components were created (Esquivias et al. 2010).
- Full cavity coherent coupling (FCCC) model: In this version, the equations are solved in the full cavity and the initial trial field is an asymmetric field $E_{\mathrm{a}}(x)$, containing odd and even components, i.e., $E_{\mathrm{a}}(x) = E_{\mathrm{e}}(x) + E_{\mathrm{o}}(x)$. The photon density is taken as proportional to $|E_{\mathrm{a}}(x)|^2$ with $E_{\mathrm{a}}(x)$ being the optical field after the propagation by the BPM through the previous slice. This photon density is in general asymmetric, although it can result in a final symmetric solution, depending on the particular device under study. This version often produces a snake-like intensity profile in the plane of the active layer. The corresponding near-field (NF) and far-field (FF) profiles are asymmetric and the intensity profile is not stable after subsequent roundtrips. This kind of behavior has been experimentally observed and attributed to the coherent coupling of frequency-locked lateral modes (Guthre et al. 1994).
- Full cavity incoherent coupling (FCIC) model: As in the FCCC version, the equations are solved in the full cavity and the initial trial field is an asymmetric field $E_{\mathrm{a}}(x)$, but in this case, the photon density is calculated as proportional to $|E_{\mathrm{s}}(x)|^2$, the addition of the field intensities of the even and odd components of $E_{\mathrm{a}}(x)$: $|E_{\mathrm{s}}(x)|^2 = |E_{\mathrm{e}}(x)|^2 + |E_{\mathrm{o}}(x)|^2$. In this way, the photon density is symmetric, and consequently also the carrier and temperature profiles. This approach is equivalent to the simultaneous propagation of an even and an odd field with slightly different frequencies and therefore only coupled through its interaction with the gain medium, i.e., incoherently coupled. This version is more appropriate than the HC model for IG lasers supporting higher order lateral modes besides the fundamental mode.

28.2.3 Model Calibration and Comparison with Experiments

A critical point in all sophisticated laser models is the high number of relatively unknown material parameters. Some of them (energy band parameters, refractive index, etc.) are specific of the different semiconductor materials used in the device. These are usually well known only for the most common binary materials (GaAs, InP), relatively well known for some ternary alloys and relatively unknown for many other alloys. Even for the best known materials, the temperature or wavelength dependence of some parameters has not been reported. In addition, other important parameters (scattering losses, trap characteristics, etc.) depend on the particular fabrication process, and can be different for nominally identical materials. As a consequence, it is always necessary to calibrate the model parameters with experimental results in order to reproduce the experiments and to predict the trends of the device performance when modifying the material composition or the device geometry.

The main idea is to find, if possible, a set of simulation parameters such that the model reproduces the main trends of relevant experimental results, such as power–voltage–current characteristics, FF and NF patterns, and evolution of M^2 and FF and NF patterns with current. If these simulation parameters conveniently account for the main physical effects causing the observed beam properties, then the model will predict qualitatively, and even quantitatively, the performance of new devices based on the same materials but with different design. In other words, it is not so important to use a complete set of correct material parameters, but rather to find out those parameters relevant for the model and related to the main internal process limiting the maximum power or degrading the beam quality.

Our calibration process includes three steps, the first one at theoretical level, the second one by comparing with experimental results in BA lasers, and the third one by comparing with results in tapered lasers. Prior to the calibration, it is important to consider the entire epilayer structure and find out in the literature as many material parameters as possible, especially in the most standard references (Adachi 1992; Vurgaftman et al. 2001). Unfortunately, as stated above, many important parameters, such as refractive index, Auger, and free-carrier absorption coefficients, have never been measured for new materials or alloy compositions, in which case we use judicious guess values based on the parameter values in other alloys with similar gap or composition.

The first step of the calibration procedure is the fitting of the parameters used to calculate the material gain spectra $g_{\text{mat}}(\lambda)$ and the spontaneous recombination rate R_{sp}, as a function of the electron and hole quasi-Fermi energies. As it is described in Tijero et al. (2017), the simulation tool operates with a parabolic band model for the calculation of these functions but takes into account band mixing effects by the following fitting procedure: first, we use a valence band (VB) mixing model (Coldren and Corzine 1995), to calculate the quantum well (QW) energy levels and the maximum gain and the spontaneous recombination rate versus carrier concentration, $g_{\max}(n)$, and $R_{\text{sp}}(n)$. Then, these calculations are fitted by the results provided by the simulation tool using same formulation but considering parabolic valence subbands. The fitting parameters are the QW energy levels, the effective mass of each level, and two scaling parameters (multiplying factors) for $g_{\text{mat}}(\lambda)$ and for R_{sp}. Finally, the tool is fed with the parameters that best fit $g_{\max}(n)$ and $R_{\text{sp}}(n)$ obtained by the VB mixing model. Very good agreement was achieved for a wide range of carrier densities.

The second step in the calibration procedure is to compare 1D simulations with experimental results in BA lasers fabricated with the same epitaxial material than the tapered lasers. It is important to include power–current (P–I) characteristics for lasers with different cavity lengths measured at different temperatures, as well as the FF patterns along the vertical axis. The goal of this step is to determine those simulation parameters related to the material quality (scattering losses and Shockley–Reed–Hall recombination parameters), and also to modify those relevant parameters, which are not well known, such as the Auger coefficient and its temperature dependence. The comparison is made between measured and simulated results in terms of the threshold current density (J_{th}) dependence on the cavity length L, internal quantum efficiency, internal losses, and characteristic temperature T_0. The comparison between measured and calculated FF patterns provides a method to modify the refractive indices, especially that of the QW, which is very important for the calculation of the optical confinement factor in the simulations. The comparison between experimental and simulated current–voltage characteristics provides information on the total resistance of the device, which is relevant for a correct estimation of the Joule heating in the simulations.

The final stage of our calibration procedure includes two parameters: one to account for the carrier-induced refractive index change, and the other one to account for the heat transfer efficiency. For the first one, we focus on its effect on the beam divergence and fit the experimental value of the beam divergence at a fixed output power to the simulated value. This can be achieved by using as fitting parameter the coefficient n_1 that relates the index change and the carrier density assuming a square root dependence (Borruel et al. 2004a). Regarding heat transfer, in order to avoid extending the thermal simulation region, we artificially consider a heat-sink area equal to the device area and use the heat-sink thickness as fitting parameter. We use it to fit the experimental average temperature increase when increasing the current to the simulated temperature increase (on average). The experimental average temperature increase is estimated from the shift of the lasing wavelength when increasing the current, assuming a standard dependence. We have checked that by changing the heat sink thickness, the average value of the temperature changes without important modifications in the lateral and longitudinal temperature profiles.

This procedure was applied to 975 nm (IG and GG with beam-spoilers) and to 735 nm (GG without beam spoilers) tapered lasers in Borruel et al. (2004a). We found a good agreement in the P–I characteristics, shape of the NF and FF patterns, and especially in the evolution of beam properties (M^2, astigmatism, widths of NF at waist and FF patterns) with the injection level. Furthermore, we also found a good agreement between the maximum measured power in the 975 nm GG devices and the maximum power with numerical convergence in the simulations, indicating that the physical mechanism limiting the power was correctly reproduced. This good agreement provided the basis for a new geometrical design, the clarinet laser (Borruel et al. 2005), which showed beam properties similar to those predicted by the simulations. Our model was also applied to simulate the 915-nm IG lasers described in Michel et al. (2005), but in this case the experiments showed a double peak in the NF and FF which was not reproduced by the simulations and gave rise to the upgraded FCIC version of the model previously described. Additional comparisons between experiments and simulations can be found in Odriozola et al. (2009), Michel et al. (2009), and Esquivias et al. (2010).

28.3 Simulation Examples

In order to illustrate the capabilities of the simulation tools for accounting for the behavior of typical tapered laser geometries and guiding mechanisms, we have selected three representative geometries and guiding mechanisms sharing the same epitaxial structure. In this section, we analyze with our simulation tool CONAN the effect of these design parameters on the device performance. We devote specific attention to the comparative analysis of the effects on the beam properties of the device geometry and injection conditions. Since this analysis is presented here mainly for illustrative purposes, some interesting effects will be just overviewed without a detailed study. The half-cavity version of the simulator was used in the three examples. The simulations were performed under isothermal conditions to concentrate the focus in carrier-related effects, since thermal effects are strongly dependent on the value of some relatively unknown material parameters. A detailed discussion on the role of thermal effects in tapered lasers can be found in Esquivias et al. (2010).

28.3.1 Device Geometries and Simulation Parameters

The epitaxial structure of the simulated devices corresponds to that of the 1060 nm GG tapered lasers reported in Ruiz et al. (2009). In brief, it consists of a strained InGaAs QW embedded in a large InGaAsP symmetric optical cavity with AlGaAs cladding regions. We will compare the beam properties of three devices: (1) a GG tapered laser with beam spoilers (GG-BS), (2) a GG tapered laser without beam spoilers (GG-NBS), and (3) a narrow IG laser without beam spoilers (IG).

The geometrical and material parameters used in the simulations are identical for the three devices except for the taper angle and guiding mechanism in the tapered section. The total cavity length is 3 mm and the taper angles are 6° and 1° for the GG and IG devices, respectively. Table 28.2 shows these parameters as well as a brief summary of the most influential material and device parameters used in the

TABLE 28.2 Geometrical Parameters of the Tapered Lasers and Summary of the Most Relevant Material and Device Parameters Used in the Simulation

Symbol	Parameter	Value	Units
L_{RW}	Length of the RW section	1	mm
W_{RW}	Width of the RW section	2.5	μm
L_{TAP}	Length of the taper section	2	mm
R_f	Front facet reflectivity	0.025	
R_b	Back facet reflectivity	0.95	
	Full aperture of beam spoilers	17	μm
	Distance from beam spoilers to back facet	1	mm
α_{tap}	Full taper angle	6 (GG); 1 (IG)	°
Δn_{eff}	Effective index step of the RW section	4.7×10^{-3}	
T_{HS}	Heat sink temperature	20	°C
Γ	Confinement factor	0.0084	
α_{scat}	Scattering losses coefficient	0.5	cm^{-1}
$C_n(C_p)$	Electron (hole) Auger recombination coefficient	$2(2) \times 10^{-30}$	cm^6s^{-1}
k_e, (k_h)	Electron (hole) free-carrier absorption coefficient	$3(7) \times 10^{-18}$	cm^2
n_I	Differential refractive index coefficient	4.5×10^{-11}	$cm^{3/2}$

simulation. These parameters were extracted from standard references or fitted after applying the calibration procedure described in Section 2.3 to the GG-BS device (Esquivias et al. 2010). The taper angle of the GG devices was selected so as to fit the calculated free diffraction angle assuming an index step $\Delta n_{eff} = 4.7 \times 10^{-3}$.

28.3.2 GG Tapered Laser with Beam Spoilers

The GG-BS device presented here is representative of GG tapered lasers with beam spoilers. At low power, these devices show single-lobed NF and FF patterns and low values of M^2. Figure 28.3a and b illustrates the evolution of the profile of the forward and backward optical field intensities along the cavity at low power (slightly above threshold). A more detailed view of the beam profiles at different longitudinal positions is provided in Figure 28.4a and b. The shape of the fundamental lateral mode of the RW section entering the gain section (see curve A in Figure 28.4a) can be approximated by a Gaussian function. In this example, the calculated full width of the mode at $1/e^2$ (W_{mode}) is 4 μm.

When entering the tapered section, the mode is subjected to two different effects: (1) amplification by the gain medium and (2) free diffraction if the full taper angle is larger than the free diffraction full angle θ_D (at $1/e^2$). The free diffraction angle of an ideal Gaussian beam is given by

$$tg\left({}^{\theta_D}\!/\!_2\right) = \frac{2 \cdot \lambda}{\pi \cdot n_{eff} \cdot w_{mode}} \tag{28.7}$$

where n_{eff} is the effective index of the vertical waveguide. In the case of the GG-BS device under analysis, the full taper angle (6°) has been chosen to match the value of θ_D. Figure 28.4a shows the lateral intensity profile of the forward traveling light at cross sections taken at several positions along the cavity (curve A at $z = 1$ mm, curve B at $z = 2$ mm, and curve C at $z = 3$ mm, $z = 0$ and $z = 3$ mm being the back and the output facet, respectively). The beam expands smoothly as it propagates along the tapered section and reaches the output facet keeping its Gaussian-like profile (curve C), although the wave front has a convex shape, and therefore the phase at the facet is far from being uniform. The reflected (or backward) field continues diffracting in its way back but now the freely diffracting beam does not overlap any more with the gain region and therefore the beam becomes narrower as it propagates backward. This evolution is

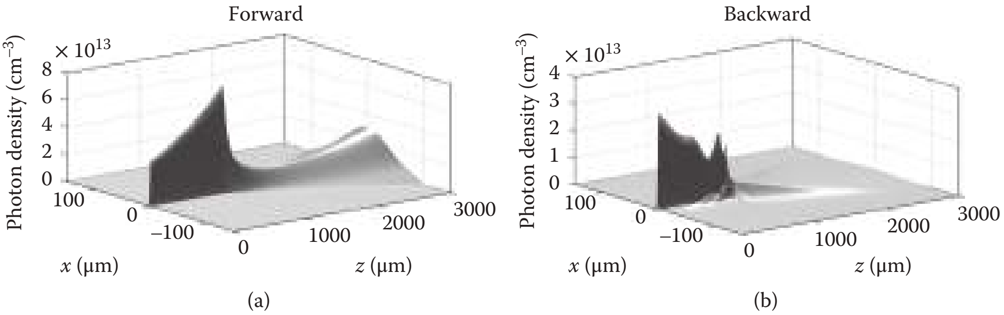


FIGURE 28.3 (a) Forward and (b) backward optical field intensity inside the cavity for the GG-BS tapered laser, when operated at low power ($P_{\text{out}} = 38$ mW).

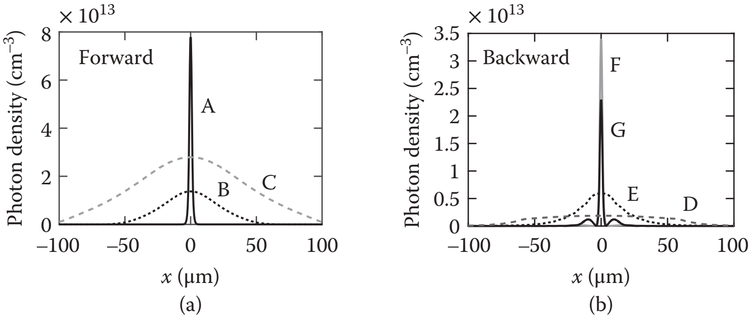


FIGURE 28.4 (a) Forward and (b) backward optical field intensities at different positions inside the cavity for the GG-BS device, when operated at low power ($P_{\text{out}} = 38$ mW). In panel (a): A at $z = 1$ mm, B at $z = 2$ mm, and C at $z = 3$ mm (output facet). In panel (b): D at $z = 2$ mm, E at $z = 1$ mm, F at $z = 0.5$ mm, and G at $z = 0$ mm (back facet).

illustrated in Figure 28.4b, where curves D and E show the intensity profiles at an intermediate position in the tapered region ($z = 2$ mm) and at the entrance of the RW section, respectively. As the beam entering the RW section is wider than the fundamental mode of this section (70 versus 4 μm at $1/e^2$), a substantial part of the power is not coupled producing the so-called coupling losses (Walpole 1996) or taper losses. The beam propagating along the RW section is filtered by the single-mode waveguide, with the help of the beam spoilers. Curves F and G are the profiles at the middle of the RW section and at the back facet of the device, respectively. The side lobes of curve F reveal that the filtering effect is still imperfect at the middle of the RW section (see also Figure 28.5b). However, at the back facet, the lobes have virtually disappeared and the beam is as narrow as the fundamental mode of the waveguide revealing that in this case the filtering effect is fully accomplished at the back facet. In other cases, the filtering could be still accomplished in the subsequent forward propagation along the RW and the second pass through the beam spoilers after the reflection at the back facet.

Even more insight into the evolution of the beam inside the cavity can be gained by the gray-scale plots in Figure 28.5. In these plots, the forward (Figure 28.5a) and backward (Figure 28.5b) photon densities in each slice perpendicular to the longitudinal axis have been normalized to their maximum value in the slice and white lines have been drawn at the border of the injected region and at the position of the beam spoilers. In comparison with the smooth and homogeneous expansion of the photon density profile of the forward field in the tapered section, the width of the backward field photon density profile increases from $z = 3$ mm to about $z = 2.5$ mm and decreases afterward due to the gain guiding in the narrower part of the tapered region. Nevertheless, at the entrance of the RW section, the photon density profile of the

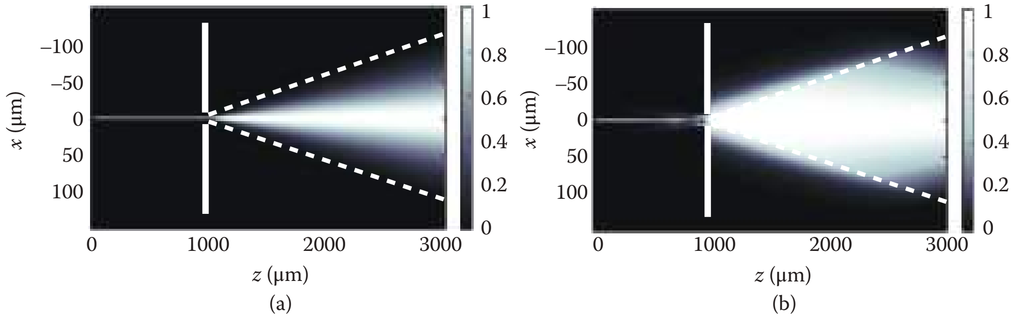


FIGURE 28.5 Normalized (a) forward and (b) backward optical field intensity inside the cavity at each z slice for the GG-BS device, when operated at low power (P_{out} = 38 mW). The white dashed lines indicate the pumped tapered region, while the white solid lines indicate the position of the beam spoilers.

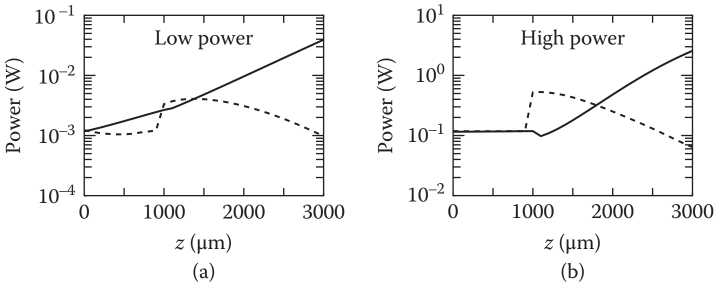


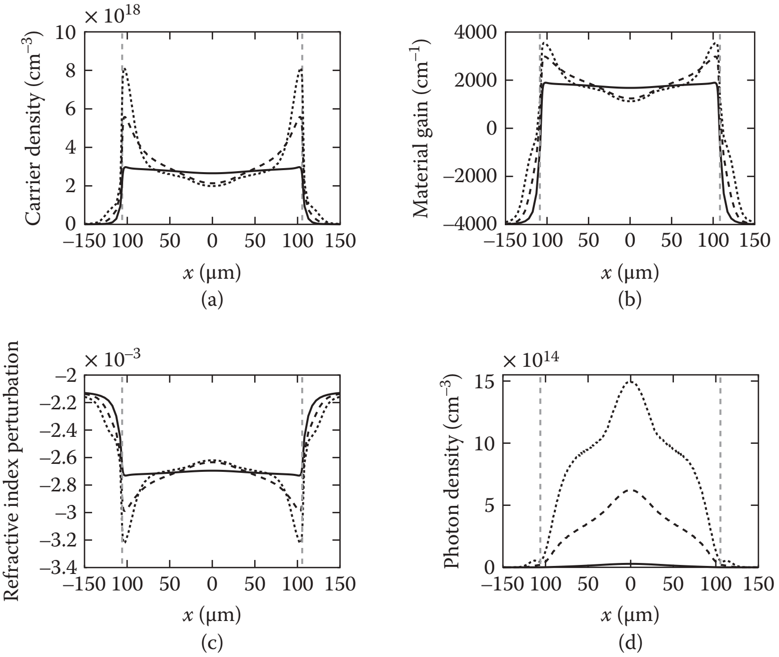
FIGURE 28.6 Optical power of the forward (solid) and backward (dashed) traveling optical fields along the cavity for the GG-BS laser, when operated at (a) low power (P_{out} = 38 mW) and (b) high power (P_{out} = 2.5 W).

backward field still expands far beyond the limits of the injected area. At this point, the filtering role of the beam spoilers is crucial and only tiny diffraction lobes survive after entering the RW section.

To complete the picture, the evolution along the cavity of the total powers carried by the forward and backward beams is shown in Figure 28.6. These powers were obtained by integration of the forward and backward photon densities in the lateral direction at each longitudinal position. In the logarithmic scale of Figure 28.6, an exponential growth of the power is represented by a straight line in which slope is proportional to the effective modal gain, defined as the difference between the modal gain and all the losses occurring in the beam propagation. At low power (Figure 28.6a), the forward field effective gain is constant in both sections. In this case, it is slightly lower in the RW section due to a lower material gain and a slightly worse overlapping with the optical mode in the RW section (not shown). The backward beam power shows initially an exponential growth up to about z = 2.5 mm due to the good overlapping with the gain region. The progressively worse overlapping makes the beam amplification to decreases down to negative values at the entrance of the RW section. At this point, the power drops down due to the filtering effect of the beam spoilers. The subsequent evolution of the backward propagating power is the result of the competing mechanisms of filtering and gain in the RW section, the balance being slightly positive at the back facet. The evolution of the forward and backward propagating powers at high power is illustrated in Figure 28.6b and is mentioned later in this section.

The nearly ideal behavior of the GG-BS device observed at low power changes dramatically when increasing the output power. The main reason for this is the mutual interaction in the semiconductor material between photons and carriers in a feedback loop leading eventually to power saturation and self-focusing at high output powers. Let us start with the optical mode profile. At high injection the mode narrows, thus concentrating a high photon density in the cavity axis. This high photon density depletes the carrier density along the cavity axis due to the higher stimulated recombination in this region, the so-called SHB effect. Figure 28.7a shows the simulated carrier density for the GG-BS device at the output facet for increasing output powers. The initially flat profile of the carrier density evolves when the injection increases to a "batman-ears" like profile, with maxima at the sides and minimum at the center. This profile has been experimentally observed through spontaneous emission measurements (Pagano et al. 2011). The carrier density minimum is limited by the transparency carrier density, 1.2×10^{18} cm^{-3} in our example. The initial simplified approach by Walpole (1996) suggested that the local gain saturation in the cavity axis caused by the SHB would induce an increase of the photon density in the side regions, leading to a top hat shape. But in semiconductor materials, a change of the carrier density produces simultaneously changes in the gain and in the refractive index, which are related by the linewidth enhancement factor. Figure 28.7b and c shows the corresponding gain and index profiles at the power levels of Figure 28.7a. The gain decreases at the cavity axis and the refractive index increases. The shape of the index profile produces a parasitic waveguide for the beam, with more important consequences on the beam shape than the gain profile. The carrier-induced waveguide produces a convergent lens effect during the beam propagation along the tapered region, which concentrates the power density at the center of the beam, thus closing the feedback loop (see Figure 28.4d).

At high power density, the strong feedback induces self-focusing of the beam, leading to saturation of the output power. This effect can be visualized with the help of Figure 28.8, where (as in Figure 28.5) we have plotted the forward (Figure 28.8a) and backward (Figure 28.8b) normalized photon densities at a high power level. In comparison with Figure 28.5a, the forward beam is much narrower, showing clearly the self-focusing. More difficult is the interpretation of the backward beam shape in the tapered section (Figure 28.8b). The expansion of the backward beam beyond the limits of the injected region adds to the gain guiding and the induced index guiding a new degree of complexity making the beam shape extremely difficult to interpret in simple terms. Again, the filtering role of the beam spoilers is apparent in the RW



(a) (b) (c) (d)

FIGURE 28.7 (a) Carrier density, (b) material gain, (c) refractive index perturbation, and (d) photon density profiles at the output facet for the GG-BS laser at different output powers: $P_{\text{out}} = 38$ mW (solid), $P_{\text{out}} = 955$ mW (dashed), and $P_{\text{out}} = 2.5$ W (dotted). Vertical dashed gray lines indicate the pumped region limits.

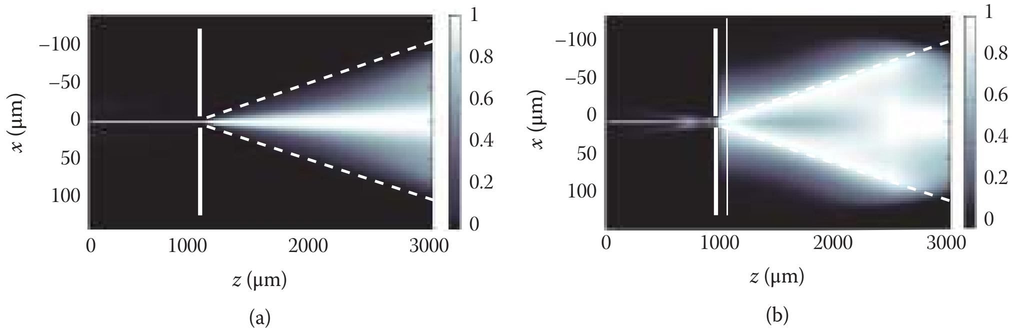


FIGURE 28.8 Normalized (a) forward and (b) backward optical field intensity inside the cavity at each z slice for the GG-BS device, when operated at high power ($P_{\text{out}} = 2.5$ W). The white dashed lines indicate the pumped tapered region, while the white solid lines indicate the position of the beam spoilers.

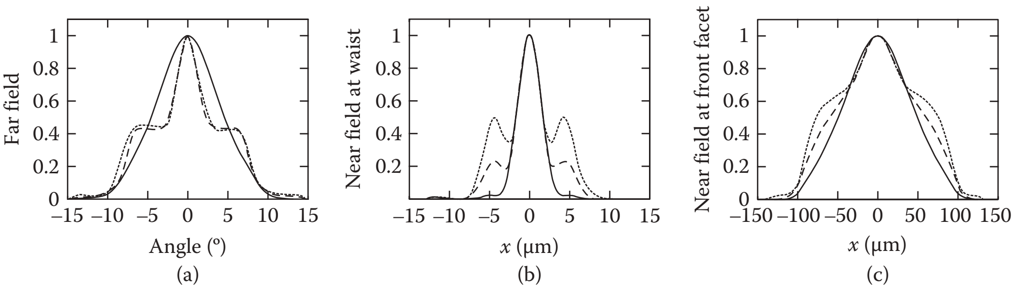


FIGURE 28.9 (a) Far field, (b) near field at waist, and (c) near field at the front facet for the GG-BS device at three power levels, $P_{\text{out}} = 38$ mW (solid), $P_{\text{out}} = 955$ mW (dashed), and $P_{\text{out}} = 2.5$ W (dotted).

section. The small diffraction lobes surviving at low power are now more noticeable and can even reach the back facet from where they are reflected as can be (hardly) seen in Figure 28.8a.

The evolution of the integrated forward propagating power at high injection (Figure 28.6b) shows also clear differences with respect to the low injection behavior. The nearly homogeneous effective gain for the forward field at low power (Figure 28.6a) becomes clearly different in the RW and tapered sections at high power. In the RW section, the effective gain vanishes revealing a strong gain saturation confirmed by a carrier density in this section slightly higher than the transparency value (not shown). As the total round-trip gain should be constant, the absence of gain in the RW section is compensated with the high values of the effective gain at the beginning of the tapered section. Further in the tapered section, the forward effective gain decreases slowly due once more to gain saturation. In contrast, the evolution of the backward propagating power at high injection is not significantly different from the low-injection behavior. In the RW section, after the sudden drop of the power due to the beam spoilers, the modal gain remains constant at a value close to zero as for the forward beam.

The SHB and self-focusing of the beam at high injection not only limit the maximum power, but also degrade the beam quality. Figure 28.9 shows the FF pattern and the NF patterns at waist and at the facet, at three power levels. Under simplifying ideal assumptions, the expected value of the FF width can be estimated by applying Snell's law to the beam at the output facet. In this case, this yields a full beam divergence angle θ_{out} (at $1/e^2$) $\sim n_{\text{eff}} \cdot \theta_{\text{D}} = 20°$ ($n_{\text{eff}} = 3.34$). The simulated value at low power is $\theta_{\text{out}} = 16.5°$, not far from the previous estimation. It is clear how the shape of the beam is modified by the carrier-induced convergent lens when increasing the power: the NF patterns at waist and at the front facet develop shoulders and the FF patterns evolve into a narrower central lobe together with the apparition of side lobes. These effects produce also an increase of the astigmatism and the value of M^2 is shown in Section 3.5.

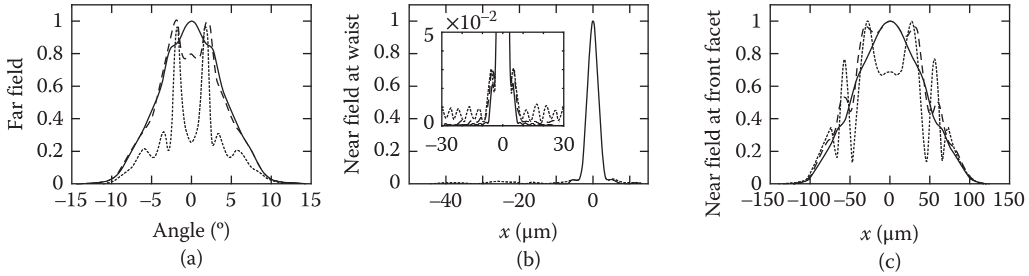
28.3.3 GG Tapered Laser without Beam Spoilers

The GG-NBS device presented here is representative of GG tapered lasers without beam spoilers. In spite of the advantages of the use of beam spoilers, they have also drawbacks due to the additional processing steps and also to the possible introduction of defects close to the active region. In fact, very high power levels have been reported for tapered lasers without beam spoilers (Sumpf et al. 2009; Müller et al. 2016). In this case, the filtering properties of the RW section had been improved by increasing its length, after a careful balance to optimize this design parameter (Wenzel et al. 2003). The main characteristic of the beam experimentally observed in tapered lasers without beam spoilers is a multilobed FF pattern, with an increasing number of lobes together with an increase of the peak to valley ratio when increasing the output power (Borruel et al. 2004a; Fiebig et al. 2009; Sumpf et al. 2009; Müller et al. 2016).

Figure 28.10 shows the simulated evolution of the FF and NF profiles of the GG-NBS device as a function of the output power. The relatively smooth profiles of the FF and NF at low power evolve when the injection increases to a more structured profile with an increasing number of more and more distinguishable lobes. The FF width at $1/e^2$ decreases slightly when the power increases, while the $1/e^2$ NF width at waist remains almost invariant, in agreement with what has been experimentally observed (Fiebig et al. 2009). These behaviors yield an almost invariant low value of $M^2(1/e^2)$. However, the value of M^2 (second moment) significantly increases with the power, mainly due to the tiny side lobes below the $1/e^2$ level in the NF at waist (see inset in Figure 28.10b). These differences will be commented later in Section 3.5. The agreement between the appearance and the evolution of the FF and NF lobes in simulation and experiments provides support for the use of the simulation tool to provide a physical understanding of the origin of these lobes in devices without beam spoilers.

Aiming at this, for an output power level of 0.26 W, Figure 28.11 shows gray-scale plots of the forward (Figure 28.11a) and backward (Figure 28.11b) propagating photon densities, normalized to their maximum value at each cavity position. The side lobes that are already apparent at the beginning of the tapered section, propagate toward the front facet, and finally result in the multilobed NF pattern shown in Figure 28.10c. During the backward propagation (Figure 28.11b), the highest intensity side lobes run away the tapered region where they extinguish without reaching the RW section. However, the backward field entering the RW section is not perfectly filtered and a fraction of it is diffracted by the RW section aperture producing side lobes in the field that reaches the back facet. This residual field at the sides of the RW section is reflected by the back facet and interferes in the tapered section with the forward field arising directly from the RW section, thus producing the multiple peaks observed in the beam characteristics.

The lack of a complete filtering in the RW section is more important at high power, as the intensity of the backward field increases and optically pumps the sides regions around the RW, thus reducing the absorption. This effect is further illustrated in Figure 28.12, where we have plotted the forward and backward photon densities (Figure 28.12a) and the carrier density (Figure 28.12b) at the interface between the



(a) (b) (c)

FIGURE 28.10 (a) Far field, (b) near field at waist, and (c) near field at the front facet for the GG-NBS device at three power levels, $P_{out} = 40$ mW (solid), $P_{out} = 260$ mW (dashed), and $P_{out} = 393$ mW (dotted). The inset in (b) is a magnification emphasizing the side lobes far from the axis appearing at $P_{out} = 260$ and 393 mW.

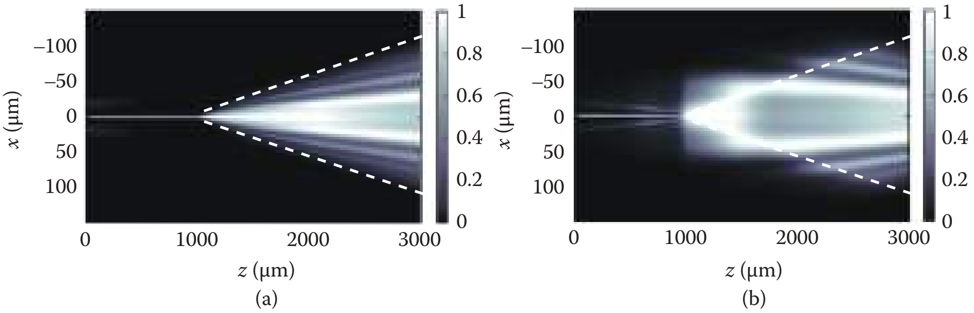


FIGURE 28.11 Normalized (a) forward and (b) backward optical field intensity inside the cavity at each z slice for the GG-NBS device, when operated at high power (P_{out} = 260 mW). The white dashed lines indicate the pumped tapered region.

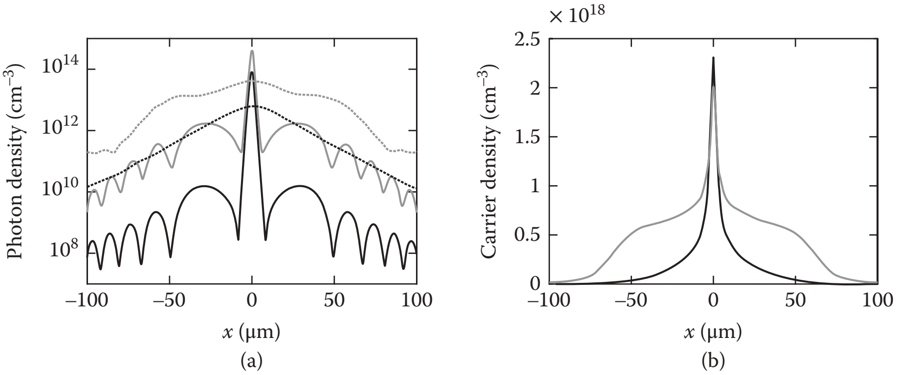


FIGURE 28.12 (a) Forward (solid) and backward (dotted) photon densities and (b) carrier density profiles at the interface between the RW (2.5μm wide) and tapered sections for the GG-NBS device at two output powers P_{out} = 40 mW (black), P_{out} = 260 mW (gray).

RW and tapered sections at two power levels. At low power, the side lobes of the forward field are around 40 dB below the main lobe in the RW (Figure 28.12a) and therefore, this results in a single-lobed NF at the facet (see Figure 28.10c). The backward field at the sides of the RW is low and therefore the induced carrier density (Figure 28.12b) is far below transparency; in consequence, this region is highly absorbent. However, at medium and high injection levels, the backward field intensity is high enough for pumping the sides of the RW to a carrier concentration close to transparency (Figure 28.12b), thus decreasing their absorption. As a result, the forward field intensity at the side lobes becomes only around 20 dB lower than the maximum (Figure 28.12a) giving rise to the multilobed profile obtained at medium and high power level.

Our previous simulations of GG tapered lasers without beam spoilers have shown the same trends as the experimental results (Borruel et al. 2004a; Odriozola et al. 2009), but the quantitative agreement is not as good as we have found in tapered laser with beam spoilers. Furthermore, the maximum output power with numerical convergence in the simulations is usually lower than the measured maximum power. We attribute these discrepancies to the limitations of our steady-state single-frequency model. We think that in the real device, there is a complex dynamics of the different lateral modes giving rise to rapidly varying NF and FF patterns. As the measured NF and FF patterns are temporal averages, it is expected that the narrow and pronounced lobes would average resulting in smoother profiles. In fact, the lack of convergence in the simulations is due to different shapes and positions of the lobes after subsequent round-trips, yielding a stable output power but a different field profile after each iteration. This is illustrated in Figure 28.13, where we have plotted the NF at the facet after two consecutive round-trips for simulation conditions in which a stable solution is not found. The averaging of the field intensity after different roundtrips would produce less pronounced lobes as they appear in the experimental results.

28.3.4 Narrow Index-Guided Tapered Laser

The IG device analyzed in this section is representative of narrow IG tapered lasers. Significant differences with respect to the behavior of large angle GG devices arise from the fact that the single lateral optical mode launched by the RW section does not just expand by free diffraction into an injected tapered region designed to match the free diffraction angle. Instead, the beam is guided in a narrow tapered section defined by the refractive index step, where injection takes place.

The gray-scale plots of Figure 28.14 are illustrative of the forward and backward beam propagation in the IG device at low power (notice the different lateral dimension with respect to the corresponding plots for GG devices). The forward beam expands preserving its shape, with most of the power (99.8% at $z = 2$ mm) inside the guiding region (Figure 28.14a). The propagation of the backward field is determined by a combination of competitive phenomena: diffraction, gain and index guiding, and reflections at the waveguide interface, resulting in a multilobed profile (Figure 28.14b). The index guiding, as well as the relatively small taper angle, produces a beam entering the RW which is narrower than that of the GG devices, hence reducing the taper losses. The RW section acts again as a spatial filter, and the beam recovers its original single-mode shape after a complete round trip.

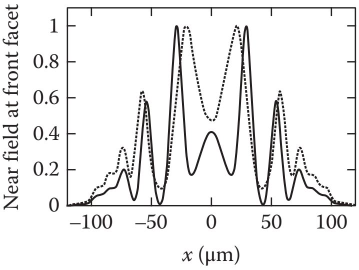
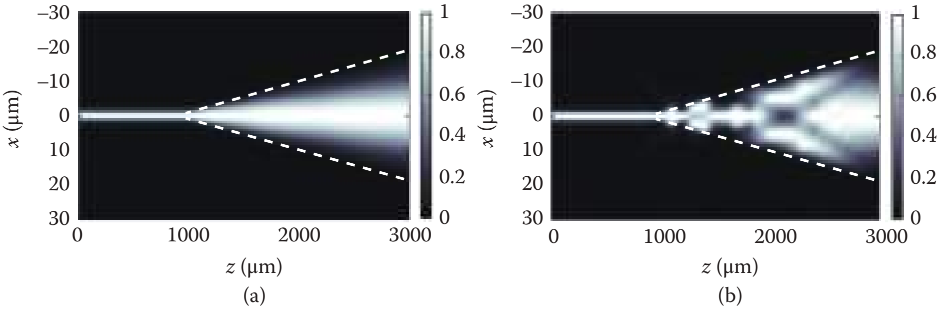


FIGURE 28.13 Example of the simulated NF profiles at the facet for two consecutive iterations of a simulation of the GG-NBS device under conditions in which a stable solution is not found.



(a)

(b)

FIGURE 28.14 Normalized (a) forward and (b) backward optical field intensity inside the cavity at each z slice for the IG device, when operated at very low power ($P_{\text{out}} = 40$ mW). The white dashed lines indicate the pumped tapered region and the index step.

The main effect degrading the beam quality of IG tapered lasers is carrier-induced lensing and self-focusing, similar to the effect analyzed in Section 28.3.2 for GG devices, but with important differences. Figure 28.15 shows the simulated evolution of the FF and NF profiles of the IG device as a function of the output power. Both the profiles as well as the evolution with power are completely different than those of the GG devices shown in Figures 28.9 and 28.10. At very low power (40 mW), since the beam expands in the tapered section covering almost all the guiding region, the simulated FF width at $1/e^2$ is $\theta_{out} = 3.28°$, close to 3.34°, the value resulting from applying Snell's law to the taper angle. In consequence, the beam is clearly astigmatic (204 μm). But the picture changes dramatically when increasing the power: the carrier lensing reduces the width of the beam at the output facet down to sizes for which diffraction effects become relevant, and consequently, the angular width of the FF patterns increases with the output power while the NF at waist and the NF at the facet narrow. In addition, the carrier lensing also produces an almost collimated beam inside the cavity (see Figure 28.16). Therefore, the virtual source position shifts toward the output facet and even beyond, giving rise to a fast decrease of the astigmatism to zero or even to negative values. The evolution of the beam inside the tapered region, as a consequence of the carrier-induced graded index profile (Figure 28.16), resembles that of a graded-index lens or an optical fiber. The strong SHB at high power produces a self-focusing of the beam leading to saturation of the output power, as will be shown in the next section.

28.3.5 Comparison between Devices

In this section, the main performance parameters of the three simulated devices are comparatively analyzed. Figure 28.17a shows the P–I characteristics, and Table 28.3 summarizes the main parameters

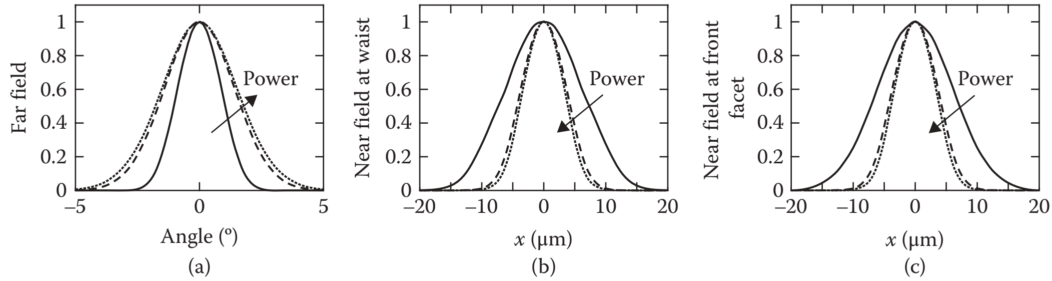


FIGURE 28.15 (a) Far field, (b) near field at waist, and (c) near field at the front facet for the IG device at three power levels, $P_{out} = 40$ mW (solid), $P_{out} = 623$ mW (dashed), and $P_{out} = 912$ mW (dotted).

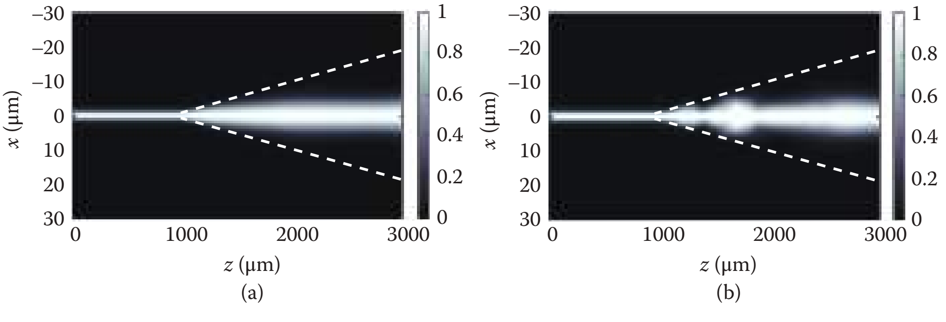


FIGURE 28.16 Normalized (a) forward and (b) backward optical field intensity inside the cavity at each z slice for the IG device, when operated at high power ($P_{out} = 912$ mW). The white dashed lines indicate the pumped tapered region and the index step.

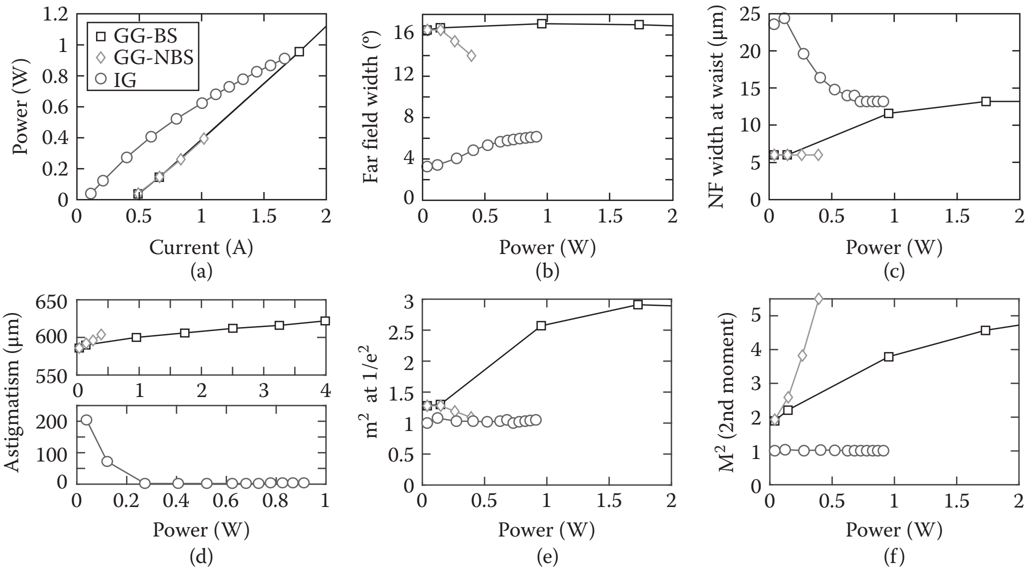


FIGURE 28.17 Comparison of the figures of merit for the different devices: GG-BS (squares), GG-NBS (diamonds), and IG (circles). (a) Power versus current characteristics, (b) far-field width versus power, taken at FWHM (open) and $1/e^2$ (closed), (c) near field at waist width versus power, taken at FWHM (open) and $1/e^2$ (closed), (d) astigmatism versus power, (e) M^2 $(1/e^2)$ versus power, and (f) M^2 (FWHM) versus power.

TABLE 28.3 Parameters Extracted from the Simulated P–I Characteristics of the Three Devices

	GG-BS	GG-NBS	IG
Threshold current (A)	0.43	0.43	0.066
Threshold current density (A/cm^2)	198	197	156
Slope efficiency at threshold (W/A)	0.625	0.625	0.85
Taper losses (cm^{-1})	8.6	8.6	5.4

extracted from these results: the threshold current and threshold current density, the slope efficiency, and the taper losses. The threshold currents of the GG-BS and GG-NBS devices are similar and much higher than that of the IG device, as expected from the comparison of the device area; however, the threshold current density of the IG device is lower, due to the lower taper losses. The taper losses at threshold can be estimated from the simulation results by considering the threshold current density and the modal gain versus current density characteristic, and taking into account the mirror and the internal losses. Values of 8.6 and 5.4 cm^{-1} are obtained for the GG and IG devices, respectively, the lower value for the IG device due to the better matching of the backward field. This difference in the taper losses is also the reason for the higher slope efficiency of the IG device. However, the strong narrowing of the beam in this device due to carrier lensing makes the slope efficiency to decrease when the injection increases (see Figure 28.17a).

Figure 28.17b and c shows the evolution of the widths of the FF and of the NF at waist (at $1/e^2$), respectively, for the three simulated devices. At low power, the FF widths of the GG and IG devices are 16.5° and 3.3°, respectively, not far from the expected value for a Gaussian beam applying Snell's law, as previously discussed. On the contrary, the NF at waist is much wider in the case of the IG device, as expected from its lower divergence. The astigmatism at low power (Figure 28.17d) is also quite different, with the virtual source (or beam waist) located at a distance behind the front facet $z_{vs} = 590$ μm and $z_{vs} = 204$ μm for the GG and IG devices, respectively. The former value is close to $L_{tap}/n_{eff} = 599$ μm, the expected value for the free diffraction of a beam at the output of the RW; the latter is lower, even at very low power, due to the index-guiding mechanism.

As previously mentioned, the devices show clear differences in the evolution of their beam characteristics when power increases. In summary: in the GG-BS device, the FF width at $1/e^2$ is approximately constant and the waist increases, in the GG-NBS the waist is constant and the FF width decreases, and in the IG device the FF broadens and the waist remains constant. These differences are also reflected in the behavior of the astigmatism (Figure 28.17d), which increases with the power for the GG devices while decreases for the IG laser. Finally, the evolution of $M^2(1/e^2)$ and M^2 (second moment) for the three devices is compared in Figure 28.17e and f, respectively. The best beam quality is obtained for the IG device with a M^2 (second moment) close to the unity, indicating a diffraction-limited beam. When power increases, the increase of the value of M^2 (second moment) is much faster for the GG-NBS device than for the GG-BS laser due to the fact that the multiple lobes in both the FF and the NF at waist of the first one strongly affect the second moment widths. On the contrary, the value of $M^2(1/e^2)$ is very low, close to the unity, for the GG-NBS device, even at the highest power in the simulations. Similar discrepancies between M^2 (second moment) and $M^2(1/e^2)$ have been reported experimentally for GG devices without beam spoilers (Fiebig et al. 2009).

28.4 Summary

We have presented an overview of the current state of the art in the modeling of high-power tapered lasers, with a detailed description of our steady-state single frequency quasi-3D simulation tool. We have explained the calibration procedure required for making meaningful comparisons with experiments and using the simulator as a predictive tool. The capabilities of the model have been illustrated by comparing the beam properties of three different types of tapered lasers emitting at 1060 nm: GG with beam spoilers, GG without beam spoilers, and narrow IG. The simulations reproduce the different behaviors experimentally observed in the three types of device and can be used for a better understanding of the interaction between carriers and photons that determines the operation of tapered lasers.

Acknowledgments

This work was supported by Ministerio de Economía y Competitividad of Spain under projects RANGER (TEC2012-38864-C03-02) and COMBINA (TEC2015-65212-C3-2-P); and by the Comunidad de Madrid under program SINFOTON-CM (S2013/MIT-2790). A. Pérez-Serrano acknowledges support from Ayudas a la Formación Posdoctoral 2013 program (FPDI-2013-15740). The authors acknowledge the contribution of L. Borruel, H. Odriozola, S. Sujecki, and E. C. Larkins to the development of the model and code for the simulation of tapered lasers.

References

Adachi, S. 1992. *Physical Properties of III-V Semiconductor Compounds: InP, InAs, GaAs, GaP, InGaAs, and InGaAsP*, New York: Wiley-Interscience.

Auzanneau, S., Krakowski, M., Berlie, F., et al. 2003. High-power and high-brightness laser diode structures at 980 nm using an Al-free active region. *Proc SPIE.* 4995:184–195.

Blume, G., Kaspari, C., Feise, D., et al. 2012. Tapered diode lasers and laser modules near 635 nm with efficient fiber coupling for flying-spot display applications. *Opt Rev.* 19(6):395–399.

Borruel, L., Esquivias, I., Moreno, P., et al. 2005. Clarinet laser: Semiconductor laser design for high-brightness applications. *Appl Phys Lett.* 87:101104.

Borruel, L., Sujecki, S., Esquivias, I., et al. 2002. A self-consistent electrical, thermal and optical model of high brightness tapered lasers. *Proc SPIE.* 4646:355–366.

Borruel, L., Sujecki, S., Moreno, P., et al. 2004a. Quasi-3-D simulation of high-brightness tapered lasers. *IEEE J Quantum Electron.* 40:463–472.

Borruel, L., Sujecki, S., Moreno, P., et al. 2004b. Modeling of patterned contacts in tapered lasers. *IEEE J Quantum Electron.* 40:1384–1388.

Coldren, L. A. and Corzine, S. W. 1995. *Diode Lasers and Photonic Integrated Circuits*. New York, NY: John Wiley & Sons.

Esquivias, I., Odriozola, H., Tijero, J. M. G., et al. 2010. Simulation of high brightness tapered lasers. *Proc SPIE.* 7616:76161E.

Fiebig, C., Blume, G., Uebernickel, M., et al. 2009. High-power DBR-tapered laser at 980 nm for single-path second harmonic generation. *IEEE J Sel Top Quantum Electron.* 15(3):978–983.

Fox, A. G. and Li, T. 1961. Resonant modes in a maser interferometer. *Bell Syst Tech J.* 40(2):453–488.

Guthre, J., Tan, G. L., Ohkubo, M., et al. 1994. Beam instability in 980 nm power laser: Experiment and analysis. *IEEE Photon Technol Lett.* 6(12):409–1411.

Hadley, G. R. 1992. Wide-angle beam propagation using Padé approximant operators. *Opt Lett.* 17(20):1426–1428.

Harold^{TM} 3.0 Reference Manual. 2001. Photon Design, Oxford (UK).

Hasler, K.-H., Sumpf, B., Adamiec, P., et al. 2008. 5-W DBR tapered lasers emitting at 1060 nm with a narrow spectral linewidth and a nearly diffraction-limited beam quality. *IEEE Photon Technol Lett.* 20:1648–1650.

ISO 11146. 2005. Lasers and laser-related equipment—Test methods for laser beam widths, divergence angles and beam propagation ratios.

Kintzer, E. S., Walpole, J. N., Chinn, S. R., et al. 1993. High-power, strained-layer amplifiers and lasers with tapered gain regions. *IEEE Photon Technol Lett.* 5:605–607.

Krakowski, M., Auzanneau, S. C., Berlie, F., et al. 2003. 1 W high brightness index guided tapered laser at 980 nm using Al-free active region materials. *Electronics Lett.* 39(15):1122–1123.

Krakowski, M., Auzanneau, S. C., Calligaro, M., et al. 2002. High power and high brightness laser diode structures at 980 nm using Al-free materials. *Proc SPIE.* 4651:80–91.

Lang, R. J., Dzurko, K., Hardy, A. A., et al. 1998. Theory of grating-confined broad-area lasers. *IEEE J Quantum Electron.* 34:2196–2210.

Mariojouls, S., Margott, S., Schmitt, A., et al. 2000. Modeling of the performance of high-brightness tapered lasers. *Proc SPIE.* 3944:395–406.

Michel, N., Hassiaoui, I., Calligaro, M., et al. 2005. High-power diode lasers with an Al-free active region at 915 nm. *Proc SPIE.* 5989:598909.

Michel, N., Odriozola, H., Kwok, C. H., et al. 2009. High modulation efficiency and high power 1060 nm tapered lasers with separate contacts. *Electron Lett.* 45(2):103–104.

Müller, A., Fricke, J., Bugge, F., et al. 2016. DBR tapered diode laser with 12.7 W output power and nearly diffraction-limited, narrowband emission at 1030 nm. *Appl Phys B.* 122:87.

O'Brien, S., Welch, D. F., Parke, R. A., et al. 1993. Operating characteristics of a high-power monolithically integrated flared amplifier master-oscillator power-amplifier. *IEEE J Quantum Electron.* 29:2052–2057.

Odriozola, H., Tijero, J. M. G., Borruel, L., et al. 2009. Beam properties of 980 nm tapered lasers with separate contacts: Experiments and simulations. *IEEE J Quantum Electron.* 45(1):42–50.

Pagano, R., Ziegler, M., Tomm, J. W., et al. 2011. Two-dimensional carrier density distribution inside a high power tapered laser diode. *Appl Phys Lett.* 98(22):221110.

Paschke, K., Bogatov, A., Drakin, A., et al. 2003. Modeling and measurements of the radiative characteristics of high-power α-DFB lasers. *IEEE J Select Topics Quantum Electron.* 9:835–843.

Paschke, K., Sumpf, B., Dittmar, F., et al. 2005. Nearly diffraction limited 980-nm tapered diode lasers with an output power of 7.7W. *IEEE J Select Topics Quantum Electron.* 11(5):1223–1227.

Pfahler, C., Eichhorn, M., Kelemen, M. T., et al. 2006. Gain saturation and high-power pulsed operation of GaSb-based tapered diode lasers with separately contacted ridge and tapered section. *Appl Phys Lett.* 89:021107.

Ruiz, M., Odriozola, H., Kwok, C. H., et al. 2009. High-brightness tapered lasers with an Al-free active region at 1060 nm. *Proc SPIE.* 7230:72301D.

Siegman, A., Nemes, G., and Serna, J. 1998. How to (maybe) measure laser beam quality. *OSA TOPS.* 17(2):184–199.

Spreemann, M., Lichtner, M., Radziunas, M., et al. 2009. Measurement and simulation of distributed-feedback tapered master-oscillator power-amplifiers. *IEEE J Quantum Electron.* 45:609–616.

Sujecki, S., Borruel, L., Wykes, J., et al. 2003. Nonlinear properties of tapered laser cavities. *IEEE J Sel Top Quantum Electron.* 9:823–834.

Sumpf, B., Hasler, K.-H., Adamiec, P., et al. 2009. High-brightness quantum well tapered lasers. *IEEE J Sel Top Quantum Electron.* 15(3):1009–1020.

Sumpf, B., Hülsewede, R., Erbert, G., et al. 2002. High-brightness 735 nm tapered diode lasers. *Electron Lett.* 38(4):183–184.

Tijero, J. M. G., Pérez-Serrano, A., del Pozo, G., and Esquivias, I. 2017. Tapered semiconductor optical amplifiers. In *Handbook of Optoelectronic Device Modeling and Simulation*. vol 1, (Piprek, J., ed.) pp.695–712. Boca Raton, FL: Taylor & Francis.

Vurgaftman, I., Meyer, J. R., and Ram-Mohan, L. R. 2001. Band parameters for III–V compound semiconductors and their alloys. *J Appl Phys.* 89(11):5815–5875.

Walpole, J. N. 1996. Semiconductor amplifiers and lasers with tapered gain regions. *Opt Quantum Electron.* 28:623–645.

Wenzel, H., Sumpf, B., and Erbert, G. 2003. High-brightness diode lasers. *C R Physique.* 4:649–661.

Wilson, F. J., Lewandowski, J. J., Nayar, B. K., et al. 1999. 9.5W cw output power from high brightness 980 nm InGaAs/AlGaAs tapered laser arrays. *Electron Lett.* 35:43–45.

Williams, K. A., Penty, R. V., White, I. H., et al. 1999. Design of high-brightness tapered laser arrays. *IEEE J Select Topics Quantum Electron.* 5:822–831.

29

High-Brightness Laser Diodes with External Feedback

Mohamad Anas Helal

Simeon N. Kaunga-Nyirenda

Steve Bull

and

Eric Larkins

29.1 Power Scaling and the Role of Beam Quality........................... 82
29.2 High-Brightness Laser Diode Sources.................................... 83
Vertical Cavity Design • Lateral Cavity Design • Evolution of Lateral Beam Quality of High-Brightness Tapered Laser Diode
29.3 External Cavity Laser Simulation 92
Laser Diode Simulation Tool "Speclase" • External Cavity Simulation Tool • Coupling of the Laser Diode Simulation Tool and Coherent Ray Tracing Tools
29.4 Case Study: The Impact of Unintentional Reflections on a DBR-Tapered Laser 95
Parasitic Reflections with On-Axis Alignment • Parasitic Reflections with Off-Axis Alignment • Analysis and Discussion
29.5 Conclusions 105

High-power laser diodes first gained interest as pump sources for solid-state lasers, with cost/watt and reliability as the primary market drivers. In recent years, new markets have been opened by increases in the brightness of the laser diode sources and by advances in the optical systems for beam shaping and combining. The work in this chapter is motivated by the development of high-brightness laser diodes for direct-diode laser systems targeting industrial applications and a desire to illustrate the important role of laser simulation tools at both the device and system level. Traditionally, welding and sheet metal cutting are the most lucrative industrial laser markets, but they are also the most demanding in terms of brightness. Laser additive manufacturing processes, such as selective laser melting, are also quickly becoming a reality and allow the fabrication of structures that cannot be made by traditional means. Their automated nature is opening the door to new manufacturing paradigms.

Direct-diode laser systems for industrial applications combine the beams from many individual laser diodes (or diode arrays) to couple them into an optical fiber for delivery to the target. The role of the optical system is to combine the individual beams without losing their brightness. High-power direct-diode lasers rely on multiple beam-combining methods, including incoherent or "side-by-side" beam combining, polarization multiplexing, and spectral beam combining. The first commercial kW-class direct-diode laser system used an external cavity to stabilize the wavelengths of, and spectrally combine, the beams of a large number of broad-area (BA) lasers to couple a 1 kW beam into a 200 μm fiber (Huang et al., 2011). The brightness of the final system is ultimately limited by the brightness of the individual sources.

All beam-combining techniques couple light back into the laser diodes—deliberately, as in external wavelength stabilization and spectral beam combining, or accidentally due to reflections off the optics. This optical feedback is known to affect both the beam quality and the degradation of the laser diodes (Hempel et al., 2013), but little has been published on how this feedback affects the operation of the laser. As the power and performance requirements increase, the role of external optical feedback is becoming increasingly interesting.

This chapter focuses on the impact of external optical feedback on high-brightness laser diodes. We start by introducing the common beam quality metrics used for lasers and briefly review the diode laser technologies most commonly considered for high-brightness direct-diode laser systems, before focusing on the tapered laser and showing why they are strong contenders for future high-power direct-diode laser systems. We then describe the coupling of our continuous wave (CW) simulation tool, *Speclase*, to commercial optical design tools to self-consistently simulate high-brightness diode lasers with external optical feedback. We conclude with a case study exploring the impact of unintentional feedback on the excitation of higher order vertical modes and lateral beam quality in a large optical cavity (LOC) tapered laser.

29.1 Power Scaling and the Role of Beam Quality

The purpose of the beam-combining optics in a direct-diode laser is usually to couple the highest amount of power from the individual laser diode sources into the end of an optical fiber. The brightness of the combined beam (units = W cm^{-2} sr^{-1}) limits the power that can be coupled into the delivery fiber and depends on the brightness of the individual diode sources.

Metrics such as the beam parameter product (BPP) Q or beam propagation factor M^2 are common metrics used by manufacturers and industries to specify the beam quality of both the individual laser diode and direct-diode laser systems. The BPP is the product of beam radius (measured at the beam waist) and the half-angle beam divergence:

$$Q = \omega_0 \times \theta_{\text{div}}, \tag{29.1}$$

where ω_0 is the beam radius measured in millimeter and θ_{div} measured in milliradian and Q is the BPP measured in mm·mrad. The BPP allows optical designers to determine the number of individual beams that can be imaged onto the end of a fiber with a fixed diameter (physical aperture) and maximum acceptance angle (numerical aperture). The power coupled into the fiber also depends on the power of the combined beam (and hence of the individual emitters). The brightness (units: W cm^{-2} sr^{-1}) is defined as

$$B = \frac{P}{\pi^2 Q^2} = \frac{P}{\lambda^2 M^2} = \frac{P}{\lambda^2 M_x^2 M_y^2}, \tag{29.2}$$

where P is the laser output power, Q is the BPP, λ is the wavelength and M^2 is the beam propagation factor. M^2 is another measure of beam quality as defined by ISO 11146. The beam quality metrics, M^2 and Q are related by Equation 29.2, as shown in Table 29.1 for $\lambda = 975$ nm. (A detailed discussion of beam quality metrics is in Chapter 28.)

Figure 29.1 shows the power and beam quality needed for different industrial laser applications. The dashed lines show the improvement in laser diodes (and direct-diode laser systems based on them) between 2000 and 2014. A laser with a power of 20 W and $M^2 \sim 20$–30 (e.g., a good BA laser diode [Thestrup et al., 2003]) is suitable for printing and material processing, but not for additive manufacturing, welding, or cutting. Conversely, a laser with a power of 10 W and $M^2 < 1.5$ (e.g., a good tapered laser diode [Fiebig et al., 2008]) is suitable for sheet metal drilling, marking, additive manufacturing, welding, and cutting of metal.

TABLE 29.1 Correspondent Values of M^2 and Q

M^2	Q mm·mrad
1.0	0.31
1.5	0.47
2.0	0.62
3.0	0.93
5.0	1.6
10.0	3.1
20.0	6.2
30.0	9.3

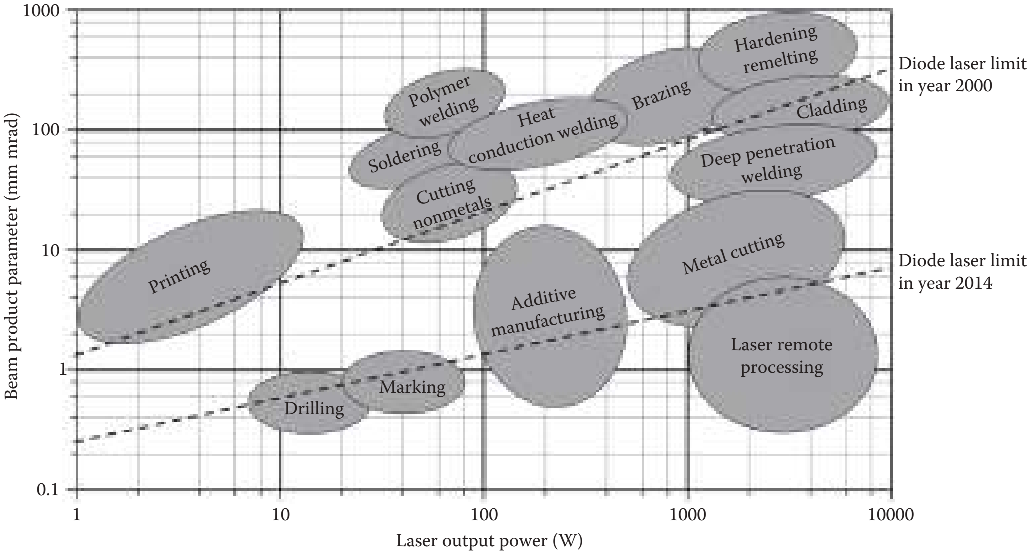


FIGURE 29.1 Beam parameter product versus laser power for different applications.

29.2 High-Brightness Laser Diode Sources

Laser diodes for the most demanding industrial applications should have high brightness (e.g., $P \geq 10$ W, $Q < 1$ mm·mrad, or $M^2 < 3$) and high-power conversion efficiency (PCE) (e.g., $PCE \geq 55\%$). The sources (and their performance) should be reliable and insensitive to reflections.

29.2.1 Vertical Cavity Design

The vertical cavity design of a laser diode for efficient, high-power operation in an external cavity is more complicated than that of an isolated high-power laser diode. First, a wide vertical mode profile (vertical near-field pattern) is needed for efficient external cavity coupling and reduced alignment tolerance. This also increases the tolerance to "smile" (a bend in the horizontal axis of a laser bar, which introduces pointing errors in the slow axis) across a laser array and allows the use of simpler, less costly optics. Second, the excitation and lasing of higher order vertical cavity modes by external optical feedback must be suppressed; photons coupled from external cavity will lead to stimulated emission. Third, a narrow vertical far-field divergence is needed to minimize the impact of optical aberrations, thereby improving the cost and performance of the optical system. Finally, the need for high-PCE means that Joule heating and free-carrier absorption (FCA) play critical roles in the vertical cavity design (Larkins et al., 2014).

High-brightness laser diodes typically use LOC waveguides to produce a wide vertical near-field pattern and narrow far-field divergence pattern. The LOC cavity reduces the power density at the facet (increasing the mirror damage threshold) and reduces the vertical confinement factor (reducing carrier-induced lensing). The confinement factor is the ratio of modal power contained in the active area to that contained in the structure. Below is the definition of the vertical confinement factor:

$$\Gamma(y) = \frac{\int_{-d/2}^{+d/2} E(y)^2 \mathrm{d}y}{\int_{-\infty}^{+\infty} E(y)^2 \mathrm{d}y}. \tag{29.3}$$

The LOC waveguide also facilitates external cavity feedback (e.g., for wavelength stabilization and/or lateral mode filtering), but also makes the device more sensitive to parasitic reflections.

As the LOC waveguide thickness increases beyond a certain point, the influence of the cladding layer diminishes and the waveguide is formed by the index contrast of the active region and the waveguide. Further increases in waveguide thickness raise the Joule heating and FCA losses, but do not significantly increase the near-field width or reduce the far-field divergence. Wider near-field patterns can be achieved (with thinner, more efficient waveguides) using other methods, including low-index quantum barriers to reduce the index contrast between the active region and the waveguide (Wang et al., 2013); or high-index optical traps to draw the field profile out into the waveguide (Buda et al., 1999).

LOC waveguides support multiple vertical modes, but, in the absence of external feedback, lasing is limited to the mode that reaches threshold first. When operated in a system with external feedback, however, back coupling to higher order modes can cause them to lase. Even if the higher order modes do not lase, the additional stimulated emission and FCA can reduce the laser's efficiency and increase self-heating. External feedback also affects the reliability of high-power laser diodes (Tomm et al., 2011). Thus, feedback is becoming important as diode laser systems grow in power and complexity.

The suppression of higher order vertical modes becomes more challenging when the laser is operated in an external cavity, where the optical feedback also excites the higher order vertical modes. First, as little power as possible must couple into the higher order vertical modes. Second, the higher order vertical modes must be prevented from reaching threshold. This can be achieved by engineering the laser cavity to make the confinement factors of higher order modes much lower than that of the fundamental mode, so that their modal gain is lower. Accordingly, we define a new figure of merit for modal discrimination (MD):

$$\mathrm{MD} = \Gamma(1)/\Gamma(n), \tag{29.4}$$

where Γ is the optical confinement factor and n refers to the higher order mode with the largest confinement factor. MD_n can also be used to describe the MD of a particular mode n. (*Note:* Propagation loss is neglected in MD, as it depends on the doping profile and operating bias—which are usually optimized *after* the initial cavity design.)

Power from the external cavity can be coupled into the higher order modes, even if they do not reach threshold. The modal gain of the higher order modes can also be reduced by increasing their propagation loss—either by FCA or by substrate leakage. Damping of the higher order vertical modes also reduces the total amplified spontaneous emission (ASE) from these parasitic modes. ASE and gain of light coupled into these modes do not contribute to the fundamental mode. Instead, they act as current leakage paths, contributing to self-heating and reducing the PCE.

LOC laser structures based on both the low-index quantum barrier and the high-index optical trap approaches were optimized (optically and electrically) for operation with external optical feedback (Larkins

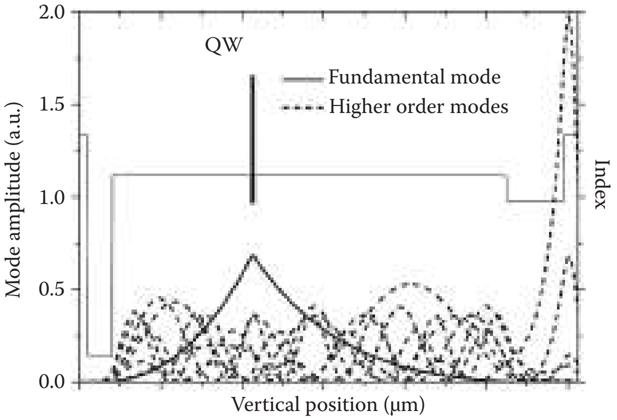


FIGURE 29.2 Vertical mode profiles in the ELoD2 LOC laser structure.

TABLE 29.2 Confinement Factor and MD Values for All Vertical Modes in ELoD2 Structure

Mode Number	Confinement Factor (Γ)	Modal Discrimination (MD)
1	0.0163	1
2	0.00376	4.33
3	0.000282	57.72
4	0.00856	1.90
5	0.00390	4.17
6	1.090E-05	1491.39
7	0.00130	12.58
8	0.00424	1.92

et al., 2014). The tapered lasers in this chapter are based on the ELoD2 vertical cavity with low-index quantum barriers (Crump et al., 2013c). This structure supports eight vertical modes, as shown in Figure 29.2. Their confinement factors and MD are given in Table 29.2.

29.2.2 Lateral Cavity Design

State-of-the-art high-power lasers need a large output power, high PCE, and excellent beam quality. Ridge waveguide (RW) lasers can produce a diffraction limited beam (single lateral mode), but have only achieved output powers of 1.6 W (Yang et al., 2004). Flared RW lasers have achieved a maximum power of 3 W and kink-free power of 2.2 W (Sverdlov et al., 2013). To achieve higher output powers, the laser diode needs a larger gain volume (energy reservoir). This can be most easily achieved by increasing the emitter width, which has the added benefit of lowering the power density at the facet—thereby increasing the catastrophic optical mirror damage (COMD) threshold.

Slab-coupled optical waveguide laser (SCOWL) diodes (Donnelly et al., 2003) use a shallower RW etch to allow the beam profile to expand laterally in the "slab" below. They have been used as high-brightness sources for power scaling by spectral beam combining (Huang et al., 2009). SCOWLs have good beam quality and low astigmatism, but their limitations are similar to those of the flared RW laser and comparatively low power (2.8 W) (Huang et al., 2007) limits their single-emitter brightness.

BA diodes have achieved powers of 29 W, reliable operation at 20 W, and record PCE (76%) at 0°C. BA lasers have simultaneously achieved high power *and* high PCE at 300 K (14.5 W, PCE >60%)

(Crump et al., 2009, 2013a). Their main limitation is their poor slow-axis beam quality ($M^2 \sim 20$–30) (Knigge et al., 2005).

Lim et al. used simulations to investigate the dependence of the beam quality of BA lasers on cavity length and width as a function of power (Lim et al., 2005). Lim's work showed that the brightness of gain-guided BA lasers improved with reduced cavity width and increased cavity length—establishing a new basis for the development of narrow broad-area (NBA) laser diodes. Since then, NBA laser diodes have received increasing attention (Crump et al., 2013b; Decker et al., 2014; Skidmore et al., 2016), with powers of 7.5 W with $Q_{\text{slow axis}} = 1.8$ mm·mrad and high efficiency (PCE = 57%) (Crump et al., 2013b). Despite the impressive performance of NBA laser diodes, they still fall short of the desired performance for the most demanding applications.

The beam quality of BA lasers can also be improved using asymmetric feedback to reduce the threshold of a particular higher order lateral mode (Pillai and Garmire, 1996; Thestrup et al., 2003; Wolff et al., 2003; Lang et al., 2008), exploiting a concept previously demonstrated for phase-locked laser arrays (). However, the usefulness of asymmetric feedback stabilization for high-brightness operation still needs to be demonstrated—in particular, whether they can maintain their high beam quality and PCE at high power

Tapered laser diodes have received great interest, since they combine the lateral mode confinement/filtering of the RW laser and the large gain volume of the BA laser. Thus, tapered lasers are able to produce a high output power (12 W) with high beam quality ($M^2 < 1.2$) and exceptional brightness ($B = 1.1$ GW·cm^{-2} sr^{-1}) (Walpole et al., 1992; Kelemen et al., 2005; Sumpf et al., 2010). This comes at the expense of lower PCE (43%–55%) and power-dependent "wandering" astigmatism (Dittmar et al., 2006; Fiebig et al., 2008). Thus, the tapered laser is promising for high-brightness direct-diode laser systems. The main challenges are to control their beam quality degradation at high power and to understand the role of optical feedback.

29.2.3 Evolution of Lateral Beam Quality of High-Brightness Tapered Laser Diode

The tapered laser diode comprises RW and tapered amplifier (TA) sections. The RW supports a single lateral mode, to inject a diffraction-limited beam into the amplifier. It also filters out higher order lateral modes from the backward traveling fields it receives from the tapered section. The TA provides a large gain volume with a large output aperture. The taper angle is chosen to match the diffraction angle from the RW (Pearson et al., 1969; Walpole et al., 1992), allowing the field to expand in the taper to produce a smooth output field profile, as illustrated in Figure 29.3.

A nearly diffraction-limited lateral beam is obtained from conventional tapered lasers at low to moderate power, but the beam quality degrades rapidly at high power. This beam quality degradation reveals degradation of the RW filter performance due to gain saturation inside the RW and absorption bleaching by the backward traveling fields outside of it (Sujecki et al., 2003; Kaunga-nyirenda et al., 2014; Larkins et al., 2014, 2016). The absorption bleaching (due to band filling by the generated carriers) renders the material transparent and allows the backward traveling fields outside the RW to reach the rear facet, where they are reflected and reenter the TA, as shown in Figure 29.4

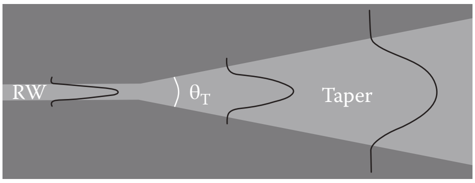


FIGURE 29.3 Schematic diagram of a tapered laser, showing the adiabatic expansion of the beam in the taper.

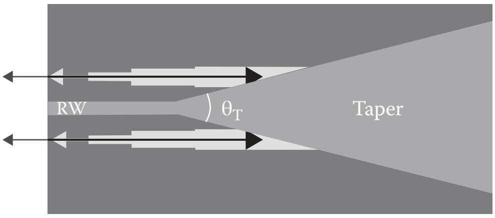


FIGURE 29.4 The beam outside the ridge waveguide (RW) bleaches the absorption in these regions, allowing the reflected beam from the back fact to travel back and couple into the tapered amplifier (TA)

These unfiltered fields create high-spatial-frequency features in the near-field pattern. These high-frequency features or filaments are caused by the propagation and diffraction of the backward-traveling fields in the absorption bleached regions outside the RW (Lim et al., 2012; Larkins et al., 2014). This diffraction is primarily due to the diffraction of the backward propagating fields that are *not* coupled into the RW. Figure 29.5a shows the back propagating field in a 975 nm-tapered laser at $I = 15$A. The backward propagating fields are amplified in the taper and gain guided at the edges of the taper to form a high-intensity spot where the RW joins the TA. The fields diffracting from this spot are clearly visible alongside of the RW. Figure 29.5b shows the backward propagating fields from this high-intensity spot in the absence of index and gain-guiding effects. The field pattern agrees with the corresponding region in Figure 29.5a (inside the white box)—except that the central lobe widens due to the absence of index guiding by the RW. In Figure 29.5c, only the central field lobe (the lobe that couples into the RW) was back propagated, while in Figure 29.5d only the outer lobes of the spot were propagated. Figure 29.5c and d shows that only the fields *not* coupled into the RW give rise to the diffraction pattern—and thus, for the absorption bleaching and beam quality degradation.

Finally, spatial hole burning at the center of the taper and electrical over-pumping at the edges creates high carrier densities and gain at the edges of the taper, as seen in quantitative intracavity spontaneous emission imaging measurements (Bull et al., 2004, 2006). Amplification and carrier-induced waveguiding in these regions cause the "batman" ears in the near-field pattern and amplify/guide the backward fields that degrade the RW filter performance (Williams et al., 1999; Sujecki et al., 2003) Different techniques have been suggested to overcome the degradation of beam quality, such as reducing the front facet reflectivity, using beam spoilers or using a longer RW section. Figure 29.6 shows the impact of reducing front facet reflectivity on the total photon density distribution (i.e. the sum of densities of both the forward and backward traveling waves). These techniques all reduce the backward traveling field and associated absorption bleaching. At the emitter powers sought for high-brightness direct-diode lasers (>10 W), however, these techniques are insufficient. (Beam quality degradation in tapered lasers is also discussed in Chapter 28.)

An integrated distributed Bragg reflector (DBR) mirror at the end of the RW provides a solution with better high-power performance. The rear facet of the laser must be antireflection (AR) coated for the DBR to work properly. The DBR (and AR-coated facet) act a spatial filter, allowing the backward traveling fields in the regions outside the RW to exit the rear facet, as illustrated in Figure 29.7. In order to study the spatial filtering performance of the RW section, and understand how the DBR RW section improves the spatial filtering performance of the conventional tapered laser, we define the RW filter response as the ratio of the forward and backward propagating field distributions at the interface between RW and TA sections:

$$H_{\mathrm{RW}} = \log_{10}\left[\frac{P_{\mathrm{forward}}}{P_{\mathrm{backward}}}\right]. \tag{29.5}$$

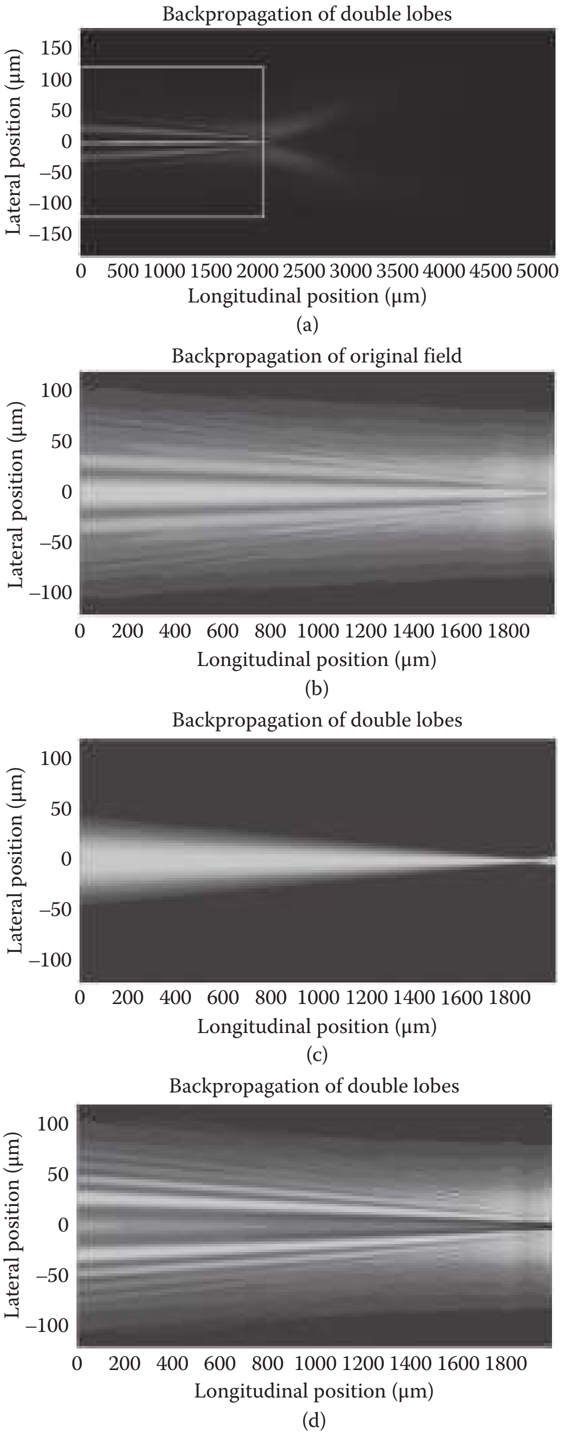


FIGURE 29.5 (a) Backward propagating field in 975 nm tapered laser ($I = 15$ A) for converged simulation. Backward-propagated photon distribution in the RW section for transparent, uniform index material for: (b) the total field; (c) the central lobe coupled into the RW; and (d) the fields *not* coupled into the RW.

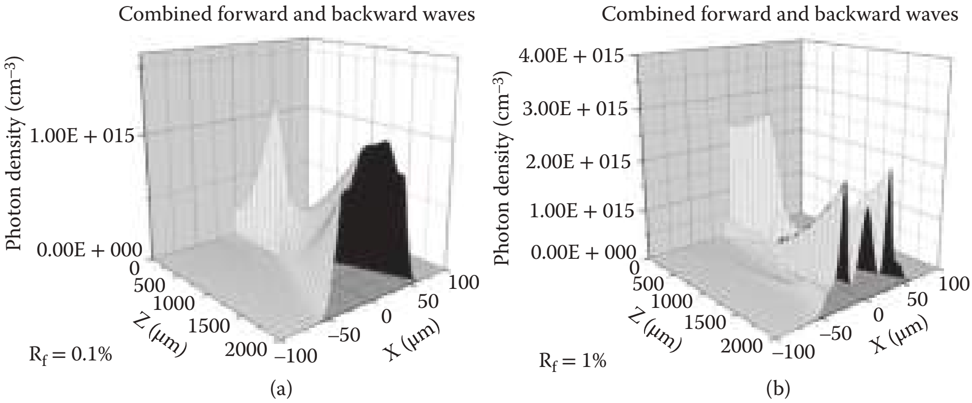


FIGURE 29.6 Impact of front facet reflectivity on the shape of forward and backward traveling waves. Both figures (a and b) show the total photon density distribution (forward + backward).

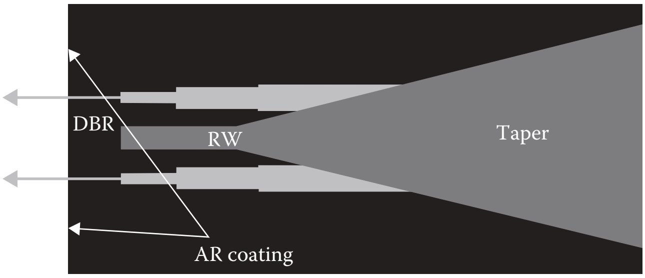


FIGURE 29.7 Schematic diagram of the DBR-tapered laser, showing how the back-propagating beams outside of the RW are allowed to leave the structure, instead of coupling back into the TA.

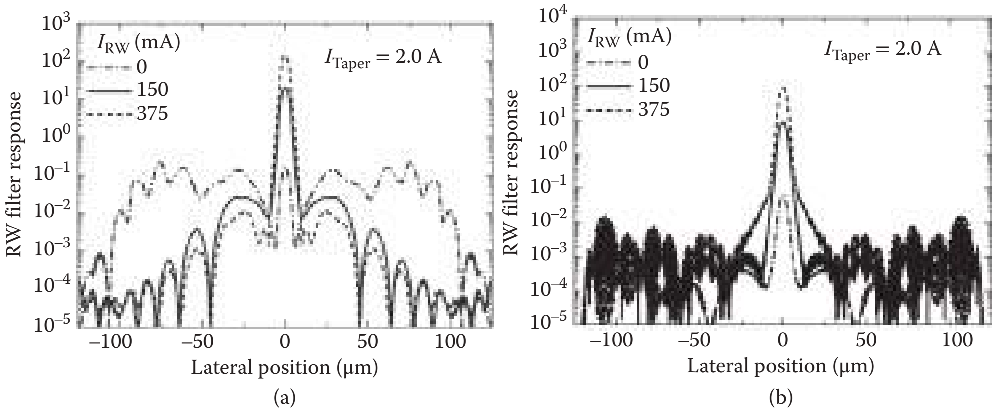


FIGURE 29.8 RW filter performance comparison of conventional (a) and DBR (b) tapered lasers.

Figure 29.8 compares the RW filter responses of DBR and conventional tapered lasers with dual contacts for a constant taper current as a function of RW current. In the conventional tapered laser, the fields in the RW are not amplified at the lowest RW current, but those outside the RW propagate and are reflected—reducing the performance of the RW filter. In the DBRtapered laser, these fields are not reflected and exit

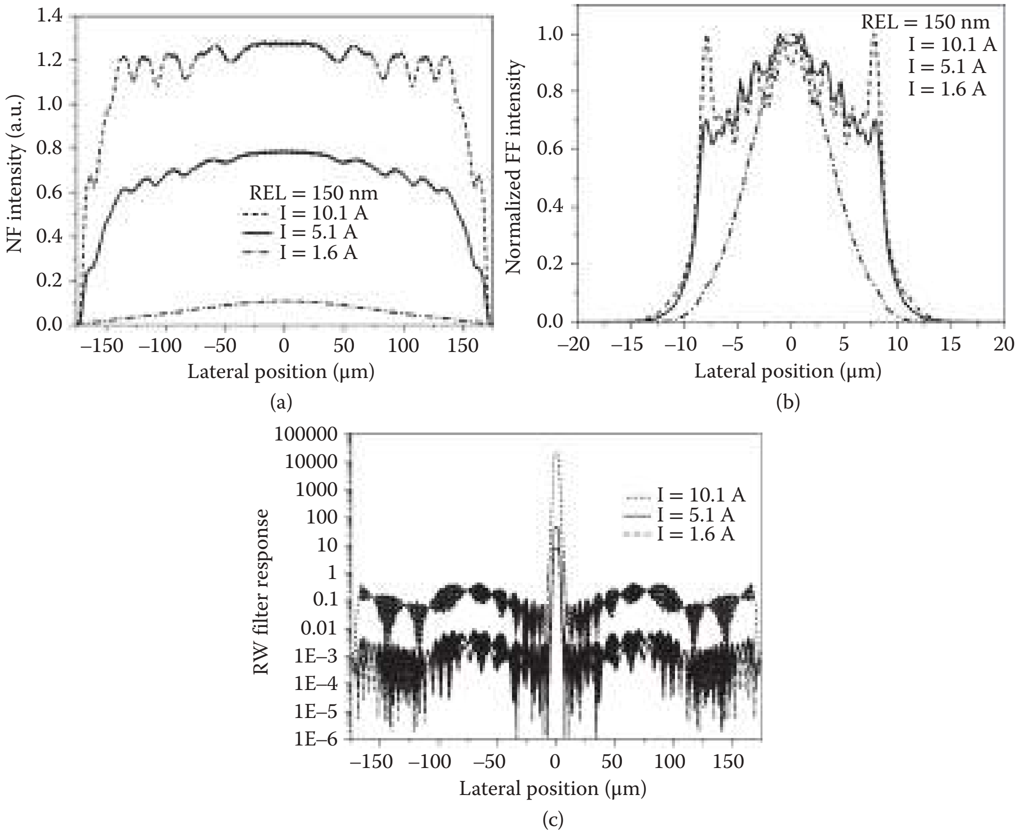


FIGURE 29.9 Evolution of near-field (a), far-field (b) and RW filter response (c) with current.

through the AR-coated rear facet. Thus, the filter performance of the DBRtapered laser remains good up to higher powers.

Despite the large performance improvement of the DBR-tapered laser at high powers, $M^2_{\text{2nd moment}}$ still increases with power. To understand this, we studied the evolution of the near- and far-field patterns and the RW response with current, as shown in Figure 29.9.

As the current increases, high spatial frequency features appear in the near-field pattern and grow in strength, causing degradation of the far-field pattern. The appearance of these features correlates with a sudden reduction in the RW filter performance. Furthermore, the RW width and etch depth have little impact on beam quality as seen in Figure 29.10. (Figure 29.10 also shows that $M^2_{e^{-2}}$ [beam radius w_0 in Equation 29.1 is measured at $1/e^2$] is independent of current, while $M^2_{\text{2nd moment}}$ [beam radius w_0 in Equation 29.1 is measured using second moment of area definition, as defined in Chapter 29] provides a more sensitive measure of the changes in the near and far-field patterns.)

Figure 29.11 shows the forward propagating field at the junction of the RW and TA sections. Although the fields outside the RW are small, they will be strongly amplified.

The fields outside the RW section explain the degradation of the filter function (and the changes in the near- and far-field patterns), but there is a problem: *If the backward propagating fields outside the RW all escape through the AR-coated rear facet, then where do these forward propagating fields come from?* The answer lies in the forward propagating photon distribution in Figure 29.12, which reveals a single aperture diffraction pattern—as confirmed by the positions of the nodes in the pattern. Although these simulations assume a simple patterned rear facet reflectivity, the DBR will also produce aperture diffraction.

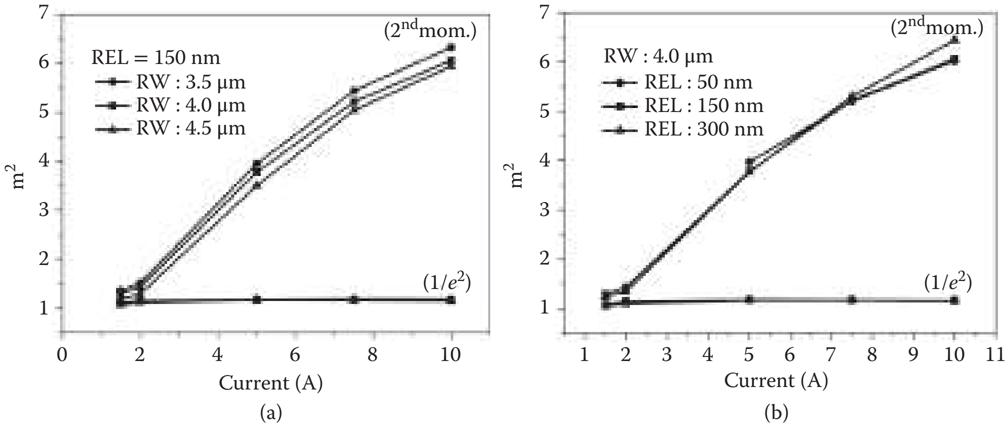


FIGURE 29.10 Impact of RW width (a) and RW etch depth (b) on beam quality.

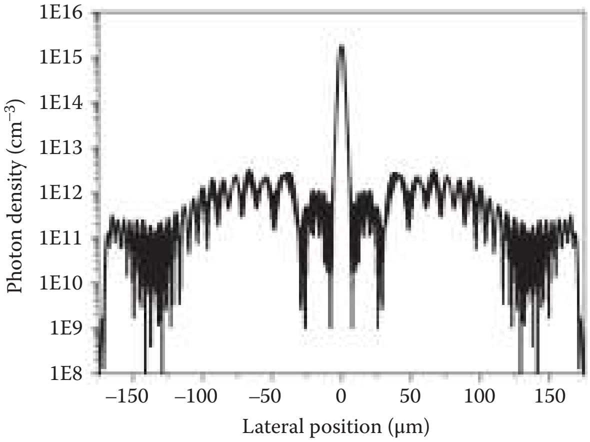


FIGURE 29.11 Forward propagating photon distribution at the RW/TA interface of a DBRtapered laser.

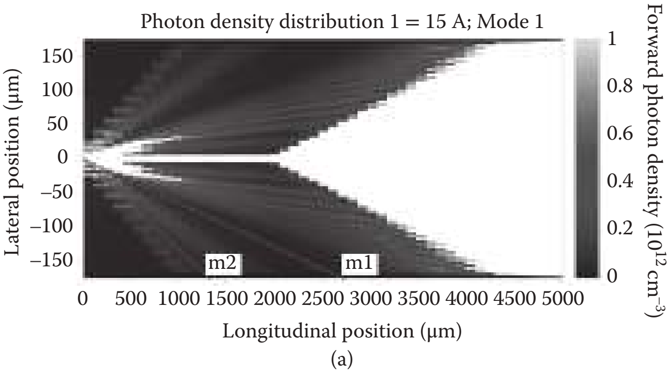


FIGURE 29.12 2D forward propagating photon density with a perfect AR-coating on the rear facet (a) and with an AR-coating with a power reflectivity of 0.1% (b). (*Continued*)

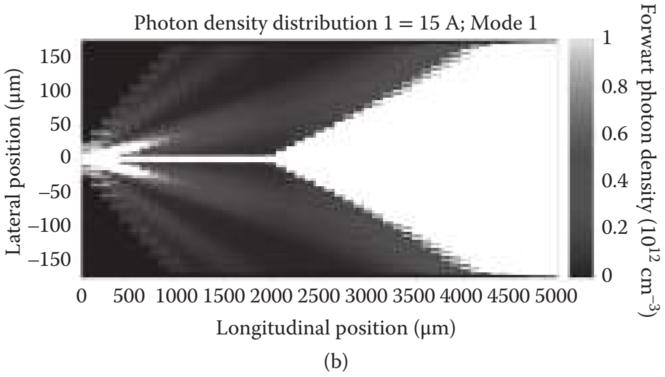


FIGURE 29.12 (Continued) 2D forward propagating photon density with a perfect AR-coating on the rear facet (a) and with an AR-coating with a power reflectivity of 0.1% (b).

29.3 External Cavity Laser Simulation

New simulation tools are needed to address the problems posed by high-brightness diode lasers operating in a system context with external optical feedback. The laser diode and the external cavity pose different modeling challenges and require the use of different modeling approaches and software tools. These approaches and tools must be brought together to model the laser diode and the external optical system self-consistently.

29.3.1 Laser Diode Simulation Tool "*Speclase*"

The simulation of high-brightness laser diodes is a challenging task, which requires self-consistent modeling of the electrical, optical, and thermal processes throughout the device (Williams et al., 1999; Lim, 2003; Sujecki et al., 2003; Lim et al., 2009). LOC devices with external feedback bring the additional challenge of simulating fields with different vertical mode profiles and their competition for the available gain.

The simulation of the laser diode is performed with an in-house laser simulation tool, *Speclase* (Lim et al., 2009), for the self-consistent quasi-3D optical, electrical, and thermal simulation of high-brightness laser diodes. *Speclase* uses the two-dimensional (2D) wide-angle finite-difference beam-propagation method (WA FD-BPM) for the optical field propagation in the longitudinal and lateral directions (x–z plane). Bipolar electrical simulations are performed in a series of planes orthogonal to the optical axis (x–y planes), neglecting heat and carrier flow in the z-direction. The electrothermal and optical solvers are coupled through the stimulated emission/absorption and ASE processes and refractive index perturbations (calculated from perturbations of the gain/absorption spectra using the Kramers–Kronig relations).

The optical solver models the spectral behavior of the device by propagating multiple wavelengths and their competition through the spectral-spatial gain distribution. Figure 29.13 shows the flow diagram of *Speclase*. The electrothermal and optical models are solved self-consistently, using an accelerated Fox–Li iterative approach (Agrawal, 1984).

In order to model the excitation and competition of fields with different vertical mode profiles, *Speclase* calculates each of the vertical field distributions and their vertical confinement factors. This allows *Speclase* to calculate the effective index and FCA distributions needed for the 2D propagation of the fields with each of the allowed vertical mode profiles. Figure 29.14 describes the flow diagram of *Speclase* for the simulation of a high-brightness LOC laser diode coupled to an external cavity.

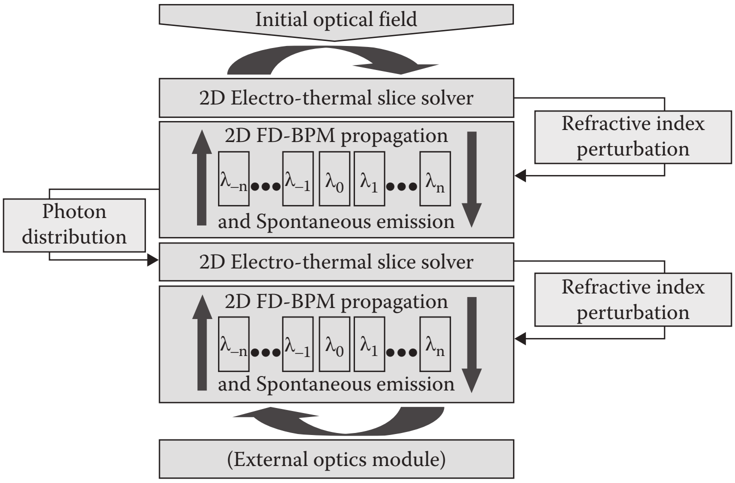

Initial optical field
2D Electro-thermal slice solver
2D FD-BPM propagation
λ_{-n} ••• λ_{-1} λ_0 λ_1 ••• λ_n
and Spontaneous emission
Refractive index perturbation
Photon distribution
2D Electro-thermal slice solver
Refractive index perturbation
2D FD-BPM propagation
λ_{-n} ••• λ_{-1} λ_0 λ_1 ••• λ_n
and Spontaneous emission
(External optics module)

FIGURE 29.13 Flow diagram of stand-alone (SA) *Speclase* with multiple wavelengths.

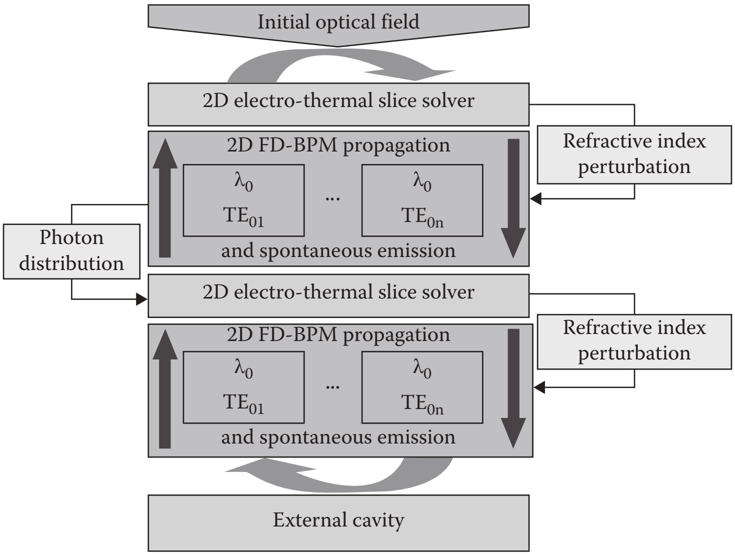

Initial optical field
2D electro-thermal slice solver
Refractive index perturbation
2D FD-BPM propagation
λ_0 TE_{01}
...
λ_0 TE_{0n}
and spontaneous emission
Photon distribution
2D electro-thermal slice solver
Refractive index perturbation
2D FD-BPM propagation
λ_0 TE_{01}
...
λ_0 TE_{0n}
and spontaneous emission
External cavity

FIGURE 29.14 Flow diagram of *Speclase* with multiple vertical modes with external cavity feedback.

29.3.2 External Cavity Simulation Tool

The modeling of the external cavity requires modeling of free-space optical propagation, including the transmission, reflection, and scattering effects. The optical models used inside the laser diode are not well suited to free-space optical propagation, where Fourier optics (FO) and coherent ray tracing methods are more appropriate. FO models preserve the wave nature of the optical propagation, but the underlying paraxial approximation is best suited to low-divergence beams. Coherent ray tracing is better suited for external cavity laser diodes with large beam divergence. Coherent ray tracing methods are also fast and flexible, making them attractive for modeling complex optical systems. We describe an external cavity modeling tool (*OpticStudio*) based on coherent ray tracing. FO models are also used in this chapter, but do not need wave-ray transformation and are not discussed here.

29.3.3 Coupling of the Laser Diode Simulation Tool and Coherent Ray Tracing Tools

The coupling of the laser simulator and the coherent ray tracing tool for the external cavity is not straight forward, since the light is treated as a wave inside the laser diode, but as rays in the external cavity. Careful analysis of the output beam of the laser diode is needed to produce the input rays for the external cavity model, while maintaining all of the properties of the beam emitted from the laser diode (e.g., near- and far-field patterns, astigmatism). Bidirectional coupling is even more challenging, requiring conversion of the rays back into an optical field distribution, which contains both intensity and phase information.

First, the field data produced by *Speclase* (near-field, far-field and wavelength) must be converted into rays, which represent the laser source in *OpticStudio*. These rays each have a set of attributes: launching coordinates, direction cosines and intensity.

Speclase simulates the lateral field distribution at the facet using 2D WA FD-BPM, so the lateral near-field pattern is multiplied by the associated vertical field profile. The horizontal and vertical far-field patterns are calculated by Fourier transformation:

$$\psi\left(\theta_H\right) = \left(\cos\theta_H\right)^2 \left| \int_{-\infty}^{+\infty} \mathrm{E}(x) e^{j(\sin\theta_H k_0 x)} \mathrm{d}x \right|^2 \tag{29.6}$$

$$\psi(\theta_V) = (\cos\theta_V)^2 \left| \int_{-\infty}^{+\infty} \mathrm{E}(y) e^{j(\sin\theta_V k_0 y)} \mathrm{d}y \right|^2, \tag{29.7}$$

where ψ is far-field profile, θ is the divergence angle, E is the electric field, and k_0 is the wavenumber. Due to the astigmatic nature of the laser beam, the output rays appear to originate from two virtual line sources (nonastigmatic sources have a virtual point source). The output rays must intersect both of these line sources—one for the vertical beam divergence and the other for the horizontal beam divergence. The longitudinal positions of these line sources are obtained by FO back propagation (horizontal line source) and by back tracing the rays (vertical line source). Their transverse positions are the first moments of the near-field patterns

The second challenge is to convert the rays returning to the laser facet back into complex field distributions, taking the ray phases into account. A fraction of the optical energy returned to the laser facet at each lateral position (x) is coupled into the laser for each of the vertical modes supported by the LOC structure, as determined by the overlap integral:

$$\eta_n = \frac{\left| \int_{-\infty}^{+\infty} \left\{ E_{\text{ext}} TE_n^* \right\} \mathrm{d}y \right|^2}{\int_{-\infty}^{+\infty} \left| E_{\text{ext}} \right|^2 \mathrm{d}y \int_{-\infty}^{+\infty} \left| TE_n \right|^2 \mathrm{d}y}, \tag{29.8}$$

where E_{ext} is the distribution of the external field incident on the facet at this lateral position, TE_n is the nth vertical field profile, and η_n is the percentage of the reflected power coupled into cavity mode n. The phase of the coupled field at each position can be calculated by

$$\varphi = \tan^{-1}\left[\frac{Im\left\{ \int_{-\infty}^{+\infty} \left\{ E_{\text{ext}} TE_n^* \right\} \mathrm{d}y \right\}}{Re\left\{ \int_{-\infty}^{+\infty} \left\{ E_{\text{ext}} TE_n^* \right\} \mathrm{d}y \right\}} \right]. \tag{29.9}$$

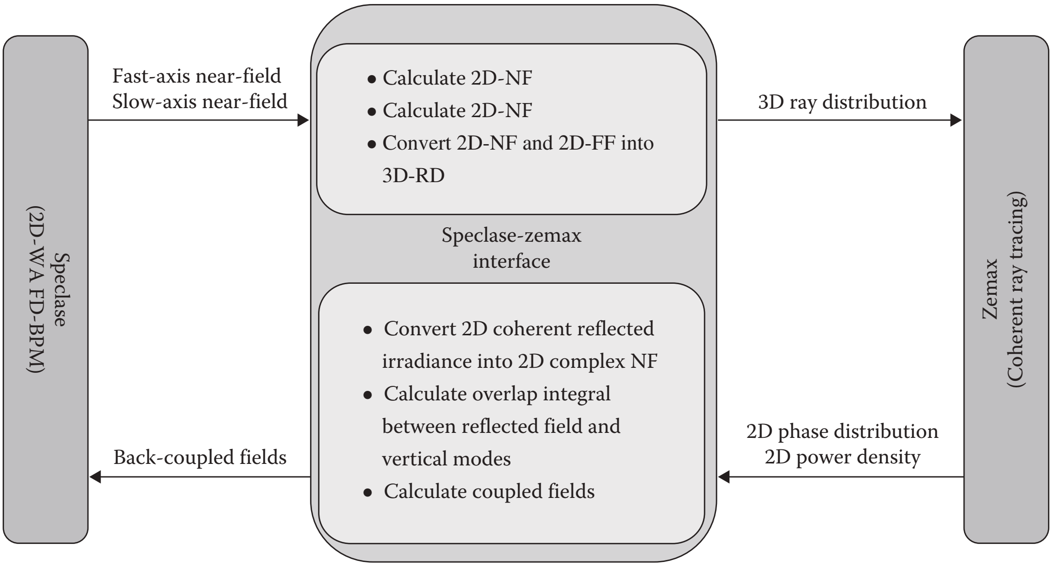

Speclase
(2D-WA FD-BPM)
Fast-axis near-field
Slow-axis near-field
• Calculate 2D-NF
• Calculate 2D-NF
• Convert 2D-NF and 2D-FF into 3D-RD
Speclase-zemax interface
• Convert 2D coherent reflected irradiance into 2D complex NF
• Calculate overlap integral between reflected field and vertical modes
• Calculate coupled fields
Back-coupled fields
3D ray distribution
2D phase distribution
2D power density
Zemax
(Coherent ray tracing)

FIGURE 29.15 Schematic of coupling *Speclase* to *OpticStudio*.

This process is repeated for the vertical field distributions at each lateral position along the facet. The resulting one-dimensional (1D) field distributions are coherently added to the internally reflected fields and propagated backward within the laser diode using 2D WA FD-BPM. The entire external cavity simulation is repeated using Fox–Li iterations until it converges. Figure 29.15 shows a schematic diagram of the coupling of *Speclase* and *OpticStudio*

29.4 Case Study: The Impact of Unintentional Reflections on a DBR-Tapered Laser

High-brightness tapered lasers are usually characterized and simulated in isolation, without back reflections to degrade the slow-axis beam quality or couple into higher order vertical modes. In this section, we simulate the simple scenario of an LOC tapered laser with fast- and slow-axis beam collimation lenses with the light incident on the end of an uncoated optical fiber ($R = 5\%$), as illustrated in Figure 29.16. The purpose of this study is to:

1. Observe how the feedback affects the power, PCE, and beam quality
2. Observe the impact of the feedback on the excitation of other vertical modes
3. Reveal the impact of self-heating from the excitation of other vertical modes
4. Investigate approaches for suppressing the impact of higher order vertical modes

The laser diode in this study is a 975-nm LOC DBR-tapered laser based on the ELoD2 structure (Crump et al., 2013c), with a 2-mm RW and a 3-mm gain-guided amplifier with a 6° taper. The DBR is represented by a patterned reflectivity ($R = 31\%$ inside the RW, $R = 0.1\%$ adjacent to it). The front facet reflectivity is 1%. Stray external cavity reflections are included using ray splitting and scattering. Figure 29.17 shows the simulated and experimental power and PCE versus current behavior of an ELoD2 BA laser.

High-power laser diodes are often fabricated as linear bar arrays or laser "bars." Laser bars often bend slightly when they are soldered to a heatsink, so that the emitter near-field patterns deviate from a straight line—an effect referred to as "smile." Figure 29.18 shows how "smile" of packaged laser bars displaces the emitters (±0.5 μm) from the axis This displacement causes the beam to propagate through the external

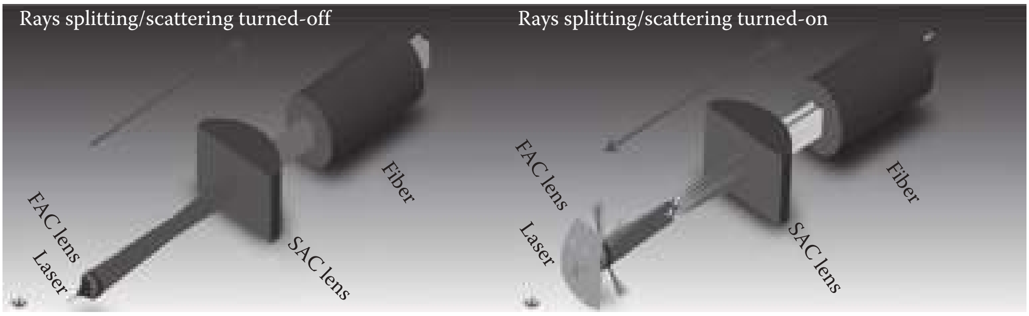

Rays splitting/scattering turned-off
Rays splitting/scattering turned-on
FAC lens
Laser
SAC lens
Fiber
FAC lens
Laser
SAC lens
Fiber

FIGURE 29.16 External cavity setup used in the simulation consists of a fast-axis collimation (FAC) lens, a slow-axis collimation (SAC) lens, and a fiber. Rays splitting and scattering turned off (left). Rays splitting and scattering turned on (right).

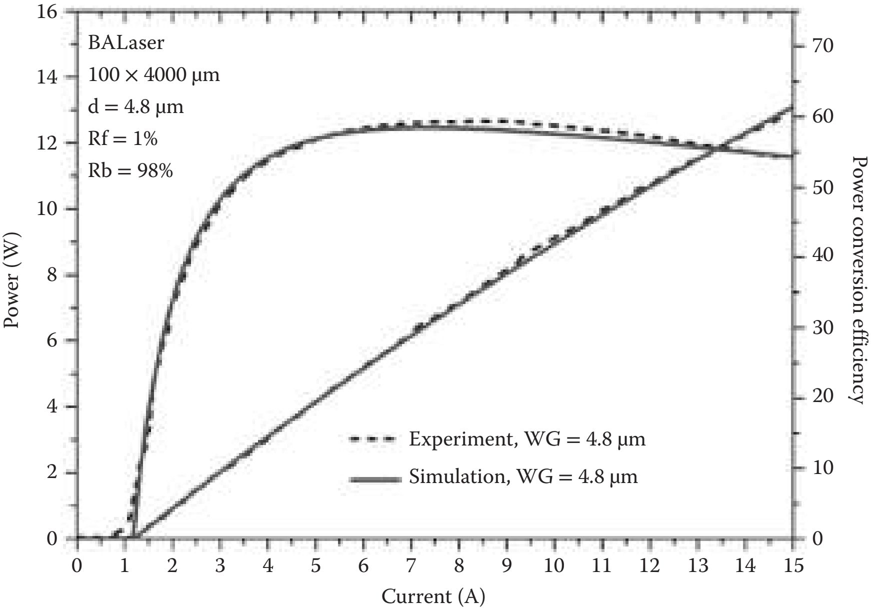


FIGURE 29.17 Simulation versus experiment for ELoD2 structure.

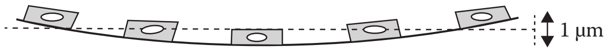
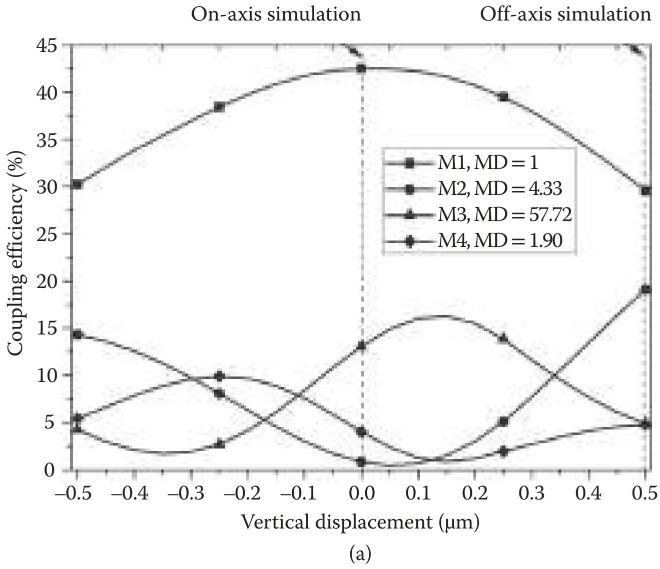


FIGURE 29.18 "Smile" of a linear laser array or bar, which is caused by bending as the bar is soldered to a heatsink.

optics at an angle to the optical axis, a phenomenon known as "pointing error" To emulate the impact of "smile" and "pointing error," the emitter position is shifted vertically with respect to the external optical system.

Two external cavity simulations were performed (on- and off-axis) and compared to a stand-alone (SA) simulation (no external feedback). The overlap integrals of the reflected field of the fundamental mode (from the SA simulation) and the cavity modes were calculated to estimate the coupling coefficients versus displacement, as shown in Figure 29.19. These curves show how displacement affects the feedback coupling to the vertical modes. The off-axis simulation was performed for a displacement of +0.5 µm,



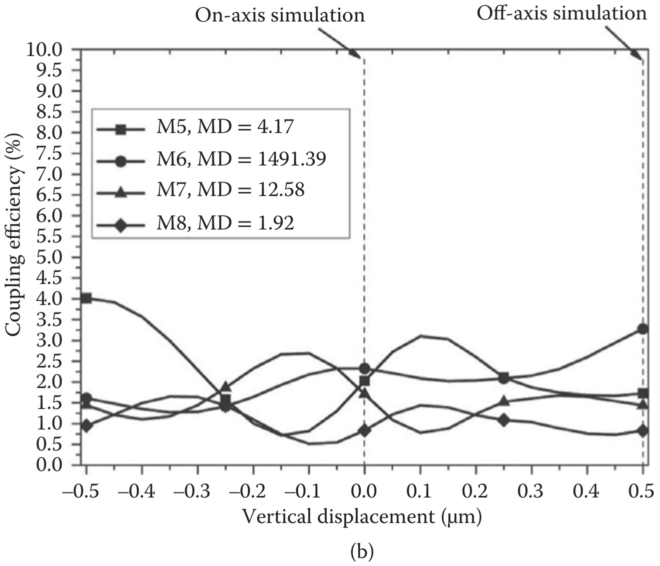
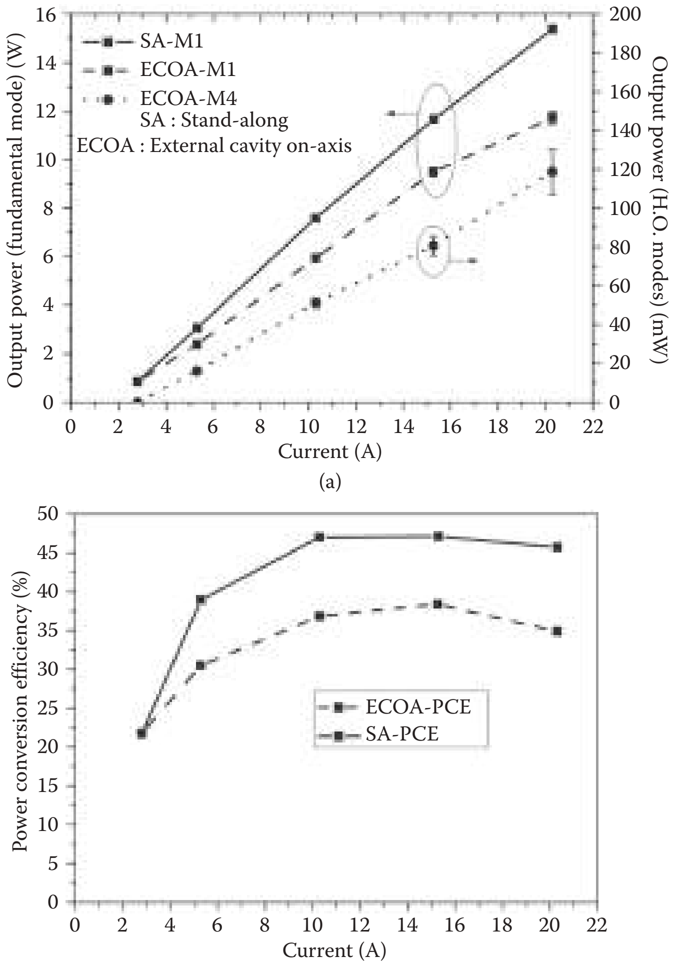


FIGURE 29.19 Impact of smile on external cavity coupling coefficients. Modes 1–4 (a) have coupling >10%, while modes 5–8 (b) have coupling <5%. An off-axis point of interest is displacement = +0.5μm.

where the fundamental mode coupling coefficient (29.5%) approaches that of a higher order mode (19%, mode 2).

29.4.1 Parasitic Reflections with On-Axis Alignment

Simulations were performed for two different external feedback conditions: no feedback; and feedback to the fields for all of the vertical mode profiles. Simulations were performed at bias currents of 2.5, 5.0, 10, 15, and 20 A. Figure 29.20 shows the power versus current and PCE versus current characteristics to compare the performance of the laser for these two feedback conditions. Our discussion will focus on the performance at a bias current of 15 A, since this produces an output power in the range of interest (10–12 W) for the targeted application.



(a)

(b)

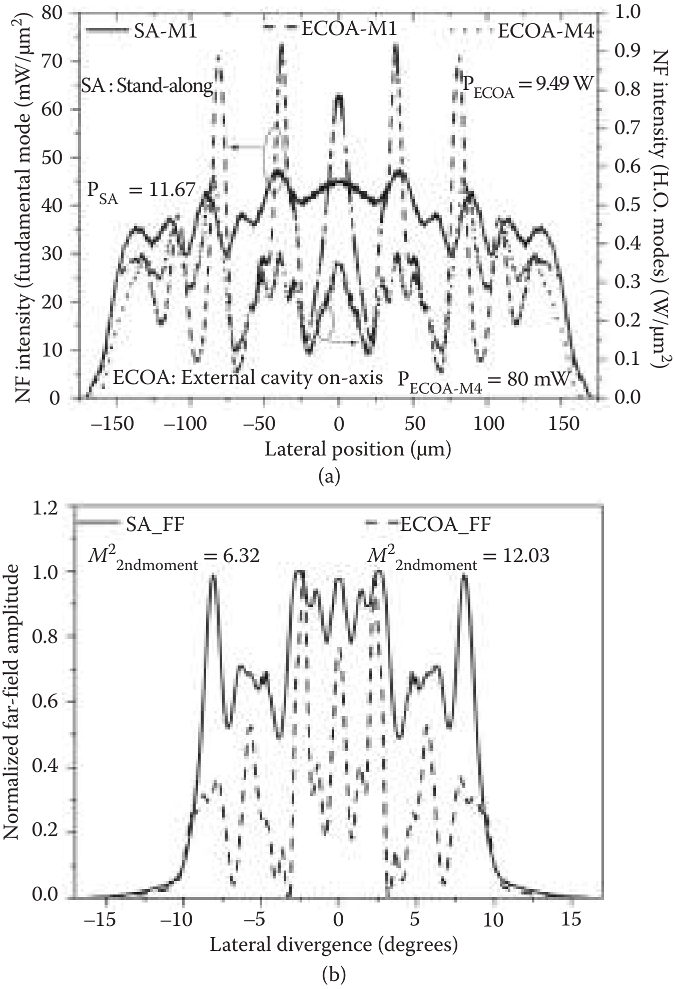
FIGURE 29.20 Simulated power (a) and power conversion efficiency (b) of the tapered laser with and without on-axis feedback.

Figure 29.20 shows that the optical feedback reduces the output power and the PCE. At $I = 15$ A, the on-axis feedback caused the fundamental mode power (mode 1) to drop from 11.67 to 9.49 W, while the PCE fell from 47.1% to 38.4%.

Figure 29.21 compares the near-field patterns (left) and far-field patterns (right) of the on-axis laser for the two feedback conditions. The feedback reduces the output power and increases the modulation depth of the high spatial frequency components in the near-field pattern, as observed experimentally (Hempel et al., 2013; Leonhäuser et al., 2014). The feedback may also lead to excitation and lasing of higher order modes—in this case mode 4.

Figure 29.22 shows the current dependence of the beam propagation parameter M^2 (both the e^{-2} and second moment definitions) for the on-axis external cavity feedback conditions. At $I = 15$ A, the feedback had little effect on $M^2_{e^{-2}}$, but $M^2_{\text{2nd moment}}$ increased from 6.32 to 12.03. This shows that aperture diffraction from the DBR dominates the beam quality degradation.

The total photon distributions (forward and backward propagating) with on-axis feedback are shown in Figure 29.23 for the first four vertical modes. Only modes 1 and 4 are lasing, but light is also coupled into the other modes. Although they do not lase or produce an output beam (and are difficult to characterize and



(a)

(b)

FIGURE 29.21 Lateral nearfield profiles of all lasing vertical modes (a) and lateral far-field profiles of the fundamental vertical mode (b) for different on-axis external cavity feedback conditions.

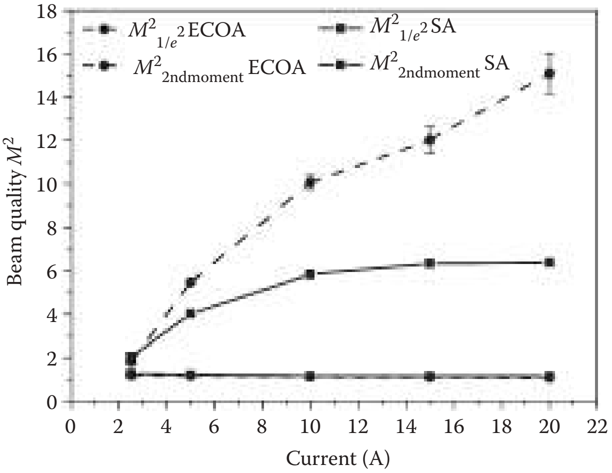


FIGURE 29.22 Dependence of M^2 on bias current for SA and external cavity on-axis simulations.

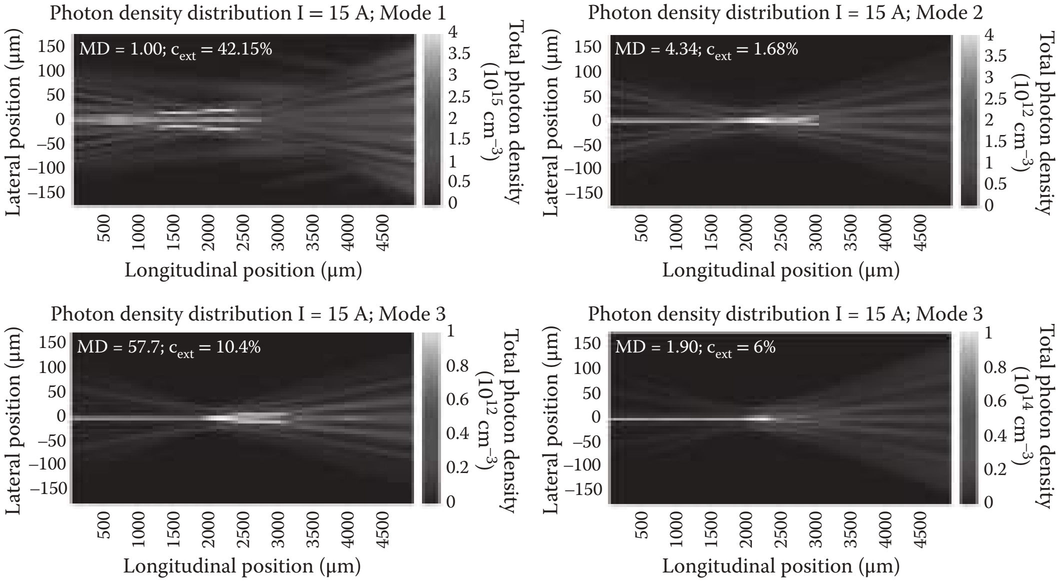


FIGURE 29.23 Total photon density. Mode 1 (top left). Mode 2 (top right). Mode 3 (bottom left). Mode 4 (bottom right). Modal discrimination (MD) and coupling coefficient (C_{ext}) are quoted for each mode.

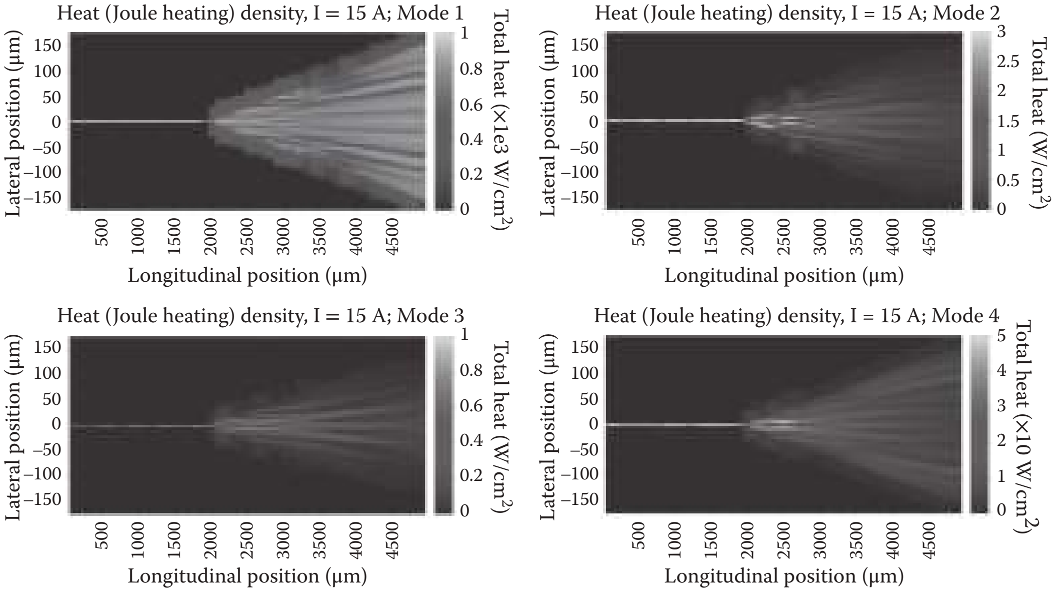


FIGURE 29.24 Self-heating due to Joule heating by each vertical mode at $I = 15$ A.

easy to overlook), they affect the laser through optical pumping (absorption bleaching), parasitic current leakage, and heat generation.

Figure 29.24 shows the Joule heating distributions caused by the currents supporting stimulated emission of the different modes at $I = 15$ A—as well as how the different vertical modes share/compete for the available gain. These distributions show where the stimulated emission occurs within the device for each of the vertical modes. For the higher order modes, this occurs solely because of the power fed back into

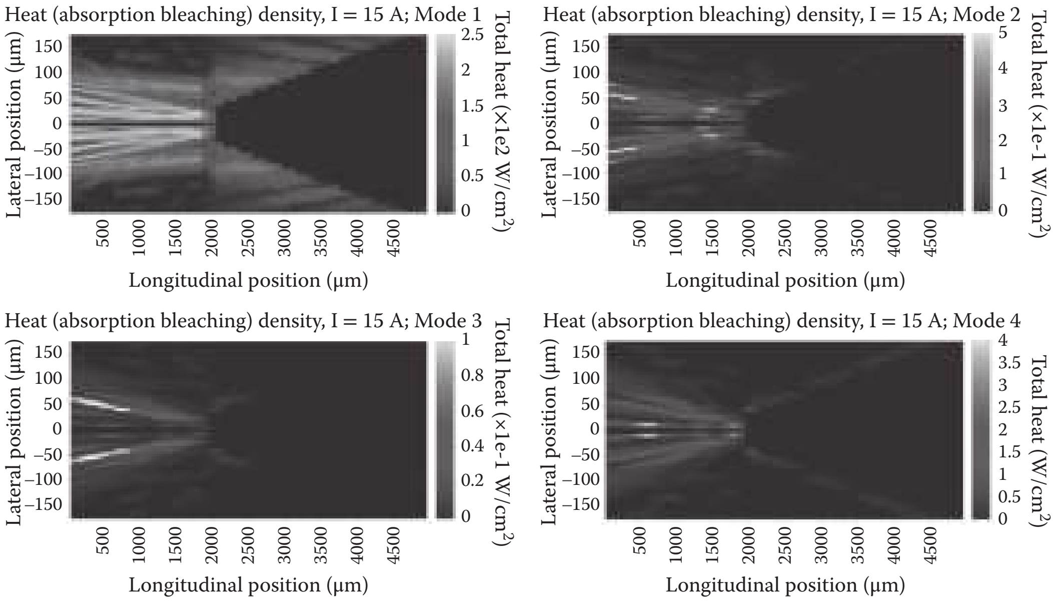


FIGURE 29.25 Self-heating due to optical pumping by each vertical mode at $I = 15$ A.

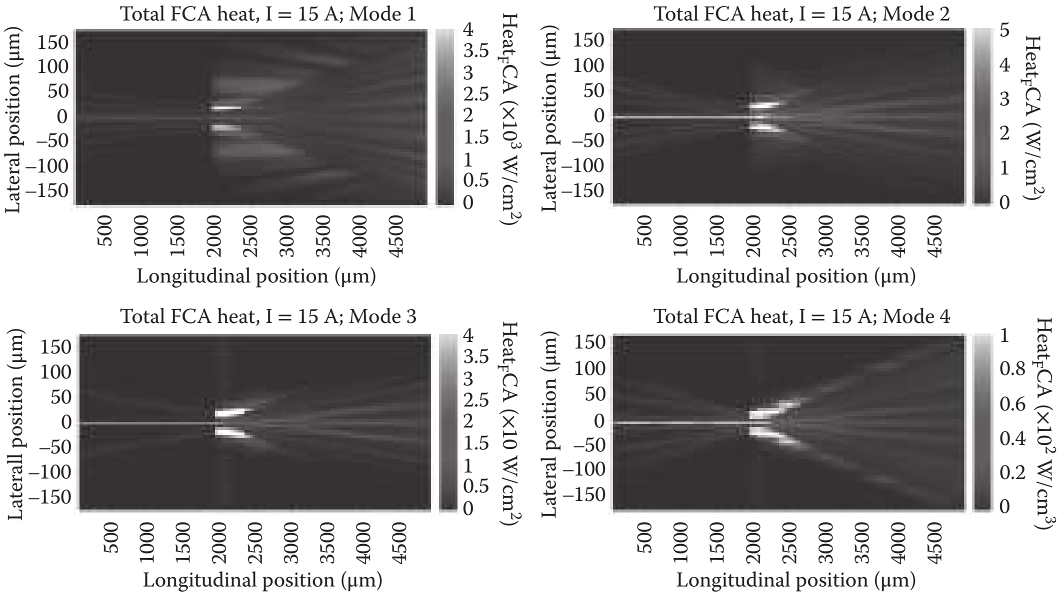


FIGURE 29.26 Self-heating due to free-carrier absorption by each vertical mode at $I = 15$ A.

them from the external cavity. Figure 29.25 shows the self-heating distributions due to optical pumping (absorption) at $I = 15$ A. This pumping causes the absorption bleaching, which plays a critical part in the degradation of the RW spatial filter performance. These distributions show the spatial distribution and relative contributions of the vertical modes to the absorption bleaching. The fundamental mode clearly plays a dominant role in the optical pumping, but the other modes also contribute. Figure 29.26 shows the spatial distributions of self-heating due to FCA. The drop in FCA to the left of the RW–TA interface is due to the RW etch. The higher order modes lose much more power due to FCA than to optical pumping.

TABLE 29.3 Impact of External Cavity Feedback on the Laser Performance (Power, Efficiency, and Beam Quality) for Coupling to All Modes and for Coupling Just to the Fundamental Mode

Simulation	Output Power (W)	PCE (%)	$M^2_{\text{2nd moment}}$
Stand-alone	11.67	47.13	6.32
External cavity (all modes)	9.49	38.39	12.04
External cavity (mode 1 only)	9.46	38.13	12.3

The excited higher order modes clearly play a role in the operation of the device, but their overall power is small compared to that of the fundamental mode (which also has the greatest feedback power). The question is, whether the degradation in PCE and beam quality is primarily controlled by feedback to the higher order vertical modes or to the fundamental mode. Table 29.3 compares the output power, PCE, and $M^2_{\text{2nd moment}}$ for the laser in isolation (SA), with on-axis feedback to all vertical modes and with on-axis feedback to just the fundamental mode. It is clear that the feedback to the fundamental mode has the greatest effect on both PCE and beam quality, while feedback to higher order vertical modes plays a smaller role. This can be understood in the context of an increased output facet reflectivity, which reduces the output power and PCE. At the same time, the optical feedback into the fundamental mode increases the absorption bleaching adjacent to the RW and also the power incident on the DBR reflector inside the RW, as discussed in Section 29.2.2. Both of these effects contribute to the beam quality degradation—the first by reducing the absorption filtering of the diffracted light outside the RW and the second by increasing the power in the side lobes of the diffraction from the DBR reflector.

29.4.2 Parasitic Reflections with Off-Axis Alignment

External feedback couples to the vertical modes differently, when its alignment to the waveguide changes. Here, we explore the impact of a vertical emitter shift of +0.5 μm due to "smile." Table 29.2 shows that modes 2 and 4 have moderate ($\text{MD}_2 = 4.33$) and low ($\text{MD}_4 = 1.90$) values of MD. Figure 29.19 shows they now have large (19%) and moderate (4.8%) coupling coefficients. The performances of these two modes should reveal the relative importance of the MD and the coupling coefficient. The coupling to fundamental mode also decreases from 42.15% (on-axis) to 29.5% (off-axis).

Figure 29.27 shows the power versus current and PCE versus current characteristics of the LOC tapered laser. With the off-axis feedback, the output power and PCE at $I = 15$ A dropped from 11.67 to 9.22 W, while the PCE fell from 47.1% to 37.3%. The coupling into mode 2 increased by a factor of ∼11, allowing both modes 2 and 4 to lase.

Figure 29.28 shows the current dependence of the beam propagation parameter M^2 (both the e^{-2} and second moment definitions) for the off-axis external cavity feedback condition. At $I = 15$ A, the off-axis feedback did not change $M^2_{e^{-2}}$, but $M^2_{\text{2nd moment}}$ increased from 6.32 to 11.72.

29.4.3 Analysis and Discussion

In this section, we analyze and discuss what happens to the power fed back into the higher order modes and how it affects the performance of the laser. We are particularly interested in how the feedback effects depend on parameters that can be influenced during the design of the laser: MD, coupling coefficient, and FCA. We are also interested in how the external cavity optics can be designed to minimize feedback effects.

Tables 29.4 and 29.5 summarize a range of power-related performance metrics for each of the vertical modes at a bias current of 15 A. This bias was selected, as it produces an output power of 11.7 W in the absence of external feedback and is within our desired range of 10–12 W. Table 29.4 is for the on-axis simulation. Table 29.5 is for the off-axis simulation when the emitter is vertically displaced by 0.5μm. Modes in boldface font are lasing. H_{stm} is the Joule heating power of the current needed to support the stimulated

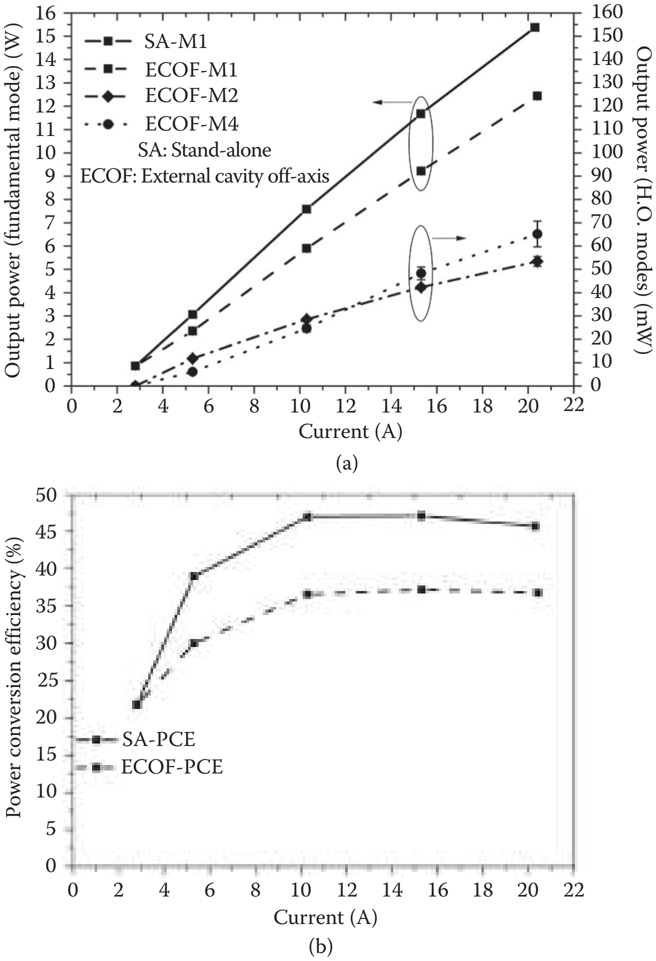


FIGURE 29.27 L–I curves of all lasing modes for tapered lasers in the SA and off-axis external cavity feedback (ECOF) configurations.

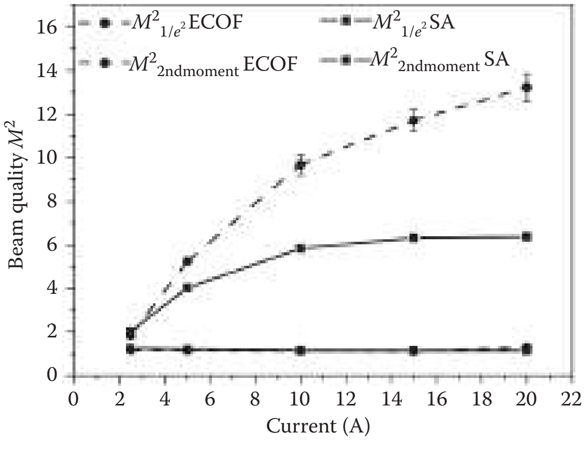


FIGURE 29.28 Dependence of M^2 on bias current for both SA and external cavity off-axis simulations.

TABLE 29.4 Summary of Impact of MD on Optical Metrics of All Modes for On-Axis Simulation for a Current of 15 A.

Mode	MD	Coupling Coefficient (%)	Coupled External Power (mW)	Output Power (W)	H_{stm} (W)	H_{abs} (W)	H_{FCA} (W)	I_{stm} (A)
1	**1.00**	**42.15**	**155.3**	**9.5**	**3.86**	**0.95**	**3.2**	**11.2**
2	4.34	1.68	6.2	3.1e-3	2.4e-3	3.14e-4	4.2e-3	7.1e-3
3	57.7	10.4	38.6	2.6e-3	5.8e-4	4.43e-5	27.7e-3	1.7e-3
4	**1.90**	**6**	**22.2**	**80.1e-3**	**47.7e-3**	**1.7e-3**	**124.6e-3**	**138.1e-3**
5	4.17	1.2	4.6	1e-3	1.5e-3	9.48e-5	10.9e-3	4.5e-3
6	1491.	2.2	8.4	4e-6	3.6e-6	4.83e-8	8e-3	7.6e-6
7	12.58	2.2	8.3	1.6e-4	4.6e-4	3.94e-6	11e-3	1.3e-3
8	1.92	0.65	2.4	5.9e-3	4.6e-3	2.43e-5	22.2e-3	13.4e-3
Total for higher order modes			90.7	92.9e-3	57.2e-3	3e-3	208.6e-3	166.1
Total for all modes			246.0	9.59	3.92	0.953	3.41	11.37

Note: Values in boldface font signify that the mode is lasing

TABLE 29.5 Summary of Impact of MD on Optical Metrics of All Modes for Off-Axis Simulation for a Current of 15 A.

Mode	MD	Coupling Coefficient (%)	Coupled External Power (mW)	Output Power (W)	H_{stm} (W)	H_{abs} (W)	H_{FCA} (W)	I_{stm} (A)
1	**1.00**	**29.5**	**142.6**	**9.22**	**3.71**	**0.96**	**3.03**	**10.96**
2	**4.34**	**19**	**92**	**42.3e-3**	**26.5e-3**	**2.7e-3**	**38.5e-3**	**76.6e-3**
3	57.7	5.23	25.28	1.5e-3	4.13e-4	3.81e-5	13.3e-3	1.2e-3
4	**1.90**	**4.8**	**23.2**	**48.2e-3**	**38.4e-3**	**4.7e-3**	**81.4e-3**	**111.2e-3**
5	4.17	1.73	8.4	1.4e-4	2.5e-3	2.62e-4	17.1e-3	7.1e-3
6	1491.	3.28	15.88	1.94e-6	4.12e-6	1.08e-7	13.3e-3	1.2e-5
7	12.58	1.44	6.97	5e-5	4.58e-4	1.08e-5	8.8e-3	1.3e-3
8	1.92	0.83	4	1.8e-3	4.6e-3	4.35e-4	31.8e-3	13.4e-3
Total for higher order modes			175.7	94.0e-3	72.9e-3	13e-3	204.2e-3	210.8
Total for all modes			318.3	9.31	3.78	0.973	3.23	11.17

Note: Values in boldface font signify that the mode is lasing

emission of the mode, I_{stm}. H_{abs} is the total heat power due to band-to-band absorption outside the gain regions and is responsible for absorption bleaching. H_{FCA} is the total heat power due to FCA. For the case of on-axis feedback, 91 mW of light is coupled into the higher order modes, which consume 166 mA of current and produce 93 mW of light and 269 mW of heat. For the case of off-axis feedback, 176 mW of light is coupled into the higher order modes, which consume 211 mA of current and produce 94 mW of light and 290 mW of heat.

The previous section showed that external feedback can cause higher order vertical modes to lase. For the laser diode and external cavity simulated here, only the fundamental mode lases without feedback, while mode 4 also lases in the case with on-axis feedback. For the case of off-axis feedback, two higher order vertical modes also lase (modes 2 and 4). The feedback reduces the power and PCE of the fundamental mode, which can be attributed partly to higher order mode lasing and partly to an increase in the effective facet reflectivity by feedback to the fundamental mode. For the device and cavity investigated here, Tables 29.4 and 29.5 suggest that modes with MD of MD < 2 are able to start lasing with comparatively small external cavity coupling coefficients of <5%, while modes with MD ~ 4 require larger coupling coefficients (e.g., 10%–20%). Additional simulations showed that the feedback required for higher order mode lasing also depends on the coupling of the reflected beam into the RW (i.e., into the RW aperture at the RW–TA interface). This is also apparent from the photon distributions in Figure 29.23.

As higher order modes began to lase more strongly, they began to compete with the fundamental mode for the spatial gain distribution and modal power oscillations were observed in the simulations. Although these are not true dynamic simulations, the oscillations have some relation to mode beating and mode partition noise. Despite that the higher order mode powers were small, they appear to affect spatial filamentation and contribute to the near- and far-field patterns of the fundamental mode—consistent with experimental observations (Hempel et al., 2013; Leonhäuser et al., 2014). For stable operation, $M^2_{e^{-2}}$ remains below 1.5 and is insensitive to feedback, while $M^2_{\text{2nd moment}} \sim 6$ for SA operation and nearly doubles with feedback (As discussed in Section 29.4.1, increased feedback to the fundamental mode has the greatest impact on beam quality, but excitation of higher order modes also affects $M^2_{\text{2nd moment}}$). For astable operation (modal power oscillations), there is significant degradation in both $M^2_{e^{-2}}$ and $M^2_{\text{2nd moment}}$.

The optical power back-coupled into the higher order vertical modes also creates heat. This occurs even if the modes are not lasing, but increases dramatically if they are. The heat power of the higher order modes is dominated FCA, followed by Joule heating associated with the current supporting stimulated emission into these modes. The heat generated per feedback photon decreases with MD. This shows the importance of both MD and FCA for suppressing feedback amplification and thus, heating by the higher order modes. At this bias level (15 A), the total heat generation in the higher order modes is 0.269–0.290 W (depending on the alignment of the laser diode), corresponding to a 3.2%–3.6% increase in heat power. The heat generation by the higher order modes alone is responsible for a ~2.3%–2.4% drop in PCE, which increases to ~2.5% if their optical output power is also considered. (This is ~20%–25% of the total drop in PCE.) The heat generated by the excitation of these higher order vertical modes is distributed along the entire cavity—irrespective of the MD. Thus, although the excitation of higher order modes affects the beam quality, it is probably not responsible for observed increases in device degradation. Instead, as shown by Kissel et al. (2016), the observed increase in degradation is probably due to feedback that is *not* coupled into the waveguide (i.e., into the substrate)—particularly when the substrate is strongly absorbing (i.e., $E_{\text{g\_substrate}} < E_{\text{photon}}$).

We have shown that external feedback can cause lasing of higher order vertical modes, resulting in self-heating, mode partition instabilities, and beam quality degradation. The question is *How can we design the laser diode and optical system to reduce the impact of optical feedback?* For the design of the laser diode, MD and FCA are key parameters for suppressing the impact of higher order modes. The external cavity coupling coefficients play a smaller role, but are still important. The stray light from aperture diffraction by the DBR at the back of the RW must be suppressed. For the design of the external cavity, the reflected power must be minimized—for example, by using good AR coatings and, perhaps, tilted facets to eliminate specular reflections. For high powers, however, other approaches may also be required, such as off-axis optical alignment to prevent reflected beams from coupling into the RW filter. Finally, care must also be taken to minimize the power coupled into the substrate—particularly if the substrate is absorbing.

29.5 Conclusions

This chapter discussed how, for many applications that employ high-brightness diode lasers, it is necessary to scale the output power through beam combination techniques, while maintaining an excellent beam quality. All beam combination techniques cause light to be fed back into the laser cavity. In some instances, this feedback can be exploited, such as in the wavelength stabilization of diode lasers and arrays. However, any laser placed in an external system will also suffer from unintentional feedback (parasitic reflections). This chapter describes a detailed method of modeling external cavity lasers with both intentional and unintentional feedback. This model, developed at the University of Nottingham, consists of an advanced laser simulation tool coupled to commercial optical design software.

In large vertical cavity lasers, as required to reach the highest output powers, the vertical cavity must be carefully considered. The structure must be designed to minimize the coupling of parasitic feedback to higher order modes. The effects of parasitic feedback on the performance (output power, PCE, and beam quality) of a tapered laser in a simple optical system are studied. The excitation and propagation of higher order modes increases the FCA causing self-heating, eventually causing degradation of the output power and efficiency. Moreover, the parasitic feedback causes more absorption bleaching, leading also to degradation of the lateral beam quality. Therefore, the inclusion of external parasitic feedback in the design process is highly significant. Coupling coefficients and MD values can be used to engineer a laser structure that is less affected by feedback. This can be accomplished by ensuring that modes with low-to-moderate MD values have the lowest coupling coefficients, thereby minimizing the light coupled into them.

The lateral cavity design also has a significant impact on brightness and the tapered laser is a strong contender for high-brightness laser systems. Tapered lasers offer, to a certain extent, the beam quality advantages of an RW laser and the high power advantages of a BA laser. However, despite this and even without external parasitic feedback, tapered lasers still suffer from beam quality degradation at higher powers. Investigations into the cause of this beam quality degradation reveal that diffraction of light at the back aperture leads to regions adjacent to the RW section becoming bleached from carriers—thereby, degrading the beam quality.

Acknowledgments

The authors thank J. Decker and P. Crump at the Ferdinand Braun Institute and N. Michel and M. Krakowski at Alcatel-Thales III–V Lab for experimental results and helpful discussions, and U. Witte and M. Traub at Fraunhofer Institute for Laser Technology for helpful discussions on optical systems for direct-diode lasers. We also thank G. Schimmel, G. Lucas-Leclin, and P. Georges at CNRS Institute d'Optique; and V. Vilokkinen and P. Uusimaa at Modulight Ltd. for helpful discussions. The authors acknowledge L. Borruel and I. Esquivias for their contributions to early collaborations on the simulation of tapered laser diodes during the EC ULTRABRIGHT project. The authors gratefully acknowledge funding from the European Commission projects: BRIDLE (IST–314719, 2012–15), WWW.BRIGHTer.EU (IST–2005–035266, 2006–10), FAST ACCESS (IST-004772, 2004–7 WWW.BRIGHT.EU (IST-511722; 2004–6), POWERPACK (IST-2000-29447, 2001–3), ULTRABRIGHT (IST-1999-10356, 2000–3).

References

Agrawal GP (1984) Fast-Fourier-transform based beam-propagation model for stripe-geometry semiconductor lasers: Inclusion of axial effects. *Journal of Applied Physics* 56:3100–3109.

Buda M, van der Vleuten WC, Iordache G, Acket GA, van de Roer TG, van Es CM, van Roy BH, Smalbrugge E (1999) Low-loss low-confinement GaAs-AlGaAs DQW laser diode with optical trap layer for high-power operation. *IEEE Photonics Technology Letters* 11:161–163.

Bull S, Andrianov A, Wykes JG, Lim JJ, Sujecki S, Auzanneau SC, Calligaro M, Lecomte M, Parillaud O, Krakowski M, Larkins EC (2006) Quantitative imaging of intracavity spontaneous emission distributions using tapered lasers fabricated with windowed n-contacts. *IEEE Proceedings—Optoelectronics* 153:2–7.

Bull S, Wykes JG, Andrianov AV, Lim JJ, Borruel L, Sujecki S, Auzanneau SC, Calligaro M, Krakowski M, Esquivias I, Larkins EC (2004) Imaging of spontaneous emission from 980 nm tapered lasers with windowed N-contacts. *European Physical Journal Applied Physics* 27:455–459.

Crump P, Blume G, Paschke K, Staske R, Pietrzak A, Zeimer U, Einfeldt S, Ginolas A, Bugge F, Häusler K, Ressel P, Wenzel H, Erbert G (2009) 20 W continuous wave reliable operation of 980 nm broad-area single emitter diode lasers with an aperture of 96 μm. *Proceedings of SPIE* 7198:719814–719819.

Crump P, Erbert G, Wenzel H, Frevert C, Schultz CM, Hasler KH, Staske R, Sumpf B, Maaβdorf A, Bugge F, Knigge S, Tränkle G (2013a) Efficient high-power laser diodes. *IEEE Journal of Selected Topics in Quantum Electronics* 19:1501211.

Crump P, Hasler KH, Wenzel H, Knigge S, Bugge F, Erbert G (2013b) High efficiency, 8 W narrow-stripe broad-area lasers with in-plane beam-parameter-product below 2 mm mrad. In: *2013 Conference on Lasers & Electro-Optics Europe & International Quantum Electronics Conference CLEO EUROPE/IQEC*, Munich, 2013, p. 1.

Crump P, Knigge S, Maaßdorf A, Bugge F, Hengesbach S, Witte U, Hoffmann H-D, Köhler B, Hubrich R, Kissel H, Biesenbach J, Erbert G, Traenkle G (2013c) Low-loss smile-insensitive external frequency-stabilization of high power diode lasers enabled by vertical designs with extremely low divergence angle and high efficiency. *Proceedings of SPIE* 8605:86050T–86013.

Decker J, Crump P, Fricke J, Maaβdorf A, Erbert G, Tränkle G (2014) Narrow stripe broad area lasers with high order distributed feedback surface gratings. *IEEE Photonics Technology Letters* 26:829–832.

Dittmar F, Sumpf B, Fricke J, Erbert G, Trankle G (2006) High-power 808-nm tapered diode lasers with nearly diffraction-limited beam quality of M/sup 2/=1.9 at P = 4.4 W. *IEEE Photonics Technology Letters* 18:601–603.

Donnelly JP, Huang RK, Walpole JN, Missaggia LJ, Harris CT, Plant JJ, Bailey RJ, Mull DE, Goodhue WD, Turner GW (2003) AlGaAs-InGaAs slab-coupled optical waveguide lasers. *IEEE Journal of Quantum Electronics* 39:289–298.

Fiebig C, Blume G, Kaspari C, Feise D, Fricke J, Matalla M, John W, Wenzel H, Paschke K, Erbert G (2008) 12 W high-brightness single-frequency DBR tapered diode laser. *Electronics Letters* 44:1253–1255.

Hempel M, Chi M, Petersen PM, Zeimer U, Tomm JW (2013) How does external feedback cause AlGaAs-based diode lasers to degrade? *Applied Physics Letters* 102:023502–023504.

Huang RK, Chann B, Glenn JD (2011) Extremely high-brightness kW-class fiber coupled diode lasers with wavelength stabilization. *Proceedings of SPIE* 8039:80390N–80310.

Huang RK, Chann B, Missaggia LJ, Augst SJ, Connors MK, Turner GW, Sanchez-Rubio A, Donnelly JP, Hostetler JL, Miester C, Dorsch F (2009) Coherent combination of slab-coupled optical waveguide lasers. *Proceedings of SPIE—the International Society for Optical Engineering* 7230:72301G–72312.

Huang RK, Donnelly JP, Missaggia LJ, Harris CT, Chann B, Goyal AK, Sanchez-Rubio A, Fan TY, Turner GW (2007) High-brightness slab-coupled optical waveguide lasers. *Proceedings of SPIE* 6485:64850F–64859.

Kaunga-nyirenda SN, Bull S, Lim JJ, Hasler KH, Fricke J, Larkins EC (2014) Factors influencing brightness and beam quality of conventional and distributed Bragg reflector tapered laser diodes in absence of self-heating. *IET Optoelectronics* 8:99–107.

Kelemen MT, Weber J, Kaufel G, Bihlmann G, Moritz R, Mikulla M, Weimann G (2005) Tapered diode lasers at 976 nm with 8 W nearly diffraction limited output power. *Electronics Letters* 41:1011–1013.

Kissel H, Leonhäuser B, Tomm JW, Hempel M, Biesenbach J (2016) Investigation of accelerated and catastrophic degradation of laser diodes caused by external optical feedback operation. European Semiconductor Laser Workshop, Darmstadt, Germany.

Knigge A, Erbert G, Jonsson J, Pittroff W, Staske R, Sumpf B, Weyers M, Trankle G (2005) Passively cooled 940 nm laser bars with 73% wall-plug efficiency at 70 W and 25/spl deg/C. *Electronics Letters* 41:250–251.

Lang L, Lim JJ, Sujecki S, Larkins EC (2008) Improvement of the beam quality of a broad-area diode laser using asymmetric feedback from an external cavity. *Optical Quantum Electronics* 40:1097–1102.

Larkins EC, Bull S, Kaunga-Nyirenda S, Helal MA, Vilokkinen V, Uusimaa P, Crump P, Erbert G (2014) Design optimisation of high-brightness laser diodes for external cavity operation in the BRIDLE Project. In: *2014 International Semiconductor Laser Conference*, Palma de Mallorca, 2014, pp. 21–22.

Larkins EC, Helal MA, Kaunga-Nyirenda SN, Bull S, Moss D (2016) Design and simulation of high-brightness diode lasers for operation in the presence of external feedback. In: Conference on Novel In-Plane Semiconductor Lasers XV, 2016. (Photonics West, San Francisco, USA) 9767

Leonhäuser B, Kissel H, Unger A, Köhler B, Biesenbach J (2014) Feedback-induced catastrophic optical mirror damage (COMD) on 976 nm broad area single emitters with different AR reflectivity. *Proceedings of SPIE* 8965:896506–896510.

Lim JJ (2003) Investigation of factors influencing the brightness of high-power laser diodes. PhD dissertation, University of Nottingham.

Lim JJ, Benson TM, Larkins EC (2005) Design of wide-emitter single-mode laser diodes. *IEEE Journal of Quantum Electronics* 41:506–516.

Lim JJ, Bull S, Kaunga-Nyirenda S, Sujecki S, Larkins EC, Hasler KH, Fricke J (2012) Factors influencing the brightness and beam quality of tapered laser diodes and bars. IEEE Summer Topical Meeting on High Power Semiconductor Lasers TuA4.1.

Lim JJ, Sujecki S, Lei L, Zhichao Z, Paboeuf D, Pauliat G, Lucas-Leclin G, Georges P, MacKenzie R, Bream P, Bull S, Hasler KH, Sumpf B, Wenzel H, Erbert G, Thestrup B, Petersen PM, Michel N, Krakowski M, Larkins EC (2009) Design and simulation of next-generation high-power, high-brightness laser diodes. *IEEE Journal of Selected Topics in Quantum Electronics* 15:993–1008.

Pearson JE, McGill TC, Kurtin S, Yariv A (1969) Diffraction of gaussian laser beams by a semi-infinite plane. *Journal of the Optical Society of America* 59:1440–1445.

Pillai RMR, Garmire EM (1996) Paraxial-misalignment insensitive external-cavity semiconductor-laser array emitting near-diffraction limited single-lobed beam. *IEEE Journal of Quantum Electronics* 32:996–1008.

Skidmore J, Peters M, Rossin V, Guo J, Xiao Y, Cheng J, Shieh A, Srinivasan R, Singh J, Wei C, Duesterberg R, Morehead JJ, Zucker E (2016) Advances in high-power 9XXnm laser diodes for pumping fiber lasers. *Proceedings of SPIE* 9733:97330B–97337.

Sujecki S, Borruel L, Wykes J, Moreno P, Sumpf B, Sewell P, Wenzel H, Benson TM, Erbert G, Esquivias I, Larkins EC (2003) Nonlinear properties of tapered laser cavities. *IEEE Journal of Selected Topics in Quantum Electronics* 9:823–834.

Sumpf B, Adamiec P, Fricke J, Ressel P, Wenzel H, Erbert G, Tränkle G (2010) Comparison of 650 nm tapered lasers with different lateral geometries at output powers up to 1 W. *Proceedings of SPIE* 7616:76161–76168.

Sverdlov B, Pfeiffer HU, Zibik E, Mohrdiek S, Pliska T, Agresti M, Lichtenstein N (2013) Optimization of fiber coupling in ultra-high power pump modules at $\lambda = 980$ nm. *Proceedings of SPIE* 8605:860508–860510.

Thestrup B, Chi M, Sass B, Petersen PM (2003) High brightness laser source based on polarization coupling of two diode lasers with asymmetric feedback. *Applied Physics Letters* 82:680–682.

Tomm JW, Ziegler M, Hempel M, Elsaesser T (2011) Mechanisms and fast kinetics of the catastrophic optical damage (COD) in GaAs-based diode lasers. *Laser & Photonics Reviews* 5:422–441.

Walpole JN, Kintzer ES, Chinn SR, Wang CA, Missaggia LJ (1992) High-power strained-layer InGaAs/AlGaAs tapered traveling wave amplifier. *Applied Physics Letters* 61:740–742.

Wang X, Erbert G, Wenzel H, Crump P, Eppich B, Knigge S, Ressel P, Ginolas A, Maaßdorf A, Tränkle G (2013) High power, high beam quality laser source with narrow, stable spectra based on truncated-tapered semiconductor amplifier. *Proceedings of SPIE* 8605:86050G–86011.

Williams KA, Penty RV, White IH, Robbins DJ, Wilson FJ, Lewandowski JJ, Nayar BK (1999) Design of high-brightness tapered laser arrays. *IEEE Journal of Selected Topics in Quantum Electronics* 5:822–831.

Wolff S, Rodionov A, Sherstobitov VE, Fouckhardt H (2003) Fourier-optical transverse mode selection in external-cavity broad-area lasers: Experimental and numerical results. *IEEE Journal of Quantum Electronics* 39:448–458.

Yaeli J, Streifer W, Scifres DR, Cross PS, Thornton RL, Burnham RD (1985) Array mode selection utilizing an external cavity configuration. *Applied Physics Letters* 47:89–91.

Yang G, Smith GM, Davis MK, Loeber DAS, Hu M, Chung-en Z, Bhat R (2004) Highly reliable high-power 980-nm pump laser. *IEEE Photonics Technology Letters* 16:2403–2405.

30

Single Longitudinal Mode Laser Diodes

30.1 Introduction.. 109
What Is a Single Longitudinal Mode Laser? • Why SLM Laser? • Current Application Status • Why Simple FP Laser Won't Work • Classification of SLM Lasers • Organization of this Chapter
30.2 Grating-Assisted SLM Laser Diodes 113
Grating Analysis • Distributed Bragg Reflector Laser • Distributed Feedback Lasers • Governing Equations for Design, Modeling, and Simulation • Examples of SLM Laser Characteristics
30.3 Coupled-Cavity Single Longitudinal Mode Laser Diodes........ 133
A General Optical (Band-Pass) Filtering Model • Cleaved-Coupled-Cavity (C^3) Laser • Etched Slotted Laser • Discrete Mode Laser • New Aspects in Governing Equations and the Solution Technique • Performance Comparisons among Coupled-Cavity SLM Lasers
30.4 Recent Development on Single Longitudinal Mode Laser Diode.. 141
Open Problems in SLM Lasers • Bragg Waveguide SLM Laser • Double-Trench Resonant Tunneling SLM Laser • Horn Ridge Waveguide (HRW) DFB Laser

Xun Li

30.1 Introduction

30.1.1 What Is a Single Longitudinal Mode Laser?

Laser is a light source that emits photons described by a "near coherent" optical field. By "coherent optical field," one means the monochromatic electromagnetic field at optical frequency (ω_0) with constant amplitude (A) and phase (θ). Therefore, a coherent optical field can be expressed as $A\cos(\omega_0 t + \theta)$, with phase θ determined by a reference starting time. Any laser built as an oscillator must take spontaneously emitted photons as its initial driven seed, since a laser doesn't have any coherent light as its input. Knowing the fact that the spontaneous emission is a random process, one therefore cannot expect that the laser will give an ideal coherent optical field output. Rather, the laser emits photons described by the ideal coherent optical field, driven by spontaneously emitted photons, for a certain amount of time τ_0 until the emerging of another group of spontaneously emitted photons. Consequently, its output field will experience a sudden change on its amplitude and phase at time τ_0 and this sequence keeps repeating indefinitely. As such, one can express the laser output as a "quasi-monochromatic" electromagnetic field at optical frequency ω_0 as $[A + \delta_a(t)]\cos[\omega_0 t + \theta(t)]$, with $\delta_a(t)$ as a random process with zero mean and $\theta(t)$ also a random process with uniform distribution between 0 and 2π. Once a laser is operated under a bias beyond its threshold, $A >> |\delta_a(t)|$ is satisfied. Hence the only nonideality of a laser output from a coherent optical field lies in its random phase. The difference becomes more apparent if one observes the output optical spectrum of a laser. Actually, as the Fourier transform of an ideal coherent optical field, its frequency-domain spectrum is a delta function that appears at ω_0 with an amplitude of $|A|$. The quasi-monochromatic field, however, has its averaged spectrum in the shape of a "broadened" delta function still peaked at ω_0, but with an amplitude in $\sqrt{1/2\pi}|A|\tau_0$ and a width of $\Delta\omega = 2\pi/\tau_0$. This is because the averaged duration between two consecutive abrupt phase changes is τ_0, which means the quasi-monochromatic field can be described as a series of truncated ideal coherent fields known as a wave train, with τ_0 as the truncation window or the averaged length of the wave train. The Fourier transform of a single truncated coherent field piece corresponds to the convolution between a delta function and a sampling function in the form of $\sin(\omega\tau_0/2)/(\omega\tau_0/2)$ as the Fourier transforms of the ideal coherent field and the flat window function (i.e., 1 for t inside τ_0 and 0 elsewhere), respectively, which gives the result as the aforementioned broadened delta function. Since distinct truncated coherent field pieces in the wave train differ by a time shift only, their Fourier transforms differ just by a phase. Hence, the frequency-domain amplitude spectra of the quasi-monochromatic field as a wave train composed of all these pieces are overlapped as a single peak as shown in Figure 30.1.

It is worth mentioning that the above description is in a "phenomenological" sense, as an accurate treatment of spontaneously emitted photons with a classical electromagnetic field theory is not possible. A physics-based description can be given by exploiting the full quantum mechanics theory, but it will have to involve the quantization of the electromagnetic field and is beyond the scope of this chapter.

The optical frequency ω_0 of the quasi-monochromatic field as the laser output is called the "lasing" frequency. Consequently, the aforementioned spectral peak is called the "lasing" mode or the "longitudinal" mode in order to make a distinction from the "transverse" mode that refers to a spatial field distribution. The term longitudinal comes from the fact that the lasing frequency is determined by the laser resonant cavity. In edge-emitting semiconductor laser diodes, the cavity is set along the longitudinal direction with its geometrical dimension significantly larger than any size in the cross-sectional area. The light is confined inside the cross-sectional area by a waveguide and is only allowed to propagate along the longitudinal direction, or along the cavity. Upon resonance, the lasting longitudinal optical field distribution gives the surviving pattern and determines the lasing frequency. Hence, we have the term longitudinal mode, as opposed to the transverse mode, indicating the confined optical field by the waveguide in the cross-sectional area. Because of the one-to-one correspondence between the optical spectral peak and the lasting longitudinal optical field distribution, conventionally the term longitudinal mode can be referred to either as the spectral peak (or simply as the lasing frequency) or as the spatial field pattern in the cavity

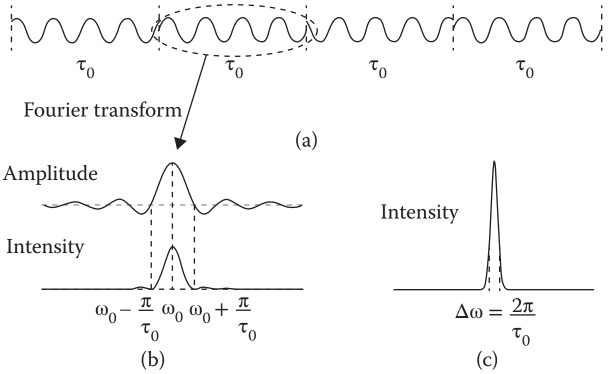

τ_0
τ_0
τ_0
τ_0
Fourier transform
(a)
Amplitude
Intensity
$\omega_0 - \frac{\pi}{\tau_0}$
ω_0
$\omega_0 + \frac{\pi}{\tau_0}$
(b)
Intensity
$\Delta\omega = \frac{2\pi}{\tau_0}$
(c)

FIGURE 30.1 (a) The quasi-monochromatic field as a wave train in the time domain and a single piece of the truncated coherent field, (b) the frequency-domain amplitude and intensity spectra as the Fourier transform of the wave train (with different pieces all overlapped), and (c) the intensity spectrum with the peak indicating a single (spectral) longitudinal mode.

(longitudinal) direction, whereas the term transverse mode can only be referred to as the spatial field pattern in the cross-sectional area, without any spectral meaning.

Generally, a cavity can support multiple resonant patterns and consequently bear with multiple resonant frequencies. Without exception, a laser can also have multiple longitudinal modes, with multiple peaks shown in its output optical spectrum and multiple lasting optical field patterns inside its cavity. With special designs, however, one is able to leave only one surviving longitudinal mode by eliminating all other resonances inside a cavity. A laser with such a cavity is therefore called a single longitudinal mode (SLM) laser.

30.1.2 Why SLM Laser?

When used as the light source in fiber-optic communication systems, a semiconductor laser operated with multiple longitudinal modes suffers the mode partition noise (MPN) [1,2], which stands as the dominant limiting factor to the transmission span of the optical signal in fiber, as the MPN jeopardizes the signal by introducing the intersymbol interference (ISI) among the pulses in the stream—an effect that cannot be simply suppressed through increasing the laser output power. If a multiple longitudinal mode laser is directly modulated, the power allocated to each of its mode fluctuates in a random manner due to the mixed homogeneous and inhomogeneous gain broadening nature of direct bandgapped semiconductors, with the randomness originated from the spontaneous emission noise. Since the signal components carried by different longitudinal modes propagate at different speed due to fiber dispersion, these components won't arrive at the destination at the same time, which results in the pulse spreading over and spilling out of its allocated time slot, and consequently causes ISI. Such ISI bears a random nature due to the random power fluctuation of the multiple longitudinal modes as the signal carriers. Hence, it cannot be eliminated through linear equalization or phase delay compensation. Also, because the random fluctuation is in proportion to the total power, increasing the laser output power won't solve the problem, if doesn't make it worse. Normally, the power penalty soars even starting from a moderate MPN level. For example, at ∼1550 nm (the center of the C-band), the maximum transmission capacity–distance product is only 5 Gbps-km, which means by using a typical multiple longitudinal mode Fabry–Pérot (FP) laser, the 2.5 and 10 Gbps optical signal can only be transmitted for 2 km and 500 m, respectively.

The only viable solution to this problem, therefore, lies in the replacement of the light source with the SLM laser where the MPN naturally disappears.

30.1.3 Current Application Status

Over the past three decades, many SLM laser structures have been proposed and demonstrated; a few dominant structures survived and became popular products on today's market. The super star of the SLM laser is no doubt the distributed feedback (DFB) laser as it takes over 99% shares of the SLM laser market—the throughput of DFB lasers have reached 10 million per year in the recent few years just by one supplier (Wuhan Telecommunication Device Co., Wuhan, China) among the few largest semiconductor laser manufactures.

30.1.4 Why Simple FP Laser Won't Work

It is well known that a piece of straight and smooth optical waveguide with its two mirror-like ends forms a simplest resonant cavity, and semiconductor laser diodes exploiting such a cavity are the FP lasers. A semiconductor optical waveguide, like the typical dielectric waveguide, has the high-pass filter characteristic that cuts off the guiding wave with wavelength longer than some critical value but imposes no constraint on the guiding waves with shorter wavelengths. Although such waveguide does only support discrete space distribution patterns in its cross section and each individual space distribution pattern is also called a transverse mode, one shouldn't mess up it (the transverse mode) with the concept of the longitudinal mode that corresponds a specific field distribution along the cavity and bears a discrete wavelength. Namely, a (dielectric) waveguide picks up the guided transverse mode, but doesn't select a single wavelength from a continuous spectrum of wavelengths that all takes the same transverse mode, except for cutting off the wavelengths below a lower bound. One can also understand it as a given waveguide structure only defines a continuous dispersion relation between the wave propagation constant (β) and the wavelength (λ), which means that, staying with a transverse mode represented by a continuous β–λ curve, one will always manage to find a β indicating a guiding (propagating along the waveguide) wave for a given wavelength λ. Nevertheless, an optical waveguide that only supports a single transverse mode stands as a necessary condition for achieving the SLM operation. As otherwise, each of the multiple nondegenerated transverse modes having its own nonidentical longitudinal mode will have to bear a different wavelength, which makes the SLM operation impossible. We are not going to dig out the waveguide concept further as the focus of this chapter is on the SLM laser. The interested readers can refer to, e.g., References [3,4].

With mirrors on the waveguide ends, the otherwise not interfered forward and backward propagating waves are coupled through the partial reflection. The waves retained inside the waveguide (i.e., the cavity) forms a standing wave pattern that corresponds to a possible longitudinal mode, as the standing wave pattern (other than a homogenous coefficient) doesn't change with time by satisfying the boundary condition at the waveguide ends and consequently bears with a static wavelength. However, in the case where the cavity length (L) is much longer than the wavelength (λ), the boundary condition can be satisfied simultaneously by multiple longitudinal modes. This is because the boundary condition of a simple FP cavity can be equivalent to a round-trip phase condition on aggregate, which reads as follows [5]:

$$2\beta L = 2\pi m \tag{30.1}$$

with m as any integer, $\beta = 2\pi n_{\text{eff}}/\lambda$ denoting the wave propagation constant and n_{eff} the effective index of the waveguide. In Equation 30.1, the facet (end mirror) phase delay has been set to zero as usually the cavity has an effective index higher than that of the surrounding medium, so that the wave experiences the internal (from high refractive index to low refractive index medium) rather than the external (from low refractive index to high refractive index medium) reflection in which the phase delay is indeed zero on reflection. From Equation 30.1, it is apparent that the number of allowed longitudinal modes in an FP cavity is directly proportional to the cavity length scaled by the wavelength, i.e., L/λ. To build an edge-emitting laser with acceptable coherence required by many applications, as well as for obtaining sufficient optical gain for achieving superior laser performance, one has to leave $L \gg \lambda$, hence multiple longitudinal modes exist in such FP cavities.

Other than the aforementioned phase condition, the amplitude condition still needs to be satisfied by the longitudinal mode that survived the phase condition to make it lase. One can therefore consider using the latter mechanism to eliminate the extra longitudinal modes. Unfortunately, the optical gain spectrum of the semiconductor material is much broader (~60 nm) than the FP longitudinal mode spacing that can readily be derived from Equation 30.1 as follows [5]:

$$\Delta\lambda = \frac{\lambda_0^2}{2n_g L} \tag{30.2}$$

with λ_0 as the center wavelength and $n_g = n_{\text{eff}} + (dn_{\text{eff}}/d\lambda)\lambda_0$ denoting the group index. For example, for typical InGaAsP/InP or AlGaInAs/InP semiconductor laser diodes, $n_g \sim 3.4$, their mode spacing vary from 0.6 to 1.2 nm in the O-band ($\lambda_0 \sim 1300$ nm), and from 0.9 to 1.8 nm in the C-band ($\lambda_0 \sim 1550$ nm), respectively, when the cavity length changes from 400 to 200 μm. As such, multiple longitudinal modes with their wavelengths closely packed adjacent to the material gain peak will share almost the same gain. Since the partially inhomogeneous gain in a semiconductor material cannot be fully clamped, small gain differences among these longitudinal modes won't be sufficient to effectively suppress the side modes, which eventually excites multiple longitudinal lasing modes. This is the main reason that the material gain dispersion in general cannot be used for longitudinal mode selection, not to mention that in the case of direct modulation, any dynamic gain change will disrupt the gain profile and consequently jeopardize any attempt of using the gain to discriminate longitudinal modes.

One therefore understands why conventional semiconductor FP laser diodes operate in multiple longitudinal modes. Above analysis also shed some light on the direction one needs to follow in the effort of developing SLM lasers, i.e., the simple FP cavity needs to be replaced by more complicated ones or modified with added structures with further built-in wavelength selectivity.

30.1.5 Classification of SLM Lasers

Canonically, there are three main approaches for semiconductor laser diodes to achieve SLM operation, all by introducing wavelength-selective cavities. These three categories of SLM lasers can be classified as the grating-assisted lasers that exploit the grating in their cavities, the coupled-cavity lasers that have extra optical (band-pass) filters in their cavities, and the external cavity lasers that use either one of the above-mentioned configurations, or the injection locking mechanism to purify the lasing spectrum but by placing the extra components outside of the main FP cavity through hybrid packaging. Theoretically, lasers in the latter category makes no difference from their counterparts in the first two categories, other than some quantitative difference on the coupling strength between the main FP cavity and the added wavelength selection components. Since they never become the mainstream product except in a few specific applications where the cost is not a concern, SLM lasers in this category are not discussed. The interested readers can refer to, e.g., References [6–8].

30.1.6 Organization of this Chapter

Sections 30.2 and 30.3 will be dedicated to discuss the grating-assisted lasers and the coupled-cavity lasers, respectively, covering their structures and working mechanisms, governing equations describing the device physics processes, and numerical simulation results on device performance.

The following section briefly describes the recent development on this topic, by showing a few advanced structures for emerging demands, their operating principles, the measured prototype device performance, in contrast to numerical simulation results.

30.2 Grating-Assisted SLM Laser Diodes

30.2.1 Grating Analysis

Noticing that a normal facet mirror formed by a dielectric (semiconductor)—dielectric (air) interface doesn't have wavelength dependence, which makes the FP cavity formed by a straight and smooth waveguide with such facet mirrors on both ends lacking the ability to select a single lasing wavelength, one would naturally think of exploiting a wavelength-selective reflector to replace the end facet mirror, which directly leads to the birth of distributed Bragg reflector (DBR) laser. This idea can logically be extended to turn the entire or part of the original straight and smooth waveguide inside the cavity into a corrugated waveguide (i.e., the waveguide grating) for lasing wavelength selection, which leads to the innovation of a whole family of DFB lasers, although the original idea of the DFB laser was independently proposed [9]. In all such grating-assisted SLM lasers, the grating obviously plays a center role. We therefore start this section by analyzing the grating itself.

A uniform (passive) grating with period Λ in any shape can be viewed as a linear superposition of many sinusoidal gratings with harmonic periods Λ/m, $m = 1, 2, 3, \ldots$, according to the Fourier expansion, simply because the refractive index change along the grating follows a periodic function. A phase-matching condition can therefore be found as follows [10]:

$$k_i \cos\theta_i = \pm\frac{2\pi}{(\Lambda/m)} + k_{om}\cos\theta_{om} \tag{30.3}$$

with k_i and k_{om} indicating the incident and mth order diffracted (by the mth order sinusoidal grating) wave numbers, respectively, θ_i and θ_{om} the incident and mth order diffracted wave angles as shown in Figure 30.2, respectively.

In this application, the incident light is propagating along the laser waveguide, $\theta_i = 0$ and $k_i = \beta = 2\pi n_{\mathrm{eff}}/\lambda$. What one wants is to have the incident wave turn around by 180°, propagating along the waveguide in the opposite direction, which requires $\theta_{om} = \pi$ and $k_{om} = \beta = 2\pi n_{\mathrm{eff}}/\lambda$, hence one has, from Equation 30.3,

$$2\beta = \frac{4\pi n_{\mathrm{eff}}}{\lambda} = \frac{2\pi m}{\Lambda}, \text{ or } \Lambda = \frac{m\lambda}{2n_{\mathrm{eff}}} \tag{30.4}$$

as the rule to select the grating period for coupling the forward and backward propagating waves in the waveguide at a given wavelength (λ), with n_{eff} still indicating the effective index of the waveguide. The wavelength that satisfies condition (Equation 30.4) is conventionally called Bragg wavelength. It is self-evident that for a given grating with fixed Λ, any wavelength deviated from λ wouldn't satisfy the phase-matching condition (Equation 30.4), consequently at such wavelength there won't be any coupling between the forward and backward propagating waves. As such, an ideal grating (with infinite length) would indeed select a single wavelength to reflect, which makes a perfect mirror as required by an SLM laser cavity. In reality, the grating cannot be made infinitely long. One would therefore expect some residue reflection at deviated wavelengths. Nevertheless, by making the grating sufficiently long, one can always expect a mirror with sufficient wavelength selection required by the SLM operation.

Yet another design consideration is how to choose the grating order. By choosing a sinusoidal grating component with order $m = M$ to be satisfied by Equation 30.4, one understands that those sinusoidal grating components with orders beyond M won't take any effect, simply because if $\lambda = 2n_{\mathrm{eff}}\Lambda/M$, for any grating orders in $M + n$, $n = 1, 2, 3, \ldots$, their diffraction angles according to Equation 30.4,

$$\theta_{o(M+n)} = \arccos\left(1 + \frac{2n}{M}\right) \tag{30.5}$$

offer no real solution at all. Physically, the distance between the two consecutive reflections of the grating in order $M + n$ is $\Lambda/(M + n)$ shorter than half of the equivalent wavelength in waveguide $\lambda/(2n_{\mathrm{eff}}) = \Lambda/M$, therefore, constructive addition of the waves on its reflection is impossible because there is no phase-matching condition (i.e., two identical phases) that can possibly be found within 2π. This can be seen more clearly if we show the round-trip phase delay between the two consecutive reflections explicitly by

$$2\frac{2\pi n_{\mathrm{eff}}[\Lambda/(M+n)]}{\lambda} = 2\frac{2\pi n_{\mathrm{eff}}[\Lambda/(M+n)]}{2n_{\mathrm{eff}}\Lambda/M} = 2\pi\frac{M}{M+n} < 2\pi.$$

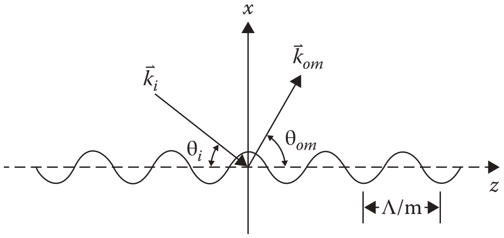


FIGURE 30.2 An illustrative diagram showing the light diffraction by a grating, with Λ/m indicating the mth order harmonic period, θ_i and θ_{om} the incident and mth order diffracted wave angles in respect to the grating plane, respectively, $\vec{k}_i$ and $\vec{k}_{om}$ the incident and mth order diffracted wave vectors, respectively; they must be in the same plane as required by the matching tangential field condition at the grating boundary.

Therefore, those sinusoidal grating components with orders higher than M virtually take no effect, because all the reflected waves by these higher order grating components cannot be added in a constructive way, rather, they all tend to be cancelled out.

For sinusoidal grating components with orders $n = 1, 2, \ldots, M - 1$, they will make the incident wave diffract into different directions, with their angles specified, according to Equation 30.4:

$$\theta_{on} = \arccos\left(1 - \frac{2n}{M}\right) \tag{30.6}$$

For example, if one chose the Bragg wavelength to match the second-order sinusoidal grating component $M = 2$ and $\lambda = n_{\text{eff}}\Lambda$. The second-order sinusoidal grating component therefore couples the forward going wave in the waveguide into the backward going wave, which offers the effect one needs, whereas the first-order sinusoidal grating component couples the forward going wave in the waveguide into a wave going along the vertical direction ($\theta_{o1} = 90°$). The latter appears to be a vertically radiating wave that leaves away from the waveguide. The rest sinusoidal components with their orders higher than 2 all take negligible effect.

In this specific application, one usually doesn't want to create any accompanying radiation that will likely bring in unwanted loss, hence one should align the Bragg wavelength with the first-order sinusoidal grating component $M = 1$ and $\lambda = 2n_{\text{eff}}\Lambda$. As such, the forward and backward waves are coupled at the lowest grating order and consequently it leaves no lower order grating components for the guided waves to be coupled to the radiating wave. Again, those higher order grating components ($M > 1$) virtually bring in no effect.

From the above qualitative analysis, one also knows that it is the period, rather than the shape, that dominates the grating characteristics. And the above analysis is valid under the assumption that the grating is sufficiently long. The required grating length to achieve certain performance, however, can only be evaluated in connection with the grating coupling strength in unit length through numerical approach. Also known as the grating coupling coefficient (κ), the unit length grating coupling strength is determined by the grating shape and the refractive index contrast, or the effective index contrast (Δn_{eff}) on aggregate. In the most general case, one needs to use the mode-matching method (MMM) [11,12] to calculate the reflection, transmission, and loss spectra of a given waveguide grating. A good approximation is to chop the waveguide grating into many short pieces along the guiding direction. By solving the mode in each piece as an eigenvalue problem defined in the local waveguide cross section, one will be able to extract the effective index in each piece as the eigenvalue solution. One can therefore use the one-dimensional (along the guiding direction) transfer matrix method (TMM) [13,14] to solve for the aforementioned grating spectra.

While numerical approaches would offer us an accurate result, one sees little physics and it is also hard to link the grating design parameters to its characteristics. For this reason, one usually uses the full numerical solver as a simulation tool only to validate or confirm a design. One therefore still needs an analytical or semi-analytical tool for the grating design purpose. For this reason, we will show following the expressions [15]. Actually, the refractive index distribution of a waveguide grating can be written in the Fourier series:

$$n^2(x, y, z) = n_0^2(x, y) + \sum_{m \neq 0} \Delta n_m^2(x, y) e^{jm\frac{2\pi z}{\Lambda}} \tag{30.7}$$

due to its periodicity along z (the guiding direction), where n_0^2 and Δn_m^2 indicate the DC and mth order coefficient of the square refractive index in the expansion, respectively, and can readily be found by taking the overlap integral (along z within one period Λ) between the square refractive index (n^2) and the respective sinusoidal grating order (in its complex form $e^{jm2\pi z/\Lambda}$, $m = 0, \pm 1, \pm 2, \ldots$) normalized by the period Λ. The cross-sectional (x, y) dependence of the refractive index reflects the waveguide structure. Consequently, the coupling coefficient of the m^{th} order sinusoidal grating component can be found by

$$\kappa_m = \frac{\pi}{\lambda n_{\text{eff}}} \frac{\iint\limits_{\Sigma} \Delta n_m^2(x, y)|\varphi(x, y)|^2 dxdy}{\iint\limits_{\Sigma} |\varphi(x, y)|^2 dxdy} \tag{30.8}$$

where n_{eff} and $\varphi(x, y)$ denote the effective index and optical field distribution of the waveguide defined by the background refractive index distribution (i.e., the DC component in the Fourier expansion) $n_0^2(x, y)$, respectively, and can be found by solving the following eigenvalue problem (with n_{eff} and φ taken as the eigenvalue and eigenfunction, respectively):

$$\left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)\varphi(x, y) + \left(\frac{2\pi}{\lambda}\right)^2 [n_0^2(x, y) - n_{\text{eff}}^2]\varphi(x, y) = 0 \tag{30.9}$$

In Equation 30.8, the integration area Σ extends to where φ can reach. In following discussions, we will stay with the first-order Bragg grating unless otherwise specified. Hence we will use $\kappa_\pm$ to indicate $\kappa_M = \kappa_{\pm 1}$. One also needs to note that Equation 30.7 is only valid for an infinitely long grating, consequently the grating coupling coefficient ($\kappa_\pm$) given in Equation 30.8 is subject to the same assumption. Practically for $n_{\text{eff}}L/\lambda > 500$, such extracted $\kappa_\pm$ in conjunction with the analytical spectrum expression (shown below) provides no appreciable difference from the accurate result obtained by the full numerical approaches (e.g., MMM or TMM).

For a fully confined guided mode (hence φ is real), both the grating symmetry and the nature of the complex material refractive index change dictate the coupling complexity. For a purely refractive index-coupled grating with real $n^2 - n_0^2$, one may easily find from Equation 30.7 that $\Delta n_{-1}^2 = (\Delta n_{+1}^2)^*$, hence $\kappa_- = \kappa_+^*$ according to Equation 30.8. One may conclude that, if the grating has a center symmetry, n^2 is an even function of z and $\Delta n_{\pm 1}^2$ are real, $\kappa_\pm$ are real and $\kappa_- = \kappa_+$, whereas if the grating has a center antisymmetry, n^2 is an odd function of z and $\Delta n_{\pm 1}^2$ are imaginary, $\kappa_\pm$ are imaginary and $\kappa_- = -\kappa_+$. For a purely gain- or loss-coupled grating with imaginary $n^2 - n_0^2$, one finds, however, $\Delta n_{-1}^2 = -(\Delta n_{+1}^2)^*$ from Equation 30.7, hence $\kappa_- = -\kappa_+^*$ following Equation 30.8. One therefore knows that, if the grating has a center symmetry, n^2 is an even function of z and $\Delta n_{\pm 1}^2$ are imaginary, $\kappa_\pm$ are imaginary and $\kappa_- = \kappa_+$, whereas if the grating has a center antisymmetry, n^2 is an odd function of z and $\Delta n_{\pm 1}^2$ are real, $\kappa_\pm$ are real and $\kappa_- = -\kappa_+$. Generally, for a complex-coupled grating with complex $n^2 - n_0^2$, or for a grating structure that is neither symmetric nor antisymmetric, $\Delta n_{\pm 1}^2$ are complex so are $\kappa_\pm$.

For a waveguide grating with a length of L without any extra end facet reflection [which can practically be realized by, e.g., antireflection (AR) coating on the end facet], one finds the amplitude reflection spectrum (defined as the ratio of the backward and forward going guided wave at the input port) and the amplitude transmission spectrum (defined as the ratio of the forward going guided wave at the output and input port) as [16]:

$$R_{DBR}^{\pm}(\lambda) = \frac{j\kappa_\mp \sinh(\gamma L)}{\gamma \cosh(\gamma L) - (\alpha + j\delta)\sinh(\gamma L)} \tag{30.10}$$

$$T_{DBR}^{\pm}(\lambda) = \frac{\gamma}{\gamma \cosh(\gamma L) - (\alpha + j\delta)\sinh(\gamma L)} \tag{30.11}$$

where $\gamma^2 = (\alpha + j\delta)^2 + \kappa_+\kappa_-$ and $\delta = 2\pi[n_{\text{eff}}/\lambda - 1/(2\Lambda)]$, with λ now indicating the wavelength variable and α the amplitude modal loss (< 0) or gain (> 0) variable of the waveguide grating. Figure 30.3

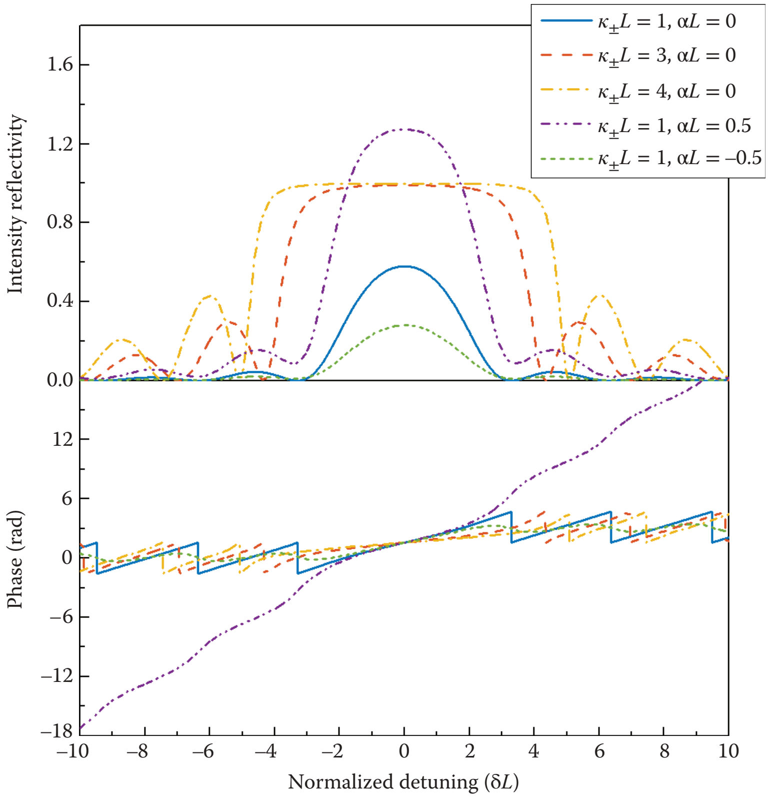


FIGURE 30.3 Reflection spectra of the Bragg grating under different normalized coupling coefficients $\kappa_{\pm}L = 1, 3, 4$ without gain or loss $\alpha L = 0$, and under $\kappa_{\pm}L = 1$ with gain $\alpha L = 0.5$ and loss $\alpha L = -0.5$.

shows respectively the amplitude and phase of the reflection spectrum of a purely refractive index-coupled grating, under different normalized (real) grating coupling coefficient ($\kappa_{\pm}L$) and normalized modal loss/gain (αL).

As can be seen from the spectrum, the maximum reflection happens at the Bragg wavelength. A Bragg stop-band can be identified by the main lobe with its width defined as the difference between the two zero-reflection wavelengths. Zero-reflection happens to the wave at the wavelength with its reflections from two adjacent grating teeth completely canceled out due to a phase delay of π, which makes the further cancel out of all reflections through the entire grating due to the equal tooth distance (i.e., the grating period Λ). Obviously, the wave at the zero-reflection wavelength will pass through the grating with 100% transmissivity as if the waveguide grating offers nothing more than a phase delay unit. For a purely passive, refractive index-coupled grating ($\kappa_+\kappa_- = \kappa_+\kappa_+^*$, hence $\kappa_+\kappa_-$ is real and can be written as $|\kappa|^2$) without any modal gain or loss ($\alpha = 0$), zero-reflections (maximum transmission) happen at $\gamma L = jk\pi$, $k = \pm 1, \pm 2, \pm 3, \ldots$, or $\delta L = \pm\sqrt{(|\kappa_+|L)^2 + (k\pi)^2}$. The Bragg stop-band width is therefore given by [16]

$$2\delta L = 2\sqrt{(|\kappa|L)^2 + \pi^2} \text{ or } \Delta\lambda_B = \frac{\lambda_B^2}{n_{\text{eff}}L}\sqrt{(|\kappa|L/\pi)^2 + 1} \tag{30.12}$$

with $\lambda_B = 2n_{\text{eff}}\Lambda$ indicating the Bragg wavelength (center of the Bragg stop-band) given by the phase-matching condition (Equation 30.4) for the first-order Bragg grating. It is apparent that the Bragg stop-band width ($\Delta\lambda_B$) increases with the normalized coupling coefficient almost linearly if $|\kappa|L >> \pi$, or quadratically if $|\kappa|L << \pi$. It is worth mentioning that the measured stop-band width of the lasing spectrum in the DFB lasers is always smaller than the Bragg stop-band width, and the former approaches $\Delta\lambda_B$ as $|\kappa|L$ increases. We explain the underlying physics later in discussion of DFB lasers.

In the near infrared fiber-optic communication wavelength range, the first-order Bragg grating has its period in the submicrometer range (e.g., ~200 nm in the O-band centralized at 1300 nm and ~240 nm in the C-band centralized at 1550 nm). It is impossible to fabricate the waveguide surface-relief grating by the standard photolithography technique. Rather, one has to exploit more complicated technologies such as the holographic lithography [17] or the electron-beam lithography (EBL) [18]. While the former doesn't need expensive facility and fits the mass production mode, it can hardly create any nonuniform grating pattern. The latter, conversely, can create arbitrary user-defined grating pattern but needs expensive machine and doesn't fit the production mode for the long hours it needs to scan-write the whole wafer. A newly emerged technique, nano-imprinting [19], seems to be very promising as the ultimate solution to the fabrication of the waveguide grating. With the imprinting mask written by EBL with complicated patterns, one can use the mask repeatedly to transfer its pattern onto wafers by a pressing-and-developing process through the spin-coated pressure-sensitive deforming-resister on the wafer surface. As such, the expensive EBL writing just needs to be done once on the mask, the later pattern transfer process from the mask to wafers (imprinting) is fast and reliable, with an even better quality compared to the widely used holographic lithography technique [20]. It is also worth mentioning that the invention of the "floating" grating idea [21] solves the problem in the precise control of the grating coupling strength and guarantees its repeatability in the mass production mode, as the depth of the grating teeth is now determined by the grating layer thickness which can be made accurate (to sub-10 nm range) in material growth, as opposed to be given by the etching depth which is not only hard to control down to 10~20 nm but also hardly repeatable. This invention sets a milestone on the mass production as well as the cost reduction of DFB lasers.

30.2.2 Distributed Bragg Reflector Laser

A typical DBR laser has a structure shown in Figure 30.4. With the first-order Bragg grating placed on one or both ends of an FP laser as the reflector, the coupling between the forward and backward going waves inside the cavity becomes wavelength dependent.

Similar to the lasing condition derived for the FP laser [5], one can readily find the resonance condition for the DBR laser by following an approach of matching the traveling wave with itself after a

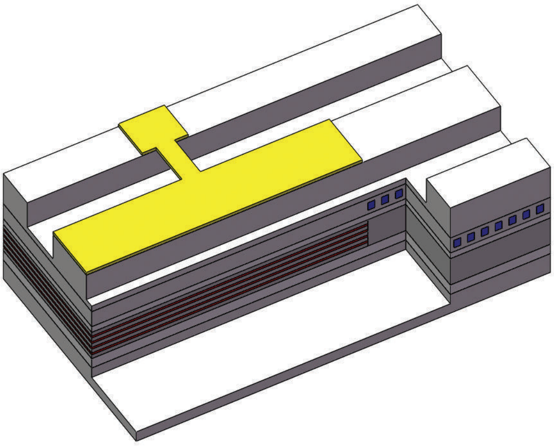


FIGURE 30.4 The DBR laser structure.

round trip inside the cavity:

$$R_l(\lambda)R_r(\lambda)e^{2[g-\alpha_i+j\frac{2\pi n_{\text{eff}}}{\lambda}]L_{ar}} = 1 \tag{30.13}$$

with R_l and R_r denoting the wavelength-dependent amplitude reflectivity of the Bragg grating reflector on the left and right end of the active region, respectively, given in the form of (Equation 30.10), g the modal gain, α_i the modal internal loss, and L_ar the active region length. One can further split the lasing condition (Equation 30.13) into the amplitude and phase condition as

$$g = g_{th} \equiv \alpha_\text{i} + \alpha_\text{c} \equiv \alpha_i + \frac{1}{2L_\text{ar}} \ln \frac{1}{|R_l(\lambda)||R_r(\lambda)|} \tag{30.14}$$

and

$$\frac{4\pi n_{\text{eff}} L_{ar}}{\lambda} + \phi_l(\lambda) + \phi_r(\lambda) = 2\pi m,\ \ m = 1, 2, 3, \ldots \tag{30.15}$$

where ϕ_l and ϕ_r indicate the phase of the amplitude reflectivity of the Bragg grating on the left and right end, respectively. It is clear that, following Equation 30.14, only the wave at the Bragg wavelength sees the highest reflection and consequently has the smallest cavity loss defined by $\alpha_\text{c} \equiv -(0.5/L_\text{ar})\ln(|R_\text{l}||R_\text{r}|)$, since at least one of the end reflectors will offer the highest reflection (i.e., either R_l, or R_r, or both will take the largest value) at the Bragg wavelength following the grating reflection spectrum shown in Figure 30.3. It then requires the smallest modal gain to reach the threshold in the neighborhood of the Bragg wavelength. It is not as clear though, following Equation 30.15, that one can find a solution to λ in the neighborhood of the Bragg wavelength, not going beyond $\lambda_B^2/(2n_\text{eff}L_\text{ar})$, with λ_B indicating the Bragg wavelength ($2n_\text{eff}\Lambda$). Actually, if one takes $\lambda = \lambda_\text{B} + \Delta\lambda$ ($\Delta\lambda \sim \lambda_\text{B}^2/(2n_\text{eff}L_\text{ar}) << \lambda_\text{B}$) in Equation 30.15 to obtain

$$2\pi\Delta\lambda\left(\frac{2n_\text{eff}L_\text{ar}}{\lambda_\text{B}^2}\right) = \frac{4\pi n_\text{eff}L_\text{ar}}{\lambda_\text{B}} + \phi_l(\lambda) + \phi_r(\lambda) - 2\pi m \tag{30.16}$$

and also to notice that the total phase on the right hand side of Equation 30.16 is a slow-varying function of λ in the neighborhood of λ_B following Figure 30.3, one immediately finds that as $\Delta\lambda$ changes from 0 to $\lambda_\text{B}^2/(2n_\text{eff}L_\text{ar})$, the total phase on the left-hand side of Equation 30.16 sweeps over an entire range of 2π, which means that the phase on the two sides of Equation 30.16 will have a matching point within the interval between λ_B and $\lambda_B \pm \lambda_B^2/(2n_\text{eff}L_{ar})$, given the fact that the phase takes 2π as its modulo. One therefore reaches the conclusion that the phase condition (Equation 30.15) always has a solution in a close neighborhood of the Bragg wavelength bound by $\lambda_\text{B}^2/(2n_\text{eff}L_\text{ar})$. Consequently, one knows that the amplitude condition (Equation 30.14) and the phase condition (Equation 30.15) jointly select a single lasing wavelength near the Bragg stop-band center of the end grating reflector. This explains why the DBR laser operates under the SLM.

The main advantage of the DBR laser lies in the separation of its gain and wavelength selection (i.e., the passive Bragg grating) region. Since the grating region is not biased, the wavelength selection mechanism suffers little change with the injection current in the gain region. One then expects a stable SLM operation with high side-mode-suppression-ratio (SMSR) and small wavelength chirp, which is evidenced by many publications [22,23]. It is also quite convenient to introduce a lasing wavelength tuning in DBR laser, as one can readily bias the grating region by a separate electrode from the gain region. As such, the current injection in the grating region introduces an effective index change, since the carrier-induced gain/loss change will have an accompany refractive index change following the Kramers–Kronig relation. Therefore, the peak reflection wavelength (the Bragg wavelength) given by Equation 30.4 will change with the current injected in the grating region, which causes the lasing wavelength change accordingly. Usually, a phase adjustment section needs to be inserted between the gain and the grating region to separately tune the phase for matching the lasing condition, as otherwise the DBR laser may cease lasing or experience a lasing mode hopping due to the mismatch of the phase condition. However, since the grating region has to be made transparent to the lasing wavelength, as otherwise the light will be absorbed by the unbiased passive grating region if it has the same band-gap with the active region, the fabrication of DBR lasers inevitably involves the monolithic integration technique that still stands as an unsolved problem up to today, for its low yield. With the birth of silicon photonics in the recent decade, it is quite promising to make the Si photonics DBR laser by bonding a direct band-gaped III-V compound semiconductor gain block on top of the silicon-on-insulator (SOI) waveguide, with the Bragg grating engraved on SOI. There is a hope that the DBR laser will be reborn with the booming Si photonics.

30.2.3 Distributed Feedback Lasers

A typical pure refractive index-coupled uniform-grating DFB laser has a structure shown in Figure 30.5. Unlike the DBR laser with a gain region made of smooth waveguide separated from the passive grating region, the DFB laser has these two regions merged into one, which makes the laser cavity substantially different from the aforementioned FP or DBR structures in the sense that the forward and backward going waves are constantly coupled in a distributive manner inside the cavity rather than in a lumped sum manner at the two ends of the cavity only. A wavelength discrimination mechanism is therefore brought in by such distributed coupling, since only the waves at the phase matched wavelength(s) will possibly add constructively to establish a standing wave pattern (i.e., a longitudinal mode) inside the cavity, which imposes a stringent condition that may likely purifies the lasing spectrum by cutting off most of the cavity modes otherwise allowed in the FP cavity.

An immediate finding is that this structure won't have its lasing wavelength at the Bragg wavelength anymore. This is because the wave near the Bragg wavelength sees the highest reflection. It therefore cannot travel far along the waveguide grating that also provides the gain. Consequently, it cannot obtain sufficient gain as required for the lasing to happen. On the contrary, at the Bragg stop band edge where the grating offers zero-reflection as shown in Figure 30.3, a maximum transmission through the grating region is obtained. Although the wave at this particular wavelength experiences the highest single-pass gain as it sees the entire length of the grating (with gain), it cannot be the lasing mode either since no reflection happens to it so that no resonance can be established. In this sense, the lasing should happen somewhere in between, i.e., the traveling waves at the lasing wavelength should be partially reflected by the grating to

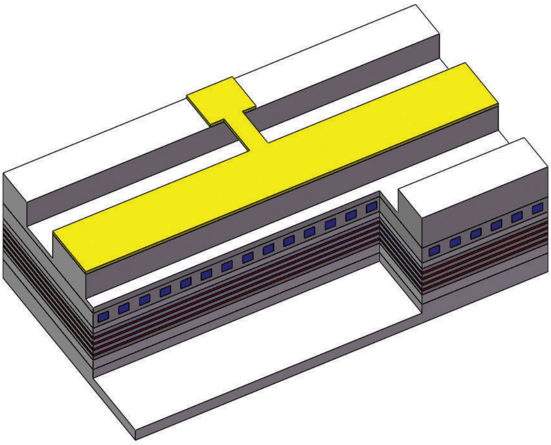


FIGURE 30.5 The DFB laser structure.

establish the resonance. It should be partially transmitted in order to experience the gain in the grating region as well. As such, the phase matching and amplitude sustenance (gain) conditions must be mixed.

Following the same thought, i.e., searching for the consistent condition between the wave and itself after a round trip inside the cavity, one can still obtain the resonance condition for the DFB cavity [24], which will stand as the lasing condition of the DFB laser. In the general case, however, one cannot expect to have analytical expressions for such lasing condition, so that we will cover the general case by the full numerical model introduced later. To gain an insight into the DFB laser, we will analyze a simple case with both of the end facet reflections and longitudinal spatial horn burning (LSHB) effect neglected. Actually, the lasing condition of such a simple DFB laser can be found by setting the denominator of Equation 30.10 and/or Equation 30.11 equal to zero. The associated infinite reflection and transmission indicate that the device can offer output without input—a feature must be carried by an oscillator. Hence one has

$$\gamma \coth(\gamma L) = \alpha + j\delta, \ \gamma^2 = (\alpha + j\delta)^2 + \kappa_+\kappa_- \tag{30.17}$$

Usually, for a given $\kappa_\pm L$ as the grating design parameter, one solves for γL from

$$(\gamma L)^2 \coth^2(\gamma L) = (\gamma L)^2 - \kappa_+\kappa_- L^2 \tag{30.18}$$

which has the same solution set as Equation 30.17. Hence the normalized gain (αL) and detuning (δL) can be found by

$$\alpha L + j\delta L = \pm\sqrt{(\gamma L)^2 - \kappa_+\kappa_- L^2} \tag{30.19}$$

from which one will be able to obtain the lasing modal threshold gain (from $g_{\text{th}} \equiv \alpha_{\text{i}} + \alpha$, α_i the modal internal loss) and the lasing wavelength (by solving $\delta = 2\pi[n_{\text{eff}}/\lambda - 1/(2\Lambda)]$ for λ). The normalized gain (αL) and detuning (δL) for different normalized grating coupling coefficient ($\kappa_\pm L$) are plotted in Figure 30.6, as the solution to Equations 30.18 and 30.19 obtained by a numerical root searching program built on Muller's algorithm [25].

The threshold gain solution indicates that, in a similar role as $\alpha_c = -(0.5/L)\ln(|R_{\text{l}}||R_{\text{r}}|)$ to the FP or DBR cavity, α, as the cavity loss to the DFB cavity, is in proportion to $1/(\kappa L)$. For a DFB grating design

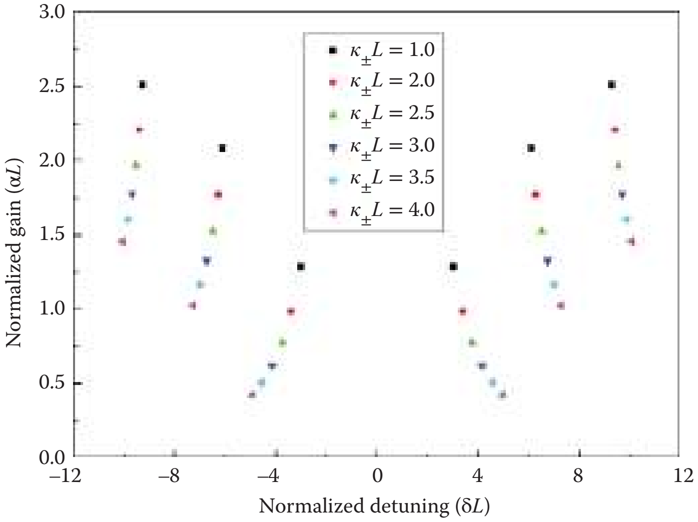


FIGURE 30.6 $\alpha L \sim \delta L$ for different $\kappa_\pm L$ as solutions to Equations 30.18 and 30.19.

with a higher $\kappa_{\pm}$, the cavity is more closed as an FP or DBR cavity with a higher end reflection R, and with the cavity length scaled in the same way. This conclusion, together with the above discussion on its lasing wavelength position, stands as the most basic design guidance for the DFB laser. Also for the DFB laser with high $\kappa_{\pm}L$ design, its cavity loss reduces so as the required modal gain to reach the lasing threshold, the lasing wavelength will therefore approach more closely to the zero-reflection wavelengths, as the wave may have sufficient gain to reach its threshold through fewer passes, and consequently less reflection is necessary. Namely, as the threshold gain approaches to zero, the lasing will happen near the Bragg stop-band edge. Hence one finds that the lasing wavelength of the DFB laser with higher $\kappa_{\pm}L$ should be closer to the zero-reflection wavelength of the unbiased grating, i.e., the Bragg stop-band edge.

A serious issue with the purely refractive index-coupled uniform-grating DFB laser is, however, its dual-mode operation nature originated from the double degeneracy caused by the center mirror-symmetry (or antisymmetry) of the cavity. It is quite obvious that, from Equation 30.18, for any real $\kappa_{+}\kappa_{-}$, once γ is a solution, $-\gamma$ and $\pm\gamma^{*}$ are all solutions. Hence for any solution set (α, δ), $(\pm\alpha, \pm\delta)$ are all possible solutions according to Equation 30.19, which means that for any modal gain required to reach the lasing threshold, also known as the DFB laser cavity loss (α as a solution to the lasing condition), there is a double degeneracy of the lasing wavelength corresponding to $\pm\delta$, respectively. These two lasing wavelengths sit on each side of the Bragg wavelength, and approach to the Bragg stop-band edges as $\kappa_{\pm}L$ increases, in consistency with previous analysis. Physically, this effect can be explained by the two equally possible standing wave patterns shown in Figure 30.7a. In a DFB laser cavity with the center mirror-symmetry ($L/\Lambda = n + 1/2$, n is an integer) or center mirror-antisymmetry ($L/\Lambda = n$ is an integer) and with zero end facet reflections (the latter condition can actually be further relaxed to identical end facet reflections), there exist two possible standing wave patterns as a result of the constructive addition of the distributively coupled forward and backward going waves. They share the same gain inside the cavity and have the identical cavity loss. The only difference between the two patterns is that one of them has its intensity peak aligned with the high refractive index region in every period, whereas the other one has its intensity peak aligned with the low refractive index region in every period. As such, the former and latter pattern will see the cavity mainly in high and low refractive index, respectively. Consequently, the former and latter will take the longer (red) and shorter (blue) lasing wavelength, respectively, on each side of the center Bragg wavelength. The double degenerated longitudinal mode with different standing wave pattern at different lasing wavelength but with

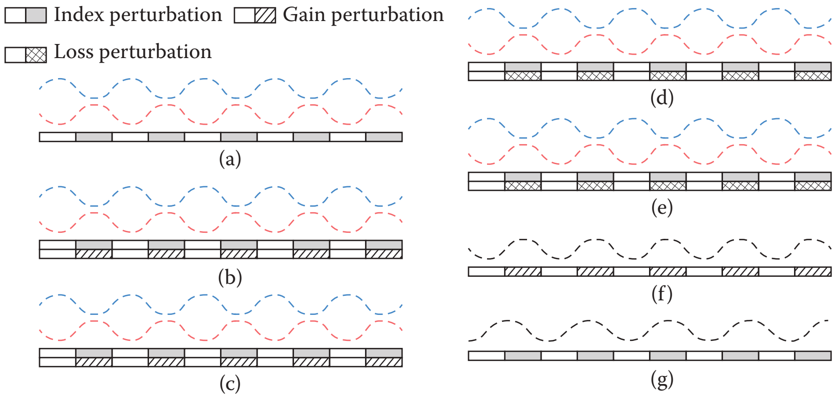


FIGURE 30.7 Schematic diagrams for (a) purely refractive index-coupled uniform-grating DFB, (b) partially gain-coupled in-phase DFB, (c) partially gain-coupled antiphase DFB, (d) partially loss-coupled in-phase DFB, (e) partially loss-coupled antiphase-grating DFB, (f) purely gain-coupled DFB, (g) quarter-wavelength phase-shifted (QWPS) DFB; the blue and red dashed lines in (a)–(e) show the intensity standing wave patterns of the longitudinal modes bearing the short and long wavelength (on the blue and red side of the Bragg wavelength), respectively; the black dashed line in (f) and (g) shows the intensity standing wave pattern of the longitudinal mode bearing the Bragg wavelength.

the same threshold gain breaches the SLM operation condition, which makes the purely refractive index-coupled uniform-grating DFB with a perfect center mirror-symmetric (or antisymmetric) cavity not a SLM laser.

This conclusion can also be generalized to any cavity with the perfect center mirror-symmetry or center mirror-anti-symmetry, once its reflection (or transmission) spectrum is symmetric, a necessary condition to reach the SLM operation is to have its lasing wavelength right in the center of the spectrum. Otherwise, dual-mode operation is inevitable. By recalling the lasing condition for the DBR laser, one may conclude that, in order to ensure its SLM operation with high SMSR, one has to either somehow break the symmetry (or anti-symmetry) of the cavity by, e.g., introducing two slightly different end reflection gratings (usually with slightly misaligned Bragg wavelengths) or use the grating reflector on one end only, or to ensure that the solution to the phase condition (Equation 30.15) is at the Bragg wavelength so that the lasing will happen exactly in the reflection spectrum center.

Since it is impossible to make the purely refractive index-coupled uniform-grating DFB lase at the Bragg wavelength, a conclusion not only drawn from the analysis at the beginning of the discussion on DFB lasers, but also known by the lasing condition (Equation 30.18). This is because for real $\kappa_+\kappa_-L^2$, real γL cannot be a solution as otherwise, the right-hand side is smaller than γL whereas the left-hand side is bigger than γL, hence γL must be complex, which leaves $\delta \neq 0$. The only viable way to reach the SLM operation for such DFB laser is then to break the cavity symmetry (or antisymmetry) by introducing asymmetric facet coatings on the two ends, usually a combination of the AR coating for the front facet and the high-reflection (HR) coating for the rear facet, respectively. However, the SLM yield with high SMSR for such DFB laser is still low since the grating phase at the two ends is usually random, given the condition that the grating period is in 1/5~1/4 of a micrometer whereas the uncertainty of the facet cleaving position is in a range of $\pm(2 \sim 3)\mu m$. Particularly for those DFB lasers with high $\kappa_\pm L$ designs, they usually suffer from very poor SLM yield (only ~15% for $SMSR > 30$ dB with $\kappa_\pm L > 3$) [26]. This is because the DFB laser with high $\kappa_\pm L$ tends to have the envelope of the standing wave pattern piled up in the middle, rather than at the two ends, of the cavity, due to the strong coupling, which makes the cavity less affected by the end facet condition as the waves inside the cavity simply don't "feel" much about the existence of the cavity ends. As a result, the end facet asymmetry applied on a strongly coupled DFB cavity is not sufficient to break the degenerated threshold gain between the two modes. Hence one finds that high $\kappa_\pm L$ often triggers dual mode lasing in DFB lasers. Figure 30.8 shows the standing wave envelope distributions inside the DFB cavity under different coupling strengths, from which one finds that the field stays at the two ends, or piles up in the middle of the cavity, corresponding to the weak (with $\kappa_\pm L < \pi/2$), or strong (with $\kappa_\pm L > \pi/2$) coupling condition, respectively. And the field takes almost an even distribution along the cavity under the critical condition $\kappa_\pm L = \pi/2$, as it has to be.

To thoroughly solve the dual-mode operation problem associated with the purely refractive index-coupled uniform-grating DFB laser, one can have different approaches along with the thought of breaking the cavity symmetry, or forcing the lase at the Bragg wavelength. While the former led to the invention of various complex-coupled DFB lasers [27–30], the latter resulted in the popular product of the QWPS DFB laser [31].

Figure 30.7b shows the working principle of the in-phase partially gain-coupled DFB laser. As an extra gain is introduced periodically inside the high refractive index region, the longitudinal mode with its intensity peak in its standing wave pattern located in the high refractive index region in each period obtains more gain than the other mode, hence the degeneracy between the two longitudinal modes breaks and only the mode with the longer wavelength (on the right/red side of the Bragg wavelength) will lase, which makes such DFB laser operate under SLM. Figure 30.7c shows the working principle by taking the other option to boost the longitudinal mode bearing the shorter wavelength (on the left/blue side of the Bragg wavelength). Known as the antiphase partially gain-coupled DFB laser, the structure has an extra gain introduced periodically inside the low refractive index region, which again breaks the degeneracy between the two longitudinal modes since the longitudinal mode with its intensity peak in its standing wave pattern located in the low refractive index region in each period has more gain than the other

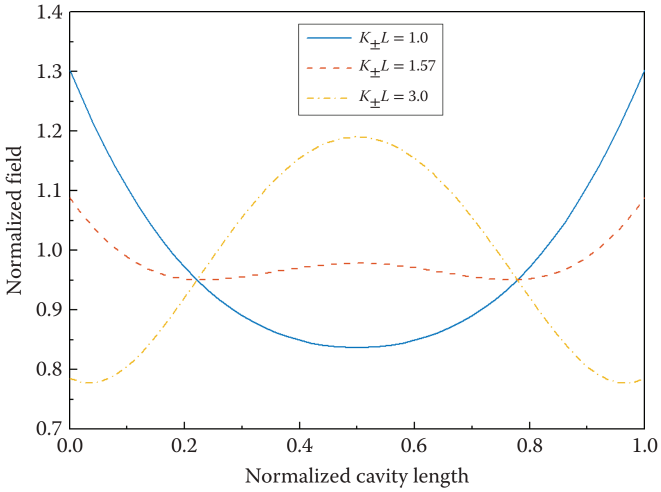


FIGURE 30.8 Field envelope distributions inside the DFB cavity for $\kappa_{\pm}L = 1.0(< 1.57)$, $\kappa_{\pm}L \sim 1.57$, and $\kappa_{\pm}L = 3.0(> 1.57)$.

mode. As a result, only the shorter wavelength mode will lase, which again makes such DFB laser operate under SLM.

Figure 30.7d and e shows the working mechanisms of the partially loss-coupled in-phase and antiphase DFB lasers, respectively. If an extra loss is introduced periodically inside the high (or low) refractive index region, the longitudinal mode with its intensity peak in its standing wave pattern located in the high (or low) refractive index region in each period suffers from more loss than the other mode, hence the degeneracy between the two longitudinal modes breaks and only the mode with the shorter (or longer) wavelength on the left/blue (or right/red) side of the Bragg wavelength will lase, due to the suppression of its counterpart on the other side of the Bragg wavelength, which makes such DFB laser operate under SLM.

One can also understand why the degeneracy breaks in such partially gain- or loss-coupled DFB lasers by the lasing condition (Equation 30.18). For complex-coupled DFBs, $\kappa_+\kappa_-L^2$ is also complex or $\kappa_+\kappa_-L^2 \neq (\kappa_+\kappa_-L^2)^*$. As such, if γ is a solution, γ^* is not necessarily a solution. Following Equation 30.19, one then finds that $\pm\delta$ cannot share the same α as the solutions.

Figure 30.7f shows the working principle of the purely gain-coupled DFB laser. Since the purely gain-coupled laser has a purely negative $\kappa_+\kappa_-$ ($\kappa_- = -\kappa_+^*, \kappa_+\kappa_- = -\kappa_+\kappa_+^* \equiv -|\kappa|^2$), which makes the lasing condition (Equation 30.18) turn into

$$(\gamma L)^2 \coth^2(\gamma L) = (\gamma L)^2 + (|\kappa|L)^2 \tag{30.20}$$

a real solution of γ is then possible which forces the lasing happen at the Bragg wavelength according to Equation 30.19, or

$$\alpha L + j\delta L = \sqrt{(\gamma L)^2 + (|\kappa|L)^2} \tag{30.21}$$

from which one readily achieve $\delta = 0$ for the right-hand side is real. The dual mode lasing in purely gain-coupled DFB laser is not possible as one cannot find two different standing wave patterns inside the cavity that share the same gain. It is obvious that only a single standing wave pattern is allowed inside the cavity with itself completely aligned with the gain grating. Any standing wave pattern in different shape with the gain grating will naturally be suppressed. Therefore, one knows that Equation 30.20 cannot have a complex

γ as its solution associated with the longitudinal mode with the lowest cavity loss (α). As otherwise, γ^* will be an allowed solution as well, according to Equation 30.21, one will then be able to have a pair of lasing wavelengths corresponding to the doubly degenerated solutions $(\alpha, \pm\delta)$, which is in conflict against our previous physics-based analysis on the purely gain-coupled DFB laser cavity. Namely, solution $(\alpha, 0)$ must take the smallest cavity loss $\alpha = \alpha_{\min}$, while any other possible doubly degenerated solutions in the form of $(\alpha, \pm\delta)$ with $\delta \neq 0$ must have a higher cavity loss $\alpha > \alpha_{\min}$ hence cannot be the lasing modes. This conclusion has been evidenced by a numerical searching for all possible solutions of Equation 30.20 [32]. Finally, one may conclude that the purely gain-coupled DFB laser will lase at the Bragg wavelength and naturally stands as the SLM laser.

Figure 30.7g shows the working mechanism of the QWPS DFB laser. The only difference from the uniform-grating purely refractive index-coupled DFB is that there is an extra half period of the grating inserted right at the cavity center. An extra half period of the grating is equivalent to a quarter-wavelength phase shift in the first-order grating following the relationship between the grating period and the Bragg wavelength $\Lambda = \lambda_B/(2n_{\text{eff}})$, which therefore introduces a π phase shift to the reflected wave or a π phase shift to the round trip wave. As such, the wave will have to make two round trips to recover its status, which leads to its resonance. This feature suggests us to fold back the cavity from the center to equivalent the original structure to a uniform-grating DFB with half of its original length and with a perfect facet with 100% reflectivity at one end (the original cavity center). In such a half cavity equivalent structure, if the standing wave pattern has its intensity peak originally aligned with the high refractive index region in every period, upon the 100% reflection at one end, the peak will have to be aligned with the low refractive index region in every period, and vice versa. As such, none of the two original standing wave patterns inside the uniform-grating DFB can exist inside the modified cavity with a quarter-wavelength phase shift introduced at the cavity center. Rather, only a standing wave pattern with its peak aligned at the edge of the grating tooth (i.e., the interface of the high and low refractive index region) in every period can stay. This standing wave pattern is unique inside the cavity and corresponds to the longitudinal mode at the Bragg wavelength, for it sees an averaged refractive index between the high and low index region in every period, which happens to be the effective index of the unperturbed waveguide by the grating. Also physically, a longitudinal mode with its envelope piled up in the middle of the cavity will see the highest reflection from both sides if it bears the Bragg wavelength. Hence this mode will be most tightly confined inside the cavity and sees the most gain. Consequently, it should have the lowest cavity loss and becomes the lasing mode. The numerically calculated longitudinal mode distribution along the cavity of such structure confirms this conclusion [33] as one finds that, with its lasing mode always at the Bragg wavelength, the QWPS DFB laser has a sharp longitudinal mode distribution peak at the center of the cavity. In a uniform-grating DFB without a phase-shifted center inside the cavity, this scenario can never happen, as the maximum reflection at the Bragg wavelength can only be possibly seen by the wave from one side rather than from both sides. Hence the wave at the Bragg wavelength cannot stay long inside the cavity to obtain sufficient gain for lasing, a same conclusion drawn earlier for the uniform-grating purely refractive index-coupled DFB laser. It is worth mentioning that a more detailed analysis would have to have a new lasing condition involved as Equation 30.17 is no longer valid for either a nonuniform grating (the grating with a phase shift) or a uniform grating with nonzero end facet reflection (the equivalently folded cavity with a uniform grating plus 100% reflection at one end). From the newly derived lasing condition for this structure, one will be able to find that $\delta = 0$ is indeed an allowed solution that has the minimum cavity loss ($\alpha = \alpha_{\min}$) [34]. Therefore, the QWPS DFB laser always operates under the SLM and has its lasing mode at the Bragg wavelength.

30.2.4 Governing Equations for Design, Modeling, and Simulation

The following one-dimensional (along the propagation direction, z) traveling wave model is often used as the optical governing equation for the edge-emitting semiconductor lasers with various waveguide-grating structures [16]:

$$
\begin{aligned}
\left(\frac{1}{v_g}\frac{\partial}{\partial t}+\frac{\partial}{\partial z}\right)e^f(z,t) &= [j\left(\frac{2\pi n_{\text{eff}}}{\lambda_0}-\frac{\pi}{\Lambda}\right)+j\frac{2\pi}{\lambda_0}\Gamma\Delta n(z,t,\omega_0) \\
&\quad +\frac{1}{2}\Gamma g(z,t,\omega_0)-\frac{1}{2}\alpha_L]e^f(z,t)+j\kappa_+e^b(z,t)+s^f(z,t) \\
\left(\frac{1}{v_g}\frac{\partial}{\partial t}-\frac{\partial}{\partial z}\right)e^b(z,t) &= [j\left(\frac{2\pi n_{\text{eff}}}{\lambda_0}-\frac{\pi}{\Lambda}\right)+j\frac{2\pi}{\lambda_0}\Gamma\Delta n(z,t,\omega_0) \\
&\quad +\frac{1}{2}\Gamma g(z,t,\omega_0)-\frac{1}{2}\alpha_L]e^b(z,t)+j\kappa_-e^f(z,t)+s^b(z,t)
\end{aligned}
\tag{30.22}
$$

where $e^{f,b}$ denote the slow-varying envelopes of the forward and backward going traveling wave field, respectively, g and Δn the material gain and refractive index change, respectively, α_L the non-interband optical modal loss, ω_0 the reference optical frequency, $\lambda_0 = 2\pi c/\omega_0$ the reference optical wavelength (i.e., the vacuum wavelength of the reference optical frequency), $v_g = c/n_g$ the group velocity, n_{eff} and Γ the effective index (of the waveguide) and confinement factor (of the active region), respectively, Λ and $\kappa_\pm$ the period, the backward-to-forward (+), and the forward-to-backward (−) coupling coefficient of the first-order waveguide Bragg grating, respectively, $s^f = s^b \equiv \tilde{s}$ the spontaneous emission contributions.

The associated optical field can be expressed as

$$
\vec{E}(x,y,z,t) = \frac{1}{2}\vec{s}\varphi(x,y)[e^f(z,t)e^{j\frac{\pi}{\Lambda}z}+e^b(z,t)e^{-j\frac{\pi}{\Lambda}z}]e^{-j\omega_0 t}+c.c. \tag{30.23}
$$

with $\vec{s}$ indicating the unit vector of along the (linear) field polarization direction, φ the field distribution of the waveguide mode as the solution of Equation 30.9. For a given waveguide-grating structure, one needs to solve Equation 30.9 first to find the effective index (n_{eff}) and the guided mode field distribution (φ). The coupling coefficients ($\kappa_\pm = \kappa_{\pm 1}$) can then be obtained from Equation 30.8 and the confinement factor is given as

$$
\Gamma = \int_{A.R.} |\varphi(x,y)|^2 dxdy \Big/ \int_{\Sigma} |\varphi(x,y)|^2 dxdy \tag{30.24}
$$

with the integration areas $A.R.$ and Σ indicating the active region (quantum wells) only and the whole area where the optical field extends.

Equation 30.22 is obtained from the optical wave equation under the slow-varying envelope approximation with the second-order derivatives of the envelope ($\partial^2 e^{f,b}/\partial z^2$ and $\partial^2 e^{f,b}/\partial t^2$) all ignored, whereas the optical wave equation is directly obtained from the Maxwell equations with the coupling among the field polarization components $\nabla(\nabla \cdot \vec{E})$ ignored. The former is true for the edge-emitting device with its cavity length L much longer than the operating wavelength λ (usually L is in a few hundreds to a thousand of λ/n_{eff}), in which the optical wave is propagating along the cavity with its traveling wave factors $\phi_\pm = e^{j(\pm\pi z/\Lambda-\omega_0 t)}$ varying much faster as z and t, hence:

$$
\begin{aligned}
|\partial^2 e^{f,b}/\partial z^2/e^{f,b}| &<< |\partial^2\phi_\pm/\partial z^2/\phi_\pm| = (\pi/\Lambda)^2 \\
|\partial^2 e^{f,b}/\partial t^2/e^{f,b}| &<< |\partial^2\phi_\pm/\partial t^2/\phi_\pm| = \omega_0^2
\end{aligned}
\tag{30.25}
$$

The latter assumption holds for the weakly confined waveguide, which is generally true of most of the III–V compound semiconductor laser diodes with either ridge waveguide or buried heterojunction structures.

In the field expansion (Equation 30.23), the single-guided-mode assumption has been invoked as a necessary condition for the SLM operation. As different guided modes have different effective indices, for the same grating period Λ, different wavelengths will be chosen to satisfy the phase-matching condition (Equation 30.4). As a result, multiple longitudinal modes in different lasing wavelengths, with each of them corresponding to a guided mode in a different effective index, will be excited, which breaches the SLM condition. A single polarization component is assumed for the same reason, but this is not necessary as the difference between the two effective indices corresponding to the two orthogonally polarized components in a guided mode is usually very small in weakly confined waveguides. As a result, the difference between their corresponding lasing wavelengths can hardly be appreciable. Actually, for most of the III–V compound semiconductor laser diodes, the active region is made of the compressively strained multiple quantum wells (CS-MQW). The compressively strained quantum well can only provide the optical gain in the (100) plane [35], which means only the TE mode with its optical E-field polarized in parallel to the slab waveguide interfaces, i.e., the interfaces of the layer stack grown in the usual <100>direction, can see the gain whereas the other polarization component, i.e., the TM mode with its E-field polarized in the perpendicular direction, will be completely suppressed. Therefore, it is usually sufficient to study the TE mode with its E-field horizontally polarized in parallel with the device top and bottom surfaces.

The material gain of semiconductors can generally be calculated by the physics-based first-principles model [16]. For performance simulation of the SLM laser, however, the following model that phenomenologically links the gain to the (minority) carrier density inside the active region is sufficient:

$$g(z,t,\omega_0) = a(\omega_0)\ln\frac{N(z,t)}{N_{\text{tr}}(\omega_0)}/[1+\varepsilon(|e^f(z,t)|^2+|e^b(z,t)|^2)] \tag{30.26}$$

with $a(\omega_0)$ and $N_{\text{tr}}(\omega_0)$ indicating the gain coefficient and the transparent (minority) carrier density for the active region comprising (strained-layer) multiple quantum wells, ε the nonlinear gain saturation factor. These parameters are usually extracted in the neighborhood of the reference optical frequency (ω_0) from the physics-based material gain model or measured experimentally.

The refractive index change and the material gain, corresponding to the real and imaginary part of the material susceptibility, are connected by the Kramers–Kronig relation. Actually, if one views the susceptibility as the frequency-domain transfer function of the material system in responding to the optical E-field as the input signal, the polarization excited by the optical E-field then becomes the output signal. Following the linear system theory, the time-domain polarization is given as the convolution of the time-domain optical E-field and the inverse Fourier transform of the susceptibility. Since a real physical system must be causal, which means the polarization at any time instant t can only be dependent on the optical E-field given before t, it imposes a strong constraint between the real and imaginary part of the susceptibility and is mathematically given in the form of the Kramers–Kronig transform. By noticing that the SLM laser is a typical narrow-band device with its lasing frequency in the neighborhood of the reference optical frequency (ω_0) and the Kramers–Kronig relation is linear, one can readily express the refractive index change as a linear function of the material gain:

$$\Delta n(z,t,\omega_0) = \alpha_{\text{LEF}}\, g(z,t,\omega_0) \tag{30.27}$$

with α_{LEF} denoting the linewidth enhancement factor.

Finally, the carrier density inside the active region can be described by the following rate equation [16,36]:

$$\begin{aligned}\frac{\partial N(z,t)}{\partial t} &= \frac{I(z,t)}{qV} - [AN(z,t)+BN^2(z,t)+CN^3(z,t)]\\ &\quad -\frac{n_{\text{eff}}}{2\hbar\omega_0}\sqrt{\frac{\varepsilon_0}{\mu_0}}\frac{L\Gamma}{V}g(z,t,\omega_0)[|e^f(z,t)|^2+|e^b(z,t)|^2]\end{aligned} \tag{30.28}$$

where I represents the bias current; q the unit electron charge; V the active region volume; A, B, and C the minority carrier Shockley–Read–Hall (SRH); bimolecular and spontaneous emission, and Auger recombination coefficients, respectively; $\hbar$ Plank constant; $\sqrt{\varepsilon_0/\mu_0} \approx 1/377[\mathrm{S}]$ the vacuum admittance; and L the active region length (along the cavity direction). Given the fact that the active region of the laser diode has a low (unintentional) doping concentration and the potential drop across it is usually negligible, a quasi-neutrality condition holds and consequently the electrons and holes have the same density inside the active region—we therefore don't have to distinguish the electron and hole density and simply use the term "carrier density" to indicate both.

As the seed of the lasing process, the spontaneous emission contribution cannot be ignored at the beginning as otherwise Equation 30.22 becomes homogenous and only the trivial zero solution exists. Once the lasing starts, i.e., for the device operated under a bias beyond the threshold, however, the contribution from the spontaneous emission becomes negligible, which is reflected as a fact that the self-consistent solution of the above set of equations always converges to the same value regardless of the excitation method of the spontaneous emission, as long as the spontaneously emitted noise power is self-consistently described. The spontaneous emission contribution is usually assigned as a Gaussian distributed zero-mean random variable with its autocorrelation function normalized by the spontaneously emitted noise power [16]:

$$< |\tilde{s}(z,t)||\tilde{s}(z',t')| > = \frac{2\hbar\omega_0}{n_{\mathrm{eff}}}\sqrt{\frac{\mu_0}{\varepsilon_0}}\gamma\Gamma n_{sp} g(z,t,\omega_0)\delta(z-z')\delta(t-t') \tag{30.29}$$

where γ denotes the dimensionless coefficient of the coupling from the spontaneous emission to the entire spatial sphere and over the whole frequency spectrum to the waveguide mode at the reference optical frequency, n_{sp} the dimensionless ratio of the spontaneous emission to stimulated emission gain, and δ the Dirac function.

Equations 30.22, and 30.26 through 30.29 form a closed loop for one to find a set of self-consistent solution on the (minority) carrier density inside the active region (N), the material gain (g), and the slow-varying envelopes of the forward and backward going traveling wave field ($e^{f,b}$), for any given bias (I), with Equations 30.9, 30.8 and 30.24 presolved for a given SLM laser structure. Consequently, the optical field inside the laser can be found through Equation 30.23 and other physics quantities for characterizing the laser, such as the output optical power and the lasing wavelength, can readily be found [16,37]. An efficient solution technique for solving Equation 30.22 is to use the time-domain split-step method [38], with Equation 30.28 solved by the well-known Runge–Kutta method [39].

To describe the thermal effect, one still needs to add on the thermal diffusion equation [40] and to modify the empirical formulas for the material gain and refractive index change by considering their temperature dependence [16]. However, we will exclude the thermal description in this model to focus our study on the SLM operation aspect of semiconductor laser diodes.

30.2.5 Examples of SLM Laser Characteristics

By exploiting the aforementioned model, one can calculate the device performance for SLM DFB lasers with a few different types of grating structures as discussed in Section 30.2.3, shown in Figures 30.5 and 30.7. Figures 30.9 through 30.14 show the numerical simulation results of the output optical power and lasing wavelength as functions of the bias, the optical spectra at a fixed bias, the normalized optical field intensity and (minority) carrier density distributions along the cavity, and the small-signal intensity modulation responses for the purely refractive index-coupled uniform-grating DFB, partially gain/loss-coupled in/antiphase DFB, purely gain-coupled DFB, and QWPS DFB, respectively, with those device parameters involved in the model summarized in Table 30.1.

From Figure 30.9, one finds that while all DFB lasers have similar threshold current, the loss-coupled and QWPS DFB lasers have lower slope efficiency. This is expected for loss-coupled DFB lasers as their loss-coupling coefficient introduces an extra contribution equivalent to the non-interband optical modal

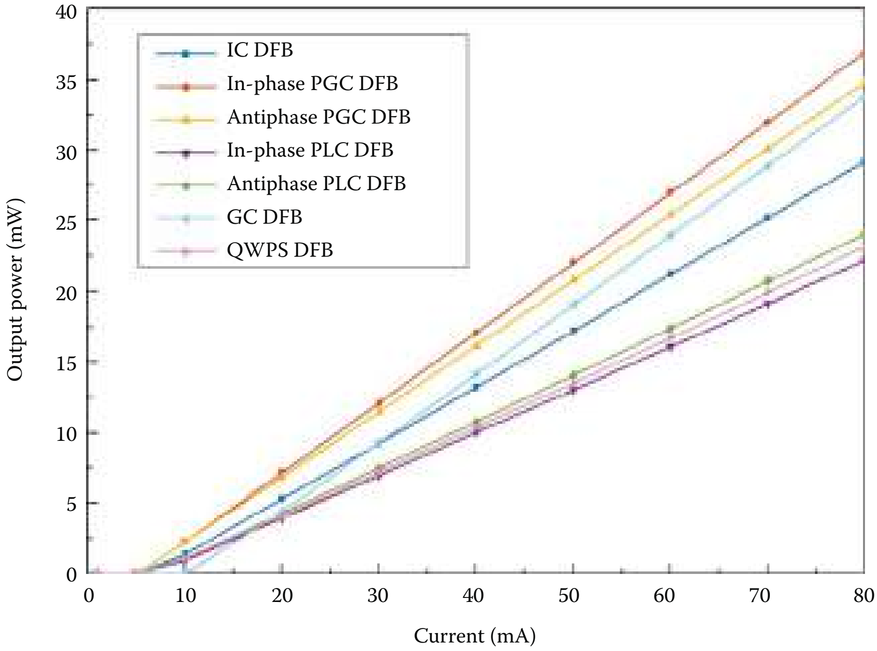


FIGURE 30.9 Output power–bias current curves for different DFB lasers. IC = (purely refractive) index-coupled (uniform grating), PGC = partially gain-coupled, PLC = partially loss-coupled, GC = (purely) gain-coupled, QWPS = quarter-wavelength phase-shifted.

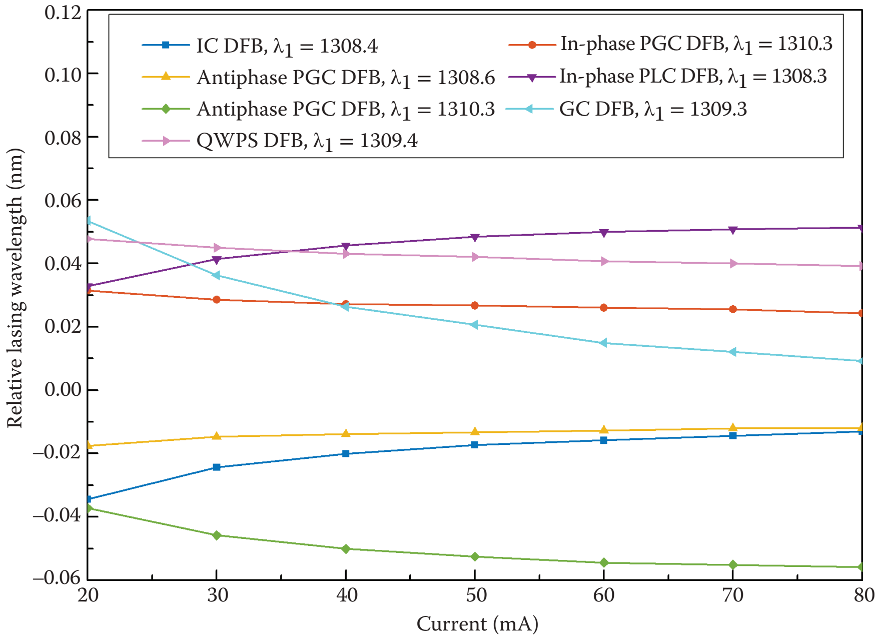


FIGURE 30.10 Relative lasing wavelength shift–bias current curves for different DFB lasers. IC = (purely refractive) index-coupled (uniform grating), PGC = partially gain-coupled, PLC = partially loss-coupled, GC = (purely) gain-coupled, QWPS = quarter-wavelength phase-shifted; the relative lasing wavelength shift is defined as λ-λ_l, with λ indicating the lasing wavelength at any bias current > 20 mA and λ_l the lasing wavelength at bias current = 20 mA for different DFB lasers under comparison.

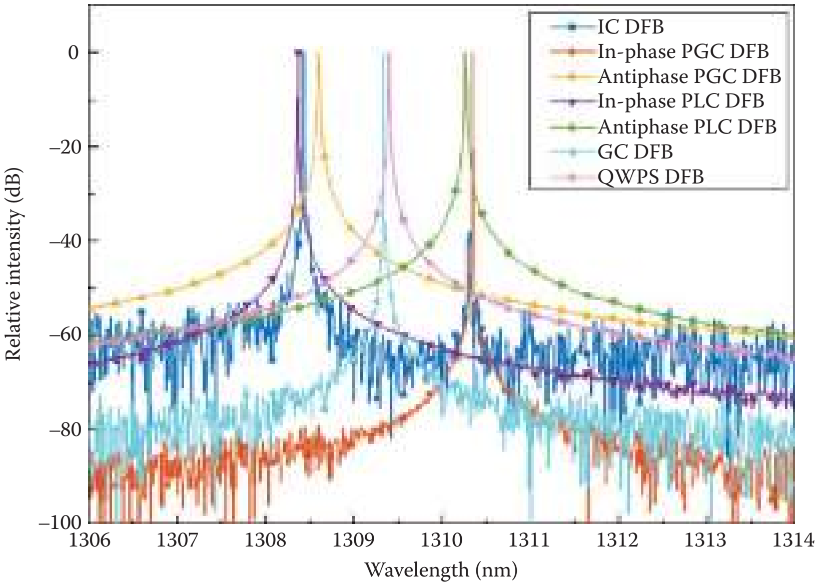


FIGURE 30.11 Lasing spectra for different DFB lasers. IC = (purely refractive) index-coupled (uniform grating), PGC = partially gain-coupled, PLC = partially loss-coupled, GC = (purely) gain-coupled, QWPS = quarter-wavelength phase-shifted.

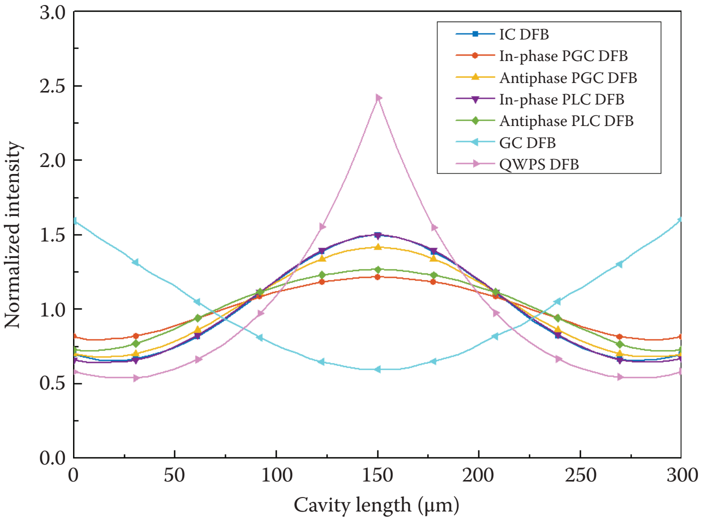


FIGURE 30.12 Normalized optical field intensity distributions for different DFB lasers. IC = (purely refractive) index-coupled (uniform grating), PGC = partially gain-coupled, PLC = partially loss-coupled, GC = (purely) gain-coupled, QWPS = quarter-wavelength phase-shifted.

loss (α_L). As shown in Table 30.1, the assumed non-interband optical modal loss is 10/cm, the normalized loss-coupling coefficient 0.5 would give an extra optical loss around 0.5/L = 0.5/(0.03 cm) ~17/cm. The total optical loss 27/cm indicates a significant increase as compared the original value in 10/cm. Since the slope efficiency is inversely proportional to the optical modal loss, higher loss would certainly cause a lower efficiency. The low slope efficiency of the QWPS DFB laser, however, is not caused by the high

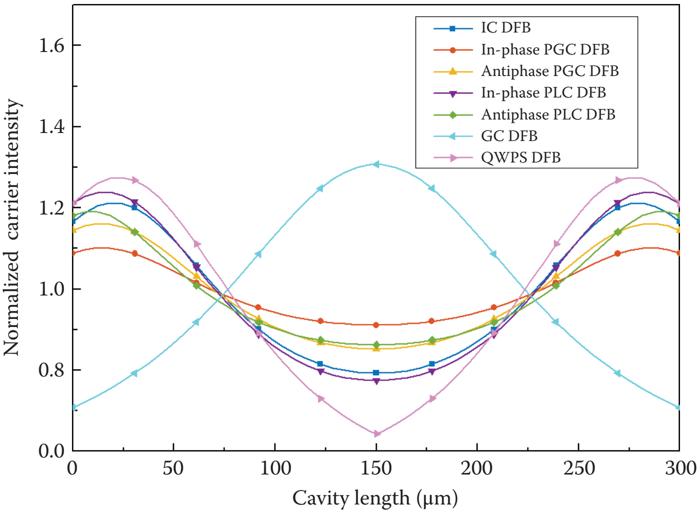


FIGURE 30.13 Normalized (minority) carrier density distributions for different DFB lasers. IC = (purely refractive) index-coupled (uniform grating), PGC = partially gain-coupled, PLC = partially loss-coupled, GC = (purely) gain-coupled, QWPS = quarter-wavelength phase-shifted.

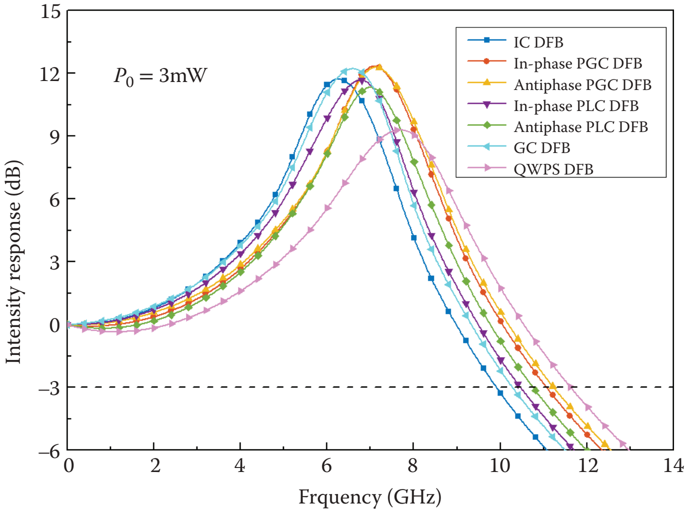


FIGURE 30.14 Small-signal intensity modulation responses for different DFB lasers. IC = (purely refractive) index-coupled (uniform grating), PGC = partially gain-coupled, PLC = partially loss-coupled, GC = (purely) gain-coupled, QWPS = quarter-wavelength phase-shifted.

optical modal loss, but by the low cavity loss. Since this structure has its field intensity mostly concentrated in the middle of the cavity, such a field pattern tends to retain more photons inside the cavity and consequently leads to a low output power by the low field distribution at both facets. This is consistent with our understanding on QWPS DFB as a high Q (quality)-factor laser with low cavity loss. Still because the slope efficiency is determined by the cavity loss over the summation of the cavity loss and the optical modal loss, apparently a low cavity loss will lead to a low slope efficiency once the optical modal loss is a nonvanishing positive value.

TABLE 30.1 Model Parameters

Model Parameter	Value
Cavity length L (μm)	300
Thickness of the quantum well (nm)	5
Number of quantum wells	6
Reflectivity of both laser facets	0, 0
Group index n_g	3.6
Effective index n_{eff}	3.2
Confinement factor Γ	0.06
Ridge width (μm)	2
Normalized index coupling coefficient	2.5
Normalized gain coupling coefficient	0.5
Normalized loss coupling coefficient	0.5
Grating period Λ (nm)	204.7
Gain coefficient a (cm^{-1})	1800
Transparent carrier density N_{tr} (10^{18} cm^{-3})	0.8
Linewidth enhancement factor α_{LEF}	−3
Nonlinear gain saturation factor ε (10^{-2} V^{-2})	9
SRH recombination coefficient A (10^9 s^{-1})	1
Bimolecular and spontaneous emission recombination coefficient B (10^{-10} $\text{cm}^{-3}\text{s}^{-1}$)	2
Auger recombination coefficient C (10^{-29} cm^{-6} s^{-1})	4
Non-interband modal loss α_L (cm^{-1})	10
Spontaneous emission coupling coefficient γ	10^{-4}
Spontaneous emission over stimulated emission ration_{sp}	1.7
Reference wavelength λ_0 (μm)	1.310

As ideal laser characteristics, the carrier density as well as the averaged gain should be clamped after lasing, but this is true only when the carrier density and optical field intensity both have uniform distributions along the cavity. In DFB lasers, however, the carrier density and optical field intensity have nonuniform distributions in "opposite" shapes as shown in Figure 30.12. Also, following the Schwartz nonequality, one knows the fact that the overlap integration between two functions with different shapes must be smaller than that between two functions in a similar shape. Under higher bias current, stronger LSHB makes the carrier density and optical field intensity distribution shapes more unlike. As a quantity directly in proportion to the overlap integration of the carrier density and optical field intensity distributions along the cavity, the averaged gain will be smaller. To maintain the lasing status, i.e., to balance the cavity loss plus the optical non-interband modal loss, the carrier density will have to increase with the bias current in order to compensate the "lost" averaged gain due to the enhanced LSHB. The increased carrier density will therefore cause a refractive index reduction of the active region following the Kramers–Kronig relation, which is simplified as Equation 30.27 with the negative linewidth enhancement factor (α_{LEF}). The reduction of the active region refractive index brings down the effective index of the waveguide and consequently makes the Bragg wavelength as well as the whole Bragg stop-band shift toward the shorter wavelength side, according to Equation 30.4 ($\lambda_B = 2n_{\text{eff}}\Lambda$). This effect is indeed reflected in Figure 30.10 by the blueshift with the increased bias current of the lasing wavelengths of the QWPS DFB and GC DFB lasers, located right in the center of the Bragg stop-band (or at the Bragg wavelength).

Figure 30.10 also shows that, without the thermal effect considered in simulation, the lasing wavelengths of the DFB lasers on the left (blue) side the Bragg stop-band (i.e., the IC DFB, the antiphase PGC DFB, and the in-phase PLC DFB; see Figure 30.11) have redshift toward the longer wavelength side with increased bias current, whereas the lasing wavelengths of the DFB lasers on the right (red) side (i.e., the in-phase PGC DFB and the antiphase PLC DFB; see Figure 30.11) take blueshift toward the shorter wavelength side. This is brought in by the shrinkage of the Bragg stop-band width due to the enhanced LSHB with the increased bias current, which further suggests an effective normalized coupling coefficient reduction or a more opened (lower Q-factor) cavity in accompanying with the stronger LSHB under higher injection level.

Figure 30.11 clearly shows that the QWPS and purely gain-coupled DFB lasers lase at the Bragg wavelength, whereas the in-phase and antiphase partially gain-coupled DFB lasers have their lasing wavelengths at the right (red) and left (blue) side of the Bragg stop-band, respectively. The in-phase and antiphase partially loss-coupled DFB lasers have their lasing wavelengths located the other way around, at the left (blue) and right (red) side of the Bragg stop-band, respectively. Finally, the purely refractive index-coupled DFB laser picks the left (blue) side wavelength rather randomly, with a relatively poor SMSR as expected. All these findings are consistent with previous analysis on these DFB lasers with different grating designs.

From Figure 30.12, one finds that the QWPS DFB laser has its field intensity mostly gathered in the middle of the cavity, as opposed to the purely gain-coupled DFB laser with its field intensity largely concentrated at both edges. The former cavity obviously bears the highest Q-factor or smallest cavity loss, whereas the latter one must correspond to the lowest Q-factor or highest cavity loss. Since the threshold current is proportional to the cavity loss, and the slope efficiency also increases with the cavity loss, the high Q-factor QWPS DFB laser bearing a low cavity loss has a low threshold current and a low slope efficiency. On the contrary, the low Q-factor purely gain-coupled DFB laser bearing a high cavity loss has a high threshold current as well as a high slope efficiency. This result has been confirmed by Figure 30.9 exactly. Other DFB lasers have their field intensities more evenly distributed and have their cavity losses in between. Consequently, they have their threshold current and slope efficiency values in between as shown in Figure 30.9. It is also worth mentioning that the sharp field turning in the middle of the QWPS DFB laser cavity is brought in by the half grating period shift exactly at the same location.

All carrier density distributions take the opposite shape as compared to their optical field intensity distributions. This result is expected as the carriers inside the cavity are consumed to generate photons, indicating a low carrier density wherever the optical field intensity is high, and vice versa.

Simulation result shows that there is no significant difference in terms of the 3 dB small-signal intensity modulation bandwidth among different DFB lasers once they are set to output the same optical power. This is quite different from the result obtained in References [41,42], in which the antiphase partially gain-coupled DFB laser shows a significantly broader 3 dB small-signal intensity modulation bandwidth in comparison with its in-phase counterpart. The inconsistency might come from the different selection on device parameters and further investigation is still needed before a final conclusion can be achieved.

30.3 Coupled-Cavity Single Longitudinal Mode Laser Diodes

30.3.1 A General Optical (Band-Pass) Filtering Model

The conventional FP cavity supports many closely packed longitudinal cavity modes with their spacing given in the form of Equation 30.2. For the edge-emitting laser with its cavity length (L) much longer than the center wavelength (λ_0), the mode spacing is much smaller than the full width at half maximum (FWHM) of the material gain peak. As such, multiple longitudinal modes can lase simultaneously, which makes a typical edge-emitting FP laser operate under the multiple longitudinal mode lasing scheme. Other than the aforementioned grating-assisted methods by introducing the wavelength-selective end reflectors (DBR), or by exploiting various DFB cavities, to achieve the SLM operation, one can still take an alternative approach by inserting an optical (band-pass) filter into the FP cavity to eliminate all other longitudinal cavity modes but leave only one for lasing. This approach is generally known as the coupled-cavity laser that forms the other category of SLM lasers, if we don't count in various external cavity SLM laser configurations.

For a coupled-cavity laser generally presented in the form of an FP cavity with an inserted OBPF shown in Figure 30.15, we have [43]:

$$\begin{bmatrix} A_{\text{out}} \\ B_{\text{out}} \end{bmatrix} = \frac{1}{T^2}\begin{bmatrix} 1 & -R \\ -R & 1 \end{bmatrix}\begin{bmatrix} e^{j\varphi_2} & 0 \\ 0 & e^{-j\varphi_2} \end{bmatrix}\frac{1}{t_{21}}\begin{bmatrix} t_{12}t_{21} - r_{12}r_{21} & r_{21} \\ -r_{12} & 1 \end{bmatrix}\begin{bmatrix} e^{j\varphi_1} & 0 \\ 0 & e^{-j\varphi_1} \end{bmatrix}\begin{bmatrix} 1 & R \\ R & 1 \end{bmatrix}\begin{bmatrix} A_{\text{in}} \\ B_{\text{in}} \end{bmatrix} \tag{30.30}$$

with $A_{\text{in,out}}$ and $B_{\text{in,out}}$ denoting the optical field amplitude of the forward and backward propagating waves, with their subscript "in" and "out" indicating the assumed input and output port on the left- and right-hand sides, respectively. One should note that the optical field is polarized in the cross-sectional plane perpendicular to the cavity direction. The arrows associated with them in Figure 30.15 indicate their propagation direction, not their polarization direction. Also in Equation 30.30, R and T indicate the amplitude reflectivity and transmissivity of the end facets when looking from inside of the FP cavity, $r_{12,21}$ the amplitude reflectivities when looking from the FP cavity to the OBPF on the left- and right-hand sides, $t_{12,21}$ the amplitude transmissivities from left to right and from right to left when staying inside the FP cavity, respectively. Finally, $\varphi_{1,2}$ are the phase delays of the wave traveling through the two sections ($L_{1,2}$) of the FP cavity separated by the OBPF.

Equation 30.30 can readily be simplified to

$$\begin{aligned} \begin{bmatrix} A_{\text{out}} \\ B_{\text{out}} \end{bmatrix} &= \frac{1}{t_{21}T^2}\begin{bmatrix} e^{j\varphi_2} & -Re^{-j\varphi_2} \\ -Re^{j\varphi_2} & e^{-j\varphi_2} \end{bmatrix}\begin{bmatrix} \delta & r_{21} \\ -r_{12} & 1 \end{bmatrix}\begin{bmatrix} e^{j\varphi_1} & Re^{j\varphi_1} \\ Re^{-j\varphi_1} & e^{-j\varphi_1} \end{bmatrix}\begin{bmatrix} A_{\text{in}} \\ B_{\text{in}} \end{bmatrix} \\ &\equiv \frac{1}{t_{21}T^2}\begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}\begin{bmatrix} A_{\text{in}} \\ B_{\text{in}} \end{bmatrix} \end{aligned}, \tag{30.31}$$

where

$$\delta \equiv t_{12}t_{21} - r_{12}r_{21} \text{ and } A_{22} \equiv e^{-j(\varphi_2+\varphi_1)} - r_{12}Re^{-j(\varphi_2-\varphi_1)} - r_{21}Re^{j(\varphi_2-\varphi_1)} - \delta R^2 e^{j(\varphi_2+\varphi_1)} \tag{30.32}$$

The lasing condition is therefore obtained as $A_{\text{in}} = B_{\text{out}} = 0$, but A_{out} and B_{in} are not zero, which requires $A_{22} = 0$, or:

$$[r_{12}e^{-j(\varphi_2-\varphi_1)} + r_{21}e^{j(\varphi_2-\varphi_1)}]Re^{j(\varphi_2+\varphi_1)} + \delta R^2 e^{2j(\varphi_2+\varphi_1)} = 1 \tag{30.33}$$

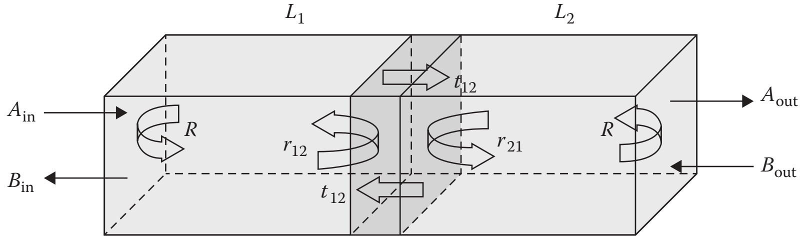


FIGURE 30.15 An illustrative diagram showing the FP cavity with an inserted optical band-pass filter (OBPF), with A_{in} and B_{in}, A_{out} and B_{out} indicating the optical field amplitudes of the forward and backward propagating waves at the input and output port, respectively; r_{12} and t_{12} the amplitude reflectivity and transmissivity from the left section to the inserted OBPF; r_{21} and t_{21} the amplitude reflectivity and transmissivity from the right section to the inserted OBPF; $L_{1,2}$ the lengths of the left and right section, respectively; R the end facet amplitude reflectivity of the FP cavity.

1. FP cavity
   It is obvious that once we remove the OBPF by letting $r_{12} = r_{21} = 0$ and $\delta = 1$, Equation 30.33 reduces to the well-known lasing condition of the FP laser:

$$R^2 e^{2j(\varphi_2+\varphi_1)} = 1 \tag{30.34}$$

2. Bidirectional symmetric filter
   Once the OBPF has the bidirectional symmetry, by letting

$$r_{12} = r_{21} = jre^{j\theta} \tag{30.35}$$

   and

$$t_{12} = t_{21} = te^{j\theta} \tag{30.36}$$

   we have

$$\delta = (t^2 + r^2)e^{2j\theta} \equiv (1 - l)e^{2j\theta} \tag{30.37}$$

   where θ and l stand for the single pass phase and loss of the filter, respectively. Hence we find from Equation 30.33:

$$2jrR\cos(\varphi_2 - \varphi_1)e^{j(\varphi_2+\varphi_1+\theta)} + (1 - l)R^2 e^{2j(\varphi_2+\varphi_1+\theta)} = 1 \tag{30.38}$$

   It is apparent that Equation 30.38 can be rewritten as

$$\bar{R}^2 e^{2j(\varphi_2+\varphi_1+\theta)} = 1, \text{ with } \bar{R} \equiv \sqrt{\frac{1-l}{1 - 2rR\cos(\varphi_2 - \varphi_1)e^{j(\varphi_2+\varphi_1+\theta+\pi/2)}}}R \tag{30.39}$$

   Unlike in Equation 30.34 where the amplitude reflectivity of the end facets R is a constant for a conventional FP laser, a coupled-cavity laser comprising an FP cavity with an inserted bidirectional symmetric filter has a wavelength-dependent effective amplitude reflectivity in its lasing condition, which, conceptually similar to the lasing condition of the DBR laser given as Equation 30.13, provides an extra wavelength selection mechanism to make a conventional FP laser single-moded. For example, if the inserted filter is lossless ($l = 0$), Equation 30.39 shows that the modification on the lasing condition from the original FP cavity is determined by the reflection (r) and length (θ) of the filter and its position inserted inside the FP cavity (φ_1 and φ_2). It is possible that, within the wavelength range set by the gain spectrum bandwidth, there is a single wavelength that makes $1 - 2rR\cos(\varphi_2 - \varphi_1)e^{j(\varphi_2+\varphi_1+\theta+\pi/2)} = 1 - 2|r|R$, and $2(\varphi_1 + \varphi_2 + \theta) = 2m\pi$, with m as an integer. Hence $\bar{R}$ will be the maximum with the round trip phase-matching condition satisfied at this wavelength. As a result, the lasing will only happen at this wavelength and such a general coupled-cavity structure supports SLM operation.

In the fabrication of real-world coupled-cavity lasers, the OBPF is usually formed by an equivalent FP etalon. The filter reflectivity and transmissivity in Equations 30.35 and 30.36 can then be derived from Equation 30.30 itself with the inserted filter removed from the FP cavity in Figure 30.15. By doing so, we find that Equation 30.30 reduces to

$$\begin{aligned}\begin{bmatrix} A'_{\text{out}} \\ B'_{\text{out}} \end{bmatrix} &= \frac{1}{T'^2}\begin{bmatrix} 1 & -R' \\ -R' & 1 \end{bmatrix}\begin{bmatrix} e^{j\varphi'} & 0 \\ 0 & e^{-j\varphi'} \end{bmatrix}\begin{bmatrix} 1 & R' \\ R' & 1 \end{bmatrix}\begin{bmatrix} A'_{\text{in}} \\ B'_{\text{in}} \end{bmatrix} \\ &= \frac{1}{T'^2}\begin{bmatrix} e^{j\varphi'}(1 - R'^2 e^{-2j\varphi'}) & 2jR'\sin\varphi' \\ -2jR'\sin\varphi' & e^{-j\varphi'}(1 - R'^2 e^{2j\varphi'}) \end{bmatrix}\begin{bmatrix} A'_{\text{in}} \\ B'_{\text{in}} \end{bmatrix}\end{aligned}$$

with the prime symbols added to the variables and parameters to differentiate them from those being used in Equation 30.30 for the most outside FP cavity in Figure 30.15. In the absence of the incident light coming from the left-hand side, $B'_{\text{out}} = 0$, we have

$$R_f \equiv \frac{B'_{\text{in}}}{A'_{\text{in}}} = j\frac{2R' \sin\varphi' e^{j\varphi'}}{1 - R'^2 e^{2j\varphi'}}, \text{ and } T_f \equiv \frac{A'_{\text{out}}}{A'_{\text{in}}} = \frac{T'^2 e^{j\varphi'}}{1 - R'^2 e^{2j\varphi'}}$$

Since the common factor $1/(1 - R'^2 e^{2j\varphi'})$ in abovementioned expressions can be expanded in a form of

$$1 + R'^2 e^{2j\varphi'} + (R'^2 e^{2j\varphi'})^2 + \ldots$$

which means that $j2R' \sin\varphi' e^{j\varphi'}$ and $T'^2 e^{j\varphi'}$ are the single pass reflectivity and transmissivity, respectively, Equations 30.35 and 30.36 are valid once we take

$$r = 2R' \sin\varphi' \tag{30.40}$$

$$t = T'^2 \tag{30.41}$$

and

$$\theta = \varphi' \tag{30.42}$$

Equations 30.35 through 30.37 and Equations 30.40 through 30.42 therefore link the model parameters to the design parameters of the physical structure.

30.3.2 Cleaved-Coupled-Cavity (C^3) Laser

A typical cleaved-coupled-cavity (C^3) laser has a structure shown in Figure 30.16.

Following the general coupled-cavity laser model derived in Section 30.3.1, we have

$$R' = (n_{\text{eff}} - 1)/(n_{\text{eff}} + 1) \tag{30.43}$$

$$T'^2 = 4n_{\text{eff}}/(n_{\text{eff}} + 1)^2 \tag{30.44}$$

$$\varphi' = 2\pi d/\lambda \tag{30.45}$$

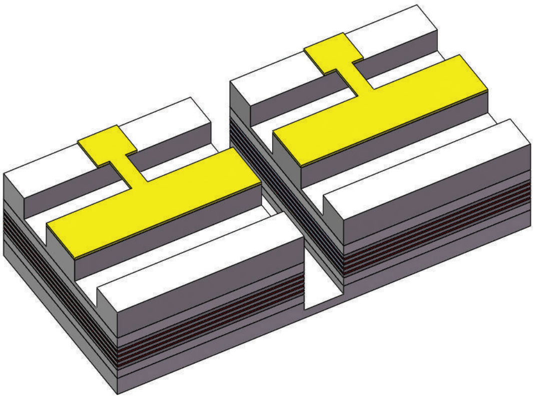


FIGURE 30.16 The C^3 laser structure.

with n_{eff} denoted as the laser waveguide effective index and d the air gap spacing between the pair of coupled cavities. By noting

$$\varphi_{1,2} = [2\pi n_{\text{eff}}/\lambda - j(g - \alpha_i)]L_{1,2} \tag{30.46}$$

and using Equations 30.40 through 30.42, one can express the lasing condition (Equation 30.39) in terms of the C^3 laser design parameters ($n_{\text{eff}}, d, g, \alpha_i, L_{1,2}$). The SLM lasing condition can therefore be found by a proper combination of the two cavity lengths ($L_{1,2}$) and the air gap spacing (d).

A major issue in the fabrication of the C^3 laser is the optical alignment between the two cavities, if the air gap will be formed by cleaving. One can certainly think of using the etching technology to form the air gap, the quality of the facet; however, cannot be guaranteed unless the chemically assisted ion beam etching (CAIBE) is exploited [44]. With CAIBE, the precise control of the air gap spacing becomes difficult, especially when the required spacing is no more than a few micrometers. An elegant approach to form the C^3 laser with a narrow air gap is the microcleaving technique [45]. In this technique, after the air gap section is defined, a selective wet (chemical) etching step is applied first to remove the material (InP) surrounding the active region (InGaAsP or AlGaInAs) inside the air gap, which leaves the latter as a thin bridge hanging over in the middle. The wafer or bar is then soaked inside some liquid and placed inside an ultrasonic cleaning bath for breaking down the active region bridge inside the air gap. The facet formed by this approach has a quality equivalent to the one obtained by the conventional cleaving technique. Since the etching of the air gap won't go very deep after passing though the active region, there is no optical alignment problem as the pair of coupled cavities still sits on top of the same substrate.

Despite its readiness in fabrication [46], the C^3 structure never becomes a popular SLM laser product for its low single-mode yield and the reliability concern. The low yield comes from the sensitive phase-matching condition, which requires precise control on cavity geometrical dimensions, including both cavity lengths and the air gap spacing, as well as the effective index of the waveguide. Since the latter depends not only on the material composition and cross-sectional geometrical dimension design, but also on operating conditions such as the ambient temperature and injection current, it naturally has a very low probability to hit the stable SLM operation condition with a high SMSR over an entire bias range. The reliability concern comes from the extra pair of facets appeared in the C^3 structure, especially when it is difficult to get them protected with dielectric coating layers, for the coating is not only difficult to be applied to the narrow air gap, but also changes the lasing condition.

Figure 30.17 shows the simulated optical spectrum of a typical SLM C^3 laser, with its structural parameter given in Table 30.2, and other parameters the same as those given in Table 30.1.

30.3.3 Etched Slotted Laser

An alternative version of the coupled-cavity structure is the etched slotted SLM laser [47–50] with its typical structure shown in Figure 30.18. The main difference lies in that the coupled cavities in such structure are formed by one or multiple etched slots across the waveguide without passing through the active region. As such, there is no optical alignment issue as the coupled-cavity device still stays as a whole piece. There is no reliability concern either as no extra active region cross section is exposed other than the usual end facets. The major drawback of this structure as compared to the C^3 laser is its much reduced reflectivity (r) of the equivalent OBPF as appeared in Equation 30.39 as the general lasing condition for coupled-cavity lasers. For an individual slot, its reflectivity is given by

$$R' = (n_{\text{eff}} - n'_{\text{eff}})/(n_{\text{eff}} + n'_{\text{eff}}) \tag{30.47}$$

with n_{eff} and n'_{eff} as the effective indices of the normal section and the waveguide inside the etched slot, respectively. As compared to the normal section, the etched slot only has its waveguide cladding layer

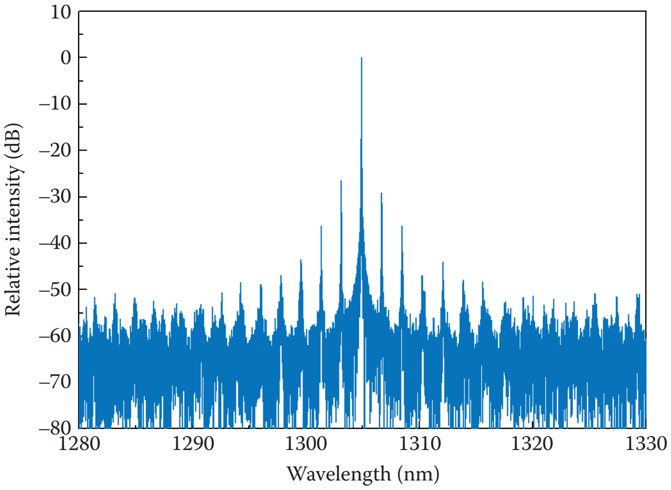


FIGURE 30.17 The simulated lasing spectrum of a typical SLM C^3 laser, with its parameters given in Table 30.2 and Table 30.1, respectively.

TABLE 30.2 C^3 Laser Parameters

Parameter	Value
Cavity lengths L_1, L_2 (μm)	136, 121 [46]
Gap between cavities d (μm)	5
Cleaved facet reflectivity r	0.565

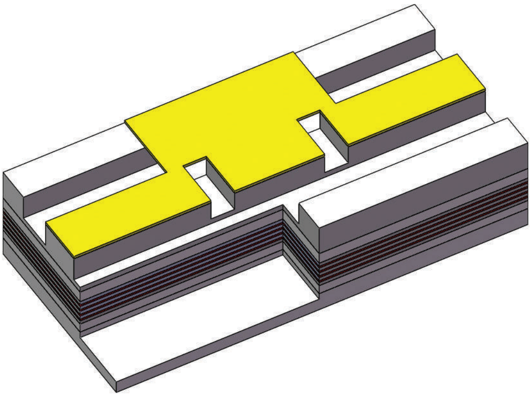


FIGURE 30.18 A typical etched slotted laser structure.

thinned to some extent, one then cannot expect a significant difference between n_{eff} and n'_{eff}. Therefore, R' given by Equation 30.47 is usually much smaller than the corresponding reflectivity in the C^3 laser given by Equation 30.43. From Equations 30.39 and 30.40, one knows that the contrast of the effective reflectivity $\bar{R}$ of the coupled-cavity laser at different wavelength will be reduce with a decreasing reflectivity R', Consequently, its ability on selecting the SLM drops significantly for the etched slotted laser.

To solve the poor mode selection problem, multiple etched slots have been introduced. A cascade of the structure in Figure 30.15 can be exploited to model such coupled-cavity laser with multiple etched slots.

A modified lasing condition similar to Equation 30.39 can also be derived with its effective reflectivity shown as a multiplication of multiple wavelength-dependent factors, with each of the factor corresponding to a single slot. As such, the wavelength spectral contrast of its effective reflectivity will be raised and its mode-selection ability will be improved. However, the multiple slotted coupled-cavity laser usually suffers an even lower single-mode yield, for the associated phase-matching condition becomes more complicated due to the introduction of multiple slots and the chance for it to be satisfied becomes rare.

30.3.4 Discrete Mode Laser

As shown in Figure 30.19, the discrete mode laser [51,52] can be viewed as a specific type of multiple slotted coupled-cavity laser with its slot's spacing properly designed to satisfy the phase-matching condition. As such, its single-mode yield can be significantly improved. With its general design rule given in Reference [52], the discrete mode laser is actually an SLM device between the grating-assisted laser and the coupled-cavity laser. It is apparent that, starting from a pair of coupled FP sections (i.e., the C^3 laser), as one increases the number of coupled FP sections with reduced reflectivities from section to section, and arranges the length of each FP section to satisfy certain phase-matching condition, one readily obtains various of grating-assisted lasers. As a special case, for a design with a uniform unit length from one FP section to another (Λ) selected to match the round trip phase-matching condition $4\pi n_{\text{eff}}\Lambda/\lambda = 2\pi$, one obtains a DFB laser with the first-order Bragg grating, as the uniform unit length design forms a periodic structure (i.e., a grating) with its periodicity (Λ) specified exactly the same as Equation 30.4, the phase-matching condition for the first-order Bragg grating. Therefore, the concept of the discrete mode laser actually bridges the grating-assisted lasers and the coupled-cavity lasers as the two main categories of SLM lasers.

30.3.5 New Aspects in Governing Equations and the Solution Technique

In general, the governing Equations 30.22 through 30.29 can still be used in modeling the coupled-cavity lasers. Since there is no grating involved, π/Λ should be replaced by $2\pi n_{\text{eff}}/\lambda$ with the grating coupling coefficients ($\kappa_{+,-}$) set to zero.

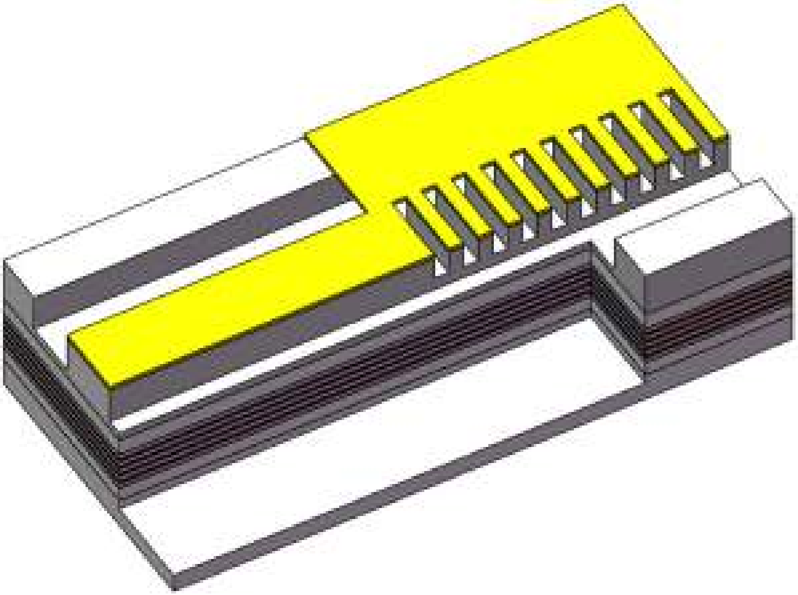


FIGURE 30.19 A typical discrete mode laser structure.

To treat the uniform sections in the coupled-cavity structure, the gain spectral dispersiveness must be considered to make sure that the lasing will happen in the neighborhood of the gain peak, as otherwise there will be no specific wavelength can be referred but the lasing cannot possibly happen at an arbitrary wavelength. Actually, the first term on the right-hand side of Equation 30.22 should be a time-domain convolution between the material gain and the field envelope, rather than be a product of them [53]. This term in its current form in Equation 30.22 is obtained under the assumption that the gain spectral bandwidth is much broader than the field envelope bandwidth, so that the time-domain gain performs as a Dirac δ-function in the convolution, which takes away the integration. For an SLM laser, the latter is usually determined by its linewidth or the modulation signal base bandwidth, whichever is broader, if the laser is under direct modulation. As the gain spectral bandwidth is usually in the range of 60 nm, whereas the field envelope bandwidth of an SLM laser usually is no more than 1 nm, the aforementioned assumption is indeed valid. To restore the gain dispersiveness in the coupled-cavity laser, however, this assumption has to be removed and a special solution technique will have to be introduced in dealing with the time-domain convolution between the material gain and the field envelope. An efficient way to treat the convolution in an initial value problem specified by a set of time-domain differential equations (or more accurately, a mixed initial-boundary value problem specified by a set of time-space domain partial differential equations) is the digital filtering method [16,54]. In this method, the material gain spectral profile is first calculated by the physics-based model within a limited wavelength range [16]. Its shape is then duplicated in the entire wavelength domain to form a periodic function, with nothing changed in its original wavelength range. The time-domain material gain can therefore found by taking the inverse Fourier transform of its spectral function. Since the latter function is turned into a periodic one, the time-domain gain therefore becomes a summation of a set of Dirac δ-functions, weighted by factors obtained as the corresponding coefficients in its Fourier expansion. Consequently, the convolution is reduced to a summation of a set of field envelopes, shifted in time domain and weighted by the said factors. For a given laser structure, these factors just need to be calculated once after the material gain spectral profile is obtained. The digital filtering method is, therefore, very efficient in handling the convolution in conjunction with the time-domain solution techniques, such as the split-step method, for solving Equation 30.22 [38,55].

Yet another problem often appearing in modeling the coupled-cavity laser is the involvement of the passive sections without injection. It is obvious that the material gain, the refractive index change, and the noise contribution in Equation 30.22 all need to be taken away for the passive sections. By noticing that Equation 30.22 become completely decoupled with the rest governing equations, one can solve these in frequency domain only for once to extract the reflectivity and transmissivity of the field envelope for these sections. In the rest (active) sections, the set of governing equations will be solved in time domain as described previously. Upon the propagating waves reach the boundary at the passive section, however, the reflected and transmitted field can be computed by taking the time-domain convolution between the incident field and the inverse Fourier transform of the reflectivity and transmissivity, respectively, as the reflectivity and transmissivity are both defined in frequency domain. Once again, the digital filtering approach can be employed in dealing with the time-domain convolution [56]. Since the reflectivity and transmissivity only need to be calculated once for a given structure, the convolution handled in such a way is very efficient. Generally, the mixed domain method, i.e., treating Equation 30.22 in time domain for active sections and in frequency domain for passive sections, is much more efficient compared with the full time-domain method, especially for passive units with long section lengths, because the frequency-domain treatment links the input and output fields of the passive section directly through transfer functions (i.e., the reflectivity and transmissivity, respectively) in a single step, without any step-by-step marching along the wave propagating direction as required by time-domain approaches.

30.3.6 Performance Comparisons among Coupled-Cavity SLM Lasers

The simulated performance of a few different types coupled-cavity SLM lasers is summarized in Table 30.3. A general tendency is that as the SMSR increases with the number of slots (i.e., the number of coupled

TABLE 30.3 Performance Comparison among Different Coupled-Cavity SLM Lasers

Laser Structure	SMSR (dB)	SLM Yield (SMSR> 15 dB)	SLM Yield (SMSR> 25 dB)
C^3 laser	12–32	60%	20%
Etched slotted laser (2 slots)	6–19	27%	0%
Discrete mode laser (5 slots)	11–29	73%	13%
Discrete mode laser (9 slots)	9–36	60%	40%
Discrete mode laser (19 slots)	7–49 (42, measured value in [52])	47%	33%

Note: The experimentally measured SMSR for a discrete mode laser with 19 slots is given in the brackets.

cavities), the SLM yield decreases. This result agrees with our conclusion drawn by the general OBPF model for the coupled-cavity structures, since the increased slot number enhances the wavelength spectral contrast of the effective reflectivity, thereby enhances the structure's mode selection ability, whereas it simultaneously introduces a more complicated phase-matching condition that is increasingly difficult to be satisfied.

30.4 Recent Development on Single Longitudinal Mode Laser Diode

30.4.1 Open Problems in SLM Lasers

Current SLM lasers are not yet ideal. Grating-assisted lasers generally suffer problems like high fabrication cost and highly sensitive to external feedback. The DBR laser needs the technology to monolithically integrate the active region with the passive grating section in different material compositions. The uniform-grating purely refractive index-coupled DFB laser has a relatively low single-mode yield, whereas the QWPS DFB laser has even poorer immunity to external feedback and wastes half of its output power, and complex-coupled DFB lasers usually have reliability issues. A variety of the grating-assisted laser structures have been proposed, but all as compromised approaches with none of them seeming to be able to solve all the aforementioned problems. Coupled-cavity lasers usually have low fabrication cost and less sensitive to external feedback. However, they all have low single-mode yield and suffer relatively low SMSR. Therefore, it still remains as an open problem on how to obtain reliable SLM lasers with high single-mode yield, high SMSR, high immunity to external feedback, and low fabrication cost.

In the past decade, some progress has been made toward the final solving of the remaining problems in SLM lasers, although most of the recent efforts on laser development have their targets on new applications found in data communication systems and telecommunication access networks with emphases on multiple wavelength accessibility and tunability, and high-speed direct modulation. Effort has also been put on developing SLM lasers with narrow linewidth to meet the strong demand in high-speed long-haul coherent telecommunication systems in advanced modulation-detection schemes with higher spectral efficiency. A thorough description on all these new developments is beyond the scope of this chapter, so that we will only give a brief introduction on the new structures that are closely relevant to SLM lasers.

30.4.2 Bragg Waveguide SLM Laser

A schematic structure of the Bragg waveguide SLM laser [57,58] is shown in Figure 30.20. Unlike the conventional SLM lasers that all utilize their cavity structure to select the single lasing mode, this structure exploits the waveguide itself to eliminate the unwanted modes in an FP cavity.

The working principle of the Bragg waveguide can be understood by a simplified model described in Figure 30.21. By exploiting the effective index method, one reduces the wave vector of an arbitrary plane traveling wave into a two-dimensional plane as shown in Figure 30.21. It can then be decomposed into a pair of orthogonal components, with the one along the waveguide direction (β) as the required wave

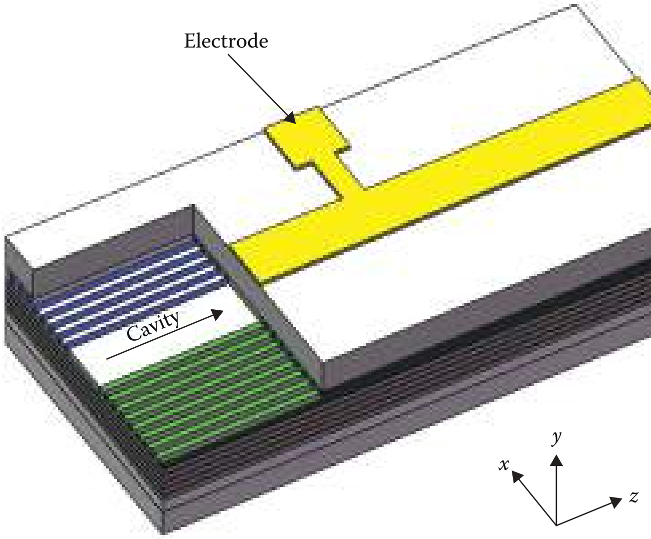


FIGURE 30.20 The Bragg waveguide SLM laser structure.

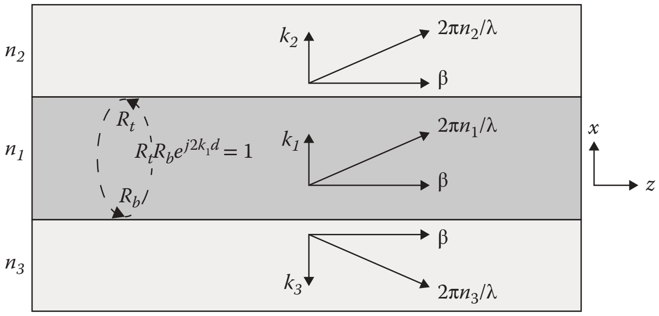


FIGURE 30.21 An illustrative diagram showing the concept of the dielectric waveguide, with $2\pi n_{1,2,3}/\lambda$ indicating the conceptual free-propagating plane wave numbers inside the core, the top cladding, and the bottom cladding layer, respectively, β their corresponding wave vectors' projection along the propagation (z) direction, $k_{1,2,3}$ their corresponding wave vectors' projections along the cross (x) direction in the core, the top cladding, and the bottom cladding layer, respectively, $R_{t,b}$ the amplitude reflectivities between the core and the top cladding layer, and between the core and the bottom cladding layer, respectively; $R_tR_be^{j2k_1d} = 1$ gives the resonance condition along the cross direction with d as the core layer thickness, which stands as the necessary and sufficient condition for guided waves in general waveguides.

propagation constant, the other along the cross direction (k_1) as the wave vector needs to be confined. For the wave component that propagates in the cross direction, a round trip travel would make the initial field E_0 become $E_0R_tR_be^{j2k_1d}$, with $R_{t,b}$ denoting the amplitude reflectivity between the core and the top cladding layer, and the core and the bottom cladding layer, respectively, and d the core layer thickness. Therefore, once the resonance happens in the cross direction, i.e.,

$$R_tR_be^{j2k_1d} = 1 \tag{30.48}$$

the field distribution along the cross direction remains the same at any cut along the waveguide direction, which means the wave is guided and propagates with β along the waveguide. This is actually a necessary and sufficient condition for a general waveguide, i.e., if and only if the wave resonates in the cross section, it will be guided by the waveguide. Consequently, the guided wave will form a standing wave pattern inside the core area of the waveguide, and will be either zero (for metallic waveguide) or evanescent (for dielectric waveguide) outside of the core. The guided wave's cross-sectional distribution doesn't change as the wave propagates along the waveguide. The only change on the wave as it propagates is its phase, scaled by $\beta z - \omega t$, with z and t indicating the space coordinate along the waveguide (i.e., in the wave propagation direction) and the time variable, respectively.

For the conventional dielectric waveguide that utilizes the total internal reflection effect, one finds that

$$k_{1,2,3} = \sqrt{n_{1,2,3}^2(2\pi/\lambda)^2 - \beta^2} \tag{30.49}$$

for the wave vector components along the cross direction in the core (with subscript 1), the top cladding (with subscript 2), and the bottom cladding (with subscript 3) layer, respectively, with $n_{1,2,3}$ denoting the refractive indices of these three layers. The propagation constant β, i.e., the wave vector component along the propagation direction (z), must be the same for the same guided wave distributed in different layers, as otherwise the guided wave would be split apart along with its propagating in z. Under the total internal reflection scheme, the refractive indices are chosen in such a way to make k_1 real ($2\pi n_1/\lambda > \beta$) but $k_{2,3}$ imaginary ($2\pi n_{2,3}/\lambda < \beta$). By noticing that $R_{t,b} = (k_1 - k_{2,3})/(k_1 + k_{2,3})$ for the TE wave with its electric field polarized in parallel with, and $R_{t,b} = (1/k_1 - 1/k_{2,3})/(1/k_1 + 1/k_{2,3})$ for the TM wave with its electric field polarized perpendicularly to, the boundary between different layers, respectively, one has $|R_{t,b}| = 1$ as the reflectivity at the boundary always takes the form of $(a - jb)/(a + jb)$. Consequently, the resonant condition (Equation 30.48) in the cross direction is reduced to

$$\varphi_t + \varphi_b + 2k_1 d = 2m\pi \tag{30.50}$$

with $\varphi_{t,b}$ indicating the phase of $R_{t,b}$, i.e.,

$$\varphi_{t,b} = -2\arctan(|k_{2,3}|/k_1) \tag{30.51}$$

for the TE wave and

$$\varphi_{t,b} = -2\arctan(k_1/|k_{2,3}|) \tag{30.52}$$

for the TM wave, respectively. Equations 30.49 through 30.51 or Equation 30.52 defines the dispersion relation of the guided wave, which tells that, for a given waveguide structure and operating wavelength λ below a maximum value known as the cutoff wavelength, one always manage to find at least one real β within the range between $2\pi\max(n_2, n_3)/\lambda$ and $2\pi n_1/\lambda$ as the solution. Therefore, the conventional dielectric waveguide has no wavelength selection ability as the dispersion relation of its guided wave shows a high-pass filter feature that supports the wave propagation with any wavelength shorter than the cutoff value, which also echoes the same statement we made at the beginning of this chapter. It is not surprising for one to reach this conclusion, once one notices that as the total internal reflection condition is satisfied, $|R_t R_b|$ is always 1 regardless of the associated wavelength. The cross-directional resonant condition, as the indication of the wave being guided, is therefore reduced to a phase-matching condition (i.e., the dispersion relation) that is too loose to select a discrete set of wavelengths, not to mention a single wavelength.

If, however, a grating-based reflection is utilized to substitute the total internal reflection, as shown in the structure given by Figure 30.20, it will be possible to make $R_t R_b$ wavelength dependent. Once not only the phase-matching condition in the form of Equation 30.50, but also a wavelength-dependent amplitude condition in the form of $|R_t(\lambda)R_b(\lambda)| = 1$, will jointly be derived from the general resonant condition

(Equation 30.48), it is possible that only a single wavelength can be found for the guided wave to satisfy both constraints imposed on the phase and amplitude.

The Bragg waveguide laser is such a device that exploits the grating reflection, rather than the total internal reflection, in its waveguide in conjunction with the FP cavity to reach the SLM operation. With a structure similar to the DBR laser arranged in the cross direction perpendicular to the waveguide for wave confinement, even a simple FP cavity makes an SLM laser [58]. A unique feature of the Bragg waveguide lies in that it allows its core layer to have a lower refractive index than that of its cladding. As such, the selection on the gain medium to build the active region for the Bragg waveguide laser will be more flexible.

30.4.3 Double-Trench Resonant Tunneling SLM Laser

As concluded by Section 30.3, the concept of the coupled cavity can generally be understood as the insertion of an OBPF into a conventional FP cavity, for purifying the lasing spectrum by eliminating all unwanted FP cavity modes. An accompanying issue is that both the transmitted light passed through the OBPF and the reflected light experienced the complementary optical band-reject filter (OBRF) stay inside the FP cavity, which add up to give no appreciable mode selection mechanism once their gains are the same, since the added spectrum of the complementary OBPF and OBRF turns out to be flat. Although the additionally introduced phase condition by the equivalent filter and any possible gain discrimination between the transmitted and reflected waves will help to eliminate the unwanted FP cavity modes, the single-mode yield and SMSR for such structures are usually low. And the SLM lasing is usually not very stable in the entire laser operating range. For example, the lasing mode often hops or the SMSR deteriorates as the bias current or ambient temperature changes. This can be attributed to the phase-sensitive lasing condition shown as Equation 30.33 or Equation 30.39, since the phase not only varies with any geometrical dimension deviation of the structure in the fabrication process, but also changes with the laser operating condition.

To ensure the SLM operation with an intensified SMSR with high yield, one needs to effectively eliminate the reflected light from the aforementioned OBPF inside the FP cavity. As shown in Figure 30.22, a double-trench structure has been proposed [59] to introduce dual reflections that cancel out the reflected light at a specific wavelength determined by the gap between the two trenches, an effect known as the resonant tunneling. As such, the transmission spectrum of the double trenches resembles that of an OBPF, which selects one of the many cavity modes in a conventional FP laser for single-mode operation. The most important design in this structure, which significantly improves the SLM laser performance, is to make the double trenches slanted so that the reflected waves with unselected wavelengths by the OBPF will escape

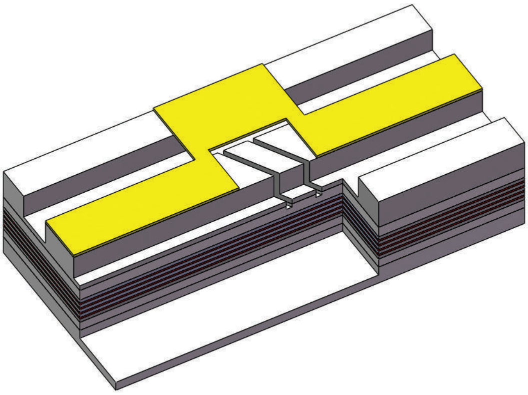


FIGURE 30.22 The double-trench resonant tunneling SLM laser structure.

from the cavity, in order to avoid jeopardizing the single-mode lasing condition, as otherwise the reflected and transmitted light in different wavelengths will compete to each other inside the cavity.

By eliminating the reflection brought in by the OBPF made of the double trenches, i.e., by letting $r_{12}, r_{21} \sim 0$, $\delta = t_{12}t_{21}$, the lasing condition (Equation 30.33) derived for general coupled-cavity lasers reduces to

$$t_{12}t_{21}R^2e^{2j(\varphi_2+\varphi_1)} = 1 \tag{30.53}$$

which indicates that the lasing will be determined by the transmission spectrum of the inserted OBPF. More specifically, the lasing will happen at the wavelength in the neighborhood of the transmission spectrum ($|t_{12}t_{21}|$) peak. Actually, once the transmission spectrum of the filter resembles that of an OBPF, the lasing condition shown in Equation 30.53 is the same as that of the DBR laser (Equation 30.13).

The simulated and measured lasing spectrum and SLM yield of a typical double-trench resonant tunneling SLM laser are given in Figures 30.23 and 30.24, respectively. The simulation result also shows, and experimental data verifies that, with varying positions of the double trenches inside the FP cavity, the slanted trench pair can always select and lock in one and only one of the many FP modes in the full laser operating range [59]. This can be attributed to the fact that, with the reflection from the double trenches effectively eliminated and consequently with the lasing condition determined by Equation 30.53 rather than Equation 30.33 or Equation 30.39 for the device, its mode selection mechanism is no longer phase sensitive. Hence the precise control of the geometrical dimension in the structure is not required in the fabrication process for such a device, which greatly enhances the single-mode yield, the SMSR, and the single-mode stability in operation, as compared to other existing coupled-cavity lasers.

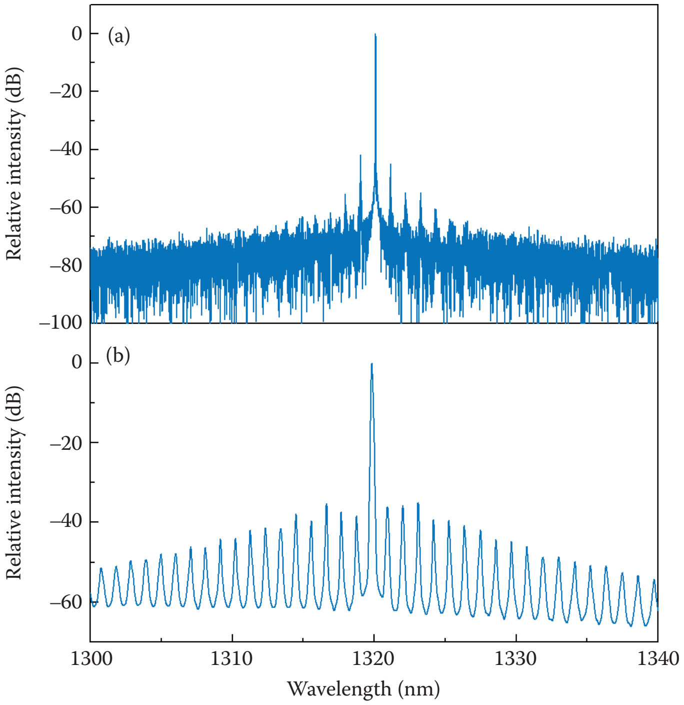


FIGURE 30.23 The (a) simulated and (b) measured lasing spectra of a typical double-trench resonant tunneling SLM laser, with its design parameters given in Reference [59].

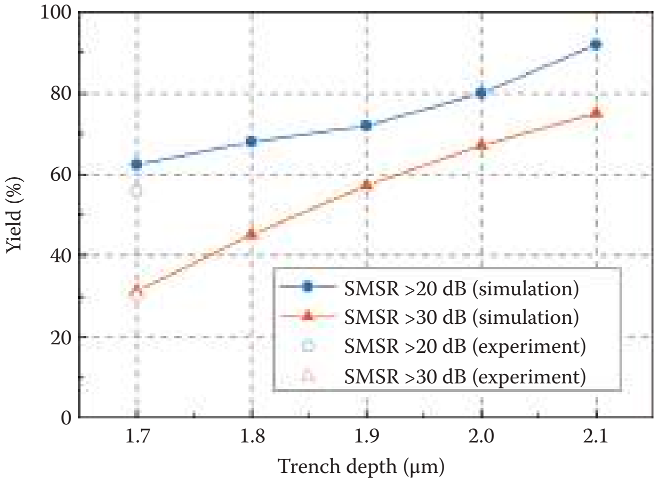


FIGURE 30.24 The SLM yield as a function of the trench depth, with device design parameters given in Reference [59].

30.4.4 Horn Ridge Waveguide (HRW) DFB Laser

With its structure shown in Figure 30.25, the horn ridge waveguide (HRW) DFB laser [60] improves the single-mode yield by breaking the inherent dual-mode degeneracy in the uniform-grating purely refractive index-coupled structure.

In a straight-waveguide purely refractive index-coupled DFB laser with uniform grating, the two longitudinal modes with their respective wavelengths located at each side of the Bragg stopband have the same reflection everywhere along the cavity due to the symmetry in the reflection spectrum, so that they have the degenerated field distribution. The longitudinal modes at the two wavelengths with the degenerated field distribution will then have the same modal gain, which causes either dual-mode operation or SLM lasing with poor SMSR if the degeneracy is somehow removed by the uncontrollable LSHB effect. In the HRW DFB laser, however, an effective chirp to the uniform grating is introduced. As such, the field of the longitudinal mode with the wavelength on the blue side will concentrate in the right section with the wider ridge width, as a consequence of the strong reflection it sees from the left section with the narrow ridge width, where the local Bragg wavelength takes a relative shift toward the shorter wavelength side due to the reduction on the local effective index. On the contrary, the field of the longitudinal mode with the wavelength on the red side will concentrate in the left section with the narrower ridge width, as a consequence of the strong reflection it sees from the right section with the wider ridge width, where the local Bragg wavelength shifts relatively toward the longer wavelength side due to the enhancement on the local effective index. Therefore, the major effect of the effectively chirped grating is to remove the degeneracy of field distributions of the two symmetrical modes on each sides of the Bragg stopband. With the help of such effectively chirped grating, the two longitudinal modes at the Bragg stop-band will take different distributions and have their associated field intensities spatially localized in different sections along the cavity. With the further help of the nonuniform modal gain simultaneously generated by the horn waveguide, the field distribution having a larger overlap with the modal gain turns out to be the only dominant lasing mode. In this particular structure, the modal gain is higher on the wide ridge side due to the larger confinement factor associated. Hence the wavelength at the blue side of the Bragg stop-band, i.e., the shorter wavelength, will be selected to lase, whereas the wavelength at the red side, i.e., the longer wavelength, will be effectively suppressed.

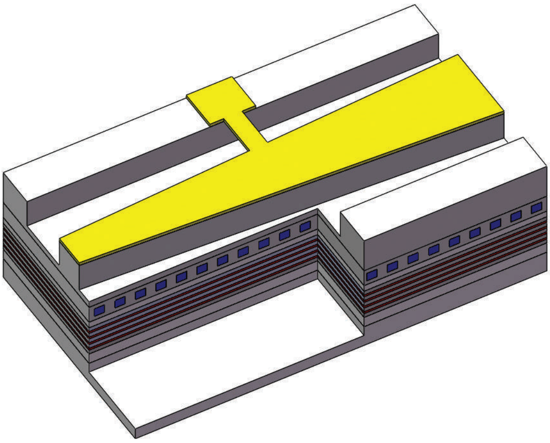


FIGURE 30.25 The horn ridge waveguide (HRW) DFB laser structure.

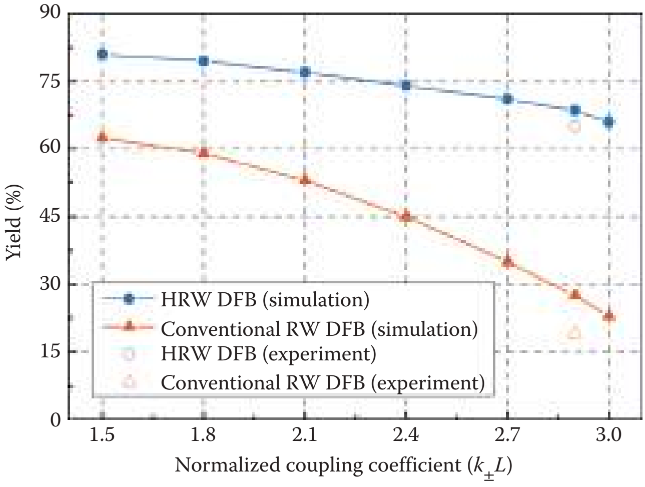


FIGURE 30.26 Comparison on the SLM yield (SMSR >30 dB) between the HRW DFB and the conventional ridge waveguide (RW) DFB with different κ ± *L*; other device design parameters are given in Reference [60].

Simulation result shows, and experimental result demonstrates, that the SLM yield can be drastically raised (by more than threefolds) with an optimized HRW design, especially for the uniform-grating purely refractive index-coupled DFB laser with high normalized coupling coefficient (κ*L*) [60]. This is further evidenced by the comparison result in Figure 30.26. Finally, measured lasing spectra of a typical HRW DFB laser under different bias current are shown in Figure 30.27.

Acknowledgment

The author would like to express his sincere thankfulness to Cheng Ke for his great help on performing most of the numerical calculations and on drawing all the figures.

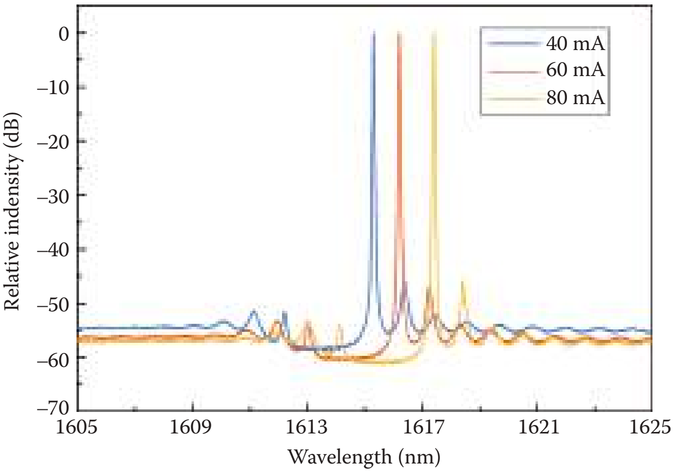


FIGURE 30.27 Measured lasing spectra of a typical HRW DFB laser under different bias current, with its design parameters given in Reference [60].

References

1. K. Ogawa, Analysis of mode partition noise in laser transmission systems, *IEEE Journal of Quantum Electronics*, vol. 18, no. 5, pp. 849–855, May 1982.
2. M. Ahmed and M. Yamada, Theoretical analysis of mode-competition noise in modulated laser diodes and its influence on the noise performance of fibre links, *Journal of Physics D Applied Physics*, vol. 45, no. 40, pp. 4172–4181, Sept. 2012.
3. M. L. Calvo and V. Lakshminarayanan, *Optical Waveguides: From Theory to Applied Technologies*, Boca Raton, FL: CRC Press, Taylor & Francis, ISBN 978-1-57444-698-2, Jan. 2007.
4. D. Marcuse, *Theory of Dielectric Optical Waveguides*, New York, NY: Academic Press, ISBN 0-12-470950-8, 1974.
5. K. Petermann, *Laser Diode Modulation and Noise*, Netherlands: Kluwer Academic Publishers, ISBN 978-0-7923-1204-8, Apr. 1988.
6. B. Mroziewicz, External cavity wavelength tunable semiconductor lasers—a review, *Opto-Electronics Review*, vol. 16, no. 4, pp. 347–366, Sept. 2008.
7. K. Vahala, K. Kyuma, A. Yariv, S. K. Kwong, M. C. Golomb, and K. Y. Lau, Narrow linewidth, single frequency semiconductor laser with a phase conjugate external cavity mirror, *Applied Physics Letters*, vol. 49, no. 23, pp. 1563–1565, Oct. 1986.
8. M. W. Fleming and A. Mooradian, Spectral characteristics of external-cavity controlled semiconductor lasers, *IEEE Journal of Quantum Electronics*, vol. 17, no. 1, pp. 44–59, Jan. 1981.
9. H. Kogelnik and C. V. Shank, Stimulated emission in a periodic structure, *Applied Physics Letters*, vol. 18, no. 4, pp. 152–154, Nov. 1971.
10. S. Wang, Principles of distributed feedback and distributed Bragg-reflector lasers, *IEEE Journal of Quantum Electronics*, vol. 10, no. 4, pp. 413–427, Apr. 1974.
11. J.-W. Mu and W.-P. Huang, Simulation of three-dimensional waveguide discontinuities by a full-vector mode-matching method based on finite-difference schemes, *Optics Express*, vol. 16, no. 22, pp. 18152–18163, Oct. 2008.
12. K. Jiang and W.-P. Huang, Finite-difference-based mode-matching method for 3-D waveguide structures under semivectorial approximation, *Journal of Lightwave Technology*, vol. 23, no. 12, pp. 4239–4248, Dec. 2005.

13. M. Yamada and K. Sakuda, Analysis of almost-periodic distributed feedback slab waveguides via a fundamental matrix approach, *Applied Optics*, vol. 26, no. 16, pp. 3474–3478, Aug. 1987.
14. T. Makino, Transfer-matrix analysis of the intensity and phase noise of multisection DFB semiconductor lasers, *IEEE Journal of Quantum Electronics*, vol. 27, no. 11, pp. 2404–2414, Nov. 1991.
15. G. P. Agrawal and N. K. Dutta, *Semiconductor Lasers*, 2nd ed. New York, NY: Van Nostrand Reinhold, ISBN 0-442-01102-4, 1993.
16. X. Li, *Optoelectronic Devices: Design, Modeling, and Simulation*, Cambridge: Cambridge University Press, ISBN 978-0-521-87510-3, Aug. 2009.
17. W. W. Ng, C. Hong, and A. Yariv, Holographic interference lithography for integrated optics, *IEEE Transactions on Electron Devices*, vol. 25, no. 10, pp. 1193–1200, Oct. 1978.
18. C. J. Armistead, B. R. Butler, S. J. Clements, A. J. Collar, D. J. Moule, S. A. Wheeler, M. J. Fice, and H. Ahmed, DFB ridge waveguide lasers at $\lambda = 1.5$ μm with first-order gratings fabricated using electron beam lithography, *Electronics Letters*, vol. 23, no. 11, pp. 592–593, May 1987.
19. S. Y. Chou, P. R. Krauss, and P. J. Renstrom, Imprint of sub-25 nm vias and trenches in polymers, *Applied Physics Letters*, vol. 67, no. 67, pp. 3114–3116, Sept. 1995.
20. L.-A. Wang, C.-H. Lin, and J.-H. Chen, Fabrication of sub-quarter-micron grating patterns by employing DUV holographic lithography, *Microelectronic Engineering*, vol. 46, no. 1–4, pp. 173–177, Aug. 1999.
21. A. Takemoto, Y. Ohkura, Y. Kawama, Y. Nakajima, T. Kimura, N. Yoshida, S. Kakimoto, and W. Susaki, 1.3-μm distributed feedback laser diode with a grating accurately controlled by a new fabrication technique, *Journal of Lightwave Technology*, vol. 7, no. 12, pp. 2072–2077, Dec. 1989.
22. Y. Abe, K. Kishino, Y. Suematsu, and S. Arai, GaInAsP/InP integrated laser with butt-jointed built-in distributed-Bragg-reflection waveguide, *Electronics Letters*, vol. 17, no. 25, pp. 945–947, Dec. 1981.
23. S. Murata, I. Mito, and K. Kobayashi, Spectral characteristics for 1.5 μm DBR laser with frequency-tuning region, *IEEE Journal of Quantum Electronics*, vol. 23, no. 6, pp. 835–838, June 1987.
24. J. E. Carroll, J. Whiteaway, and D. Plumb, *Distributed Feedback Semiconductor Lasers*, London: Institution of Electrical Engineers, ISBN 0-85296-917-1, Apr. 1998.
25. D. E. Muller, A method for solving algebraic equations using an automatic computer, *Mathematical Tables and Other Aids to Computation*, vol. 10, no. 56, pp. 208–215, Oct. 1956.
26. K. David, G. Morthier, P. Vankwikelberge, R. G. Baets, T. Wolf, and B. Borchert, Gain-coupled DFB lasers versus index-coupled and phase shifted DFB lasers: A comparison based on spatial hole burning corrected yield, *IEEE Journal of Quantum Electronics*, vol. 27, no. 6, pp. 1714–1723, June 1991.
27. H. Su, L. Zhang, A. L. Gray, R. Wang, T. C. Newell, K. J. Malloy, and L. F. Lester, High external feedback resistance of laterally loss-coupled distributed feedback quantum dot semiconductor lasers, *IEEE Photonics Technology Letters*, vol. 15, no. 11, pp. 1504–1506, Nov. 2003.
28. G.-P. Li, T. Makino, R. Moore, N. Puetz, K. Leong, and H. Lu, Partly gain-coupled 1.55 μm strained-layer multiquantum-well DFB lasers, *IEEE Journal of Quantum Electronics*, vol. 29, no. 6, pp. 1736–1742, Jun. 1993.
29. Y. Luo, Y. Nakano, K. Tada, T. Inoue, H. Hosomatsu, and H. Iwaoka, Purely gain-coupled distributed feedback semiconductor lasers, *Applied Physics Letters*, vol. 56, no. 17, pp. 1620–1622, Feb. 1990.
30. Y. Nakano, Y. Luo, and K. Tada, Facet reflection independent, single longitudinal mode oscillation in a GaAlAs/GaAs distributed feedback laser equipped with a gain-coupling mechanism, *Applied Physics Letters*, vol. 55, no. 16, pp. 1606–1608, Aug. 1989.
31. H. A. Haus and C. V. Shank, Antisymmetric taper of distributed feedback lasers, *IEEE Journal of Quantum Electronics*, vol. 12, no. 9, pp. 532–539, Sept. 1976.
32. E. Kapon, A. Hardy, and A. Katzir, The effect of complex coupling coefficients on distributed feedback lasers, *IEEE Journal of Quantum Electronics*, vol. 18, no. 1, pp. 66–71, Jan. 1982.
33. K. Utaka, S. Akiba, K. Sakai, and Y. Matsushima, λ/4-shifted InGaAsP/InP DFB lasers, *IEEE Journal of Quantum Electronics*, vol. 22, no. 7, pp. 1042–1051, July 1986.

34. S. Akiba, M. Usami, and K. Utaka, 1.5-μmλ/4-shifted InGaAsP/InP DFB lasers, *Journal of Lightwave Technology*, vol. 5, no. 11, pp. 1564–1573, Nov. 1987.
35. S.-L. Chuang, *Physics of Photonic Devices*, 2nd ed. New York, NY: Wiley, ISBN 9781118585658, 2012.
36. A. D. Sadovnikov, X. Li, and W.-P. Huang, A two-dimensional DFB laser model accounting for carrier transport effects, *IEEE Journal of Quantum Electronics*, vol. 31, no. 10, pp. 1856–1862, Oct. 1995.
37. X. Li, A. D. Sadovnikov, W.-P. Huang, and T. Makino, A physics-based three-dimensional model for distributed feedback laser diodes, *IEEE Journal of Quantum Electronics*, vol. 34, no. 9, pp. 1545–1553, Sept. 1998.
38. B. Kim, Y. Chung, and J. Lee, An efficient split-step time-domain dynamic modeling of DFB/DBR laser diodes, *IEEE Journal of Quantum Electronics*, vol. 36, no. 7, pp. 787–794, July 2000.
39. W. H. Press, S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes: The Art of Scientific Computing*, 3rd ed. Cambridge: Cambridge University Press, ISBN 978-0-511-33555-6, Sept. 2007.
40. X. Li and W.-P. Huang, Simulation of DFB semiconductor lasers incorporating thermal effects, *IEEE Journal of Quantum Electronics*, vol. 31, no. 10, pp. 1848–1855, Oct. 1995.
41. A. J. Lowery and D. Novak, Enhanced maximum intrinsic modulation bandwidth of complex-coupled DFB semiconductor lasers, *Electronics Letters*, vol. 29, no. 5, pp. 461–463, Mar. 1993.
42. L.-M. Zhang and J. E. Carroll, Enhanced AM and FM modulation response of complex coupled DFB lasers, *IEEE Photonics Technology Letters*, vol. 5, no. 5, pp. 506–508, Aug. 1993.
43. S. Hansmann, Transfer matrix analysis of the spectral properties of complex distributed feedback laser structures, *IEEE Journal of Quantum Electronics*, vol. 28, no. 11, pp. 2589–2595, Nov. 1992.
44. C. Youtsey, R. Grundbacher, R. Panepucci, I. Adesida, and C. Caneau, Characterization of chemically assisted ion beam etching of InP, *Journal of Vacuum Science & Technology B*, vol. 12, no. 6, pp. 3317–3321, Dec. 1994.
45. H. Blauvelt, N. Bar Chaim, D. Fekete, S. Margalit, and A. Yariv, AlGaAs lasers with micro-cleaved mirrors suitable for monolithic integration, *Applied Physics Letters*, vol. 40, no. 4, pp. 289–290, Feb. 1982.
46. W. T. Tsang, N. A. Olsson, and R. A. Logan, High-speed direct single-frequency modulation with large tuning rate and frequency excursion in cleaved-coupled-cavity semiconductor lasers, *Applied Physics Letters*, vol. 42, no. 8, pp. 650–652, Feb. 1983.
47. Q.-Y. Lu, W.-H. Guo, D. Byrne, and J. F. Donegan, Design of slotted single-mode lasers suitable for photonic integration, *IEEE Photonics Technology Letters*, vol. 22, no. 11, pp. 787–789, Mar. 2010.
48. Y. Wang, Y.-G. Yang, S. Zhang, L. Wang, and J.-J. He, Narrow linewidth single-mode slotted Fabry–Pérot laser using deep etched trenches, *IEEE Photonics Technology Letters*, vol. 24, no. 14, pp. 1233–1235, May 2012.
49. T.-T. Yu, L. Zou, L. Wang, and J.-J. He, Single-mode and wavelength tunable lasers based on deep-submicron slots fabricated by standard UV-lithography, *Optics Express*, pp. 16291–16299, July 2012.
50. J.-L. Zhao, K. Shi, Y.-L. Yu, and L. P. Barry, Theoretical analysis of tunable three-section slotted Fabry–Perot lasers based on time-domain traveling-wave model, *IEEE Journal of Selected Topics in Quantum Electronics*, vol. 19, no. 5, pp. 1–8, Mar. 2013.
51. S. O'Brien and E. P. O'Reilly, Theory of improved spectral purity in index patterned Fabry–Pérot lasers, *Applied Physics Letters*, vol. 86, no. 20, pp. 201101, May 2005.
52. S. O'Brien, A. Amann, R. Fehse, S. Osborne, E. P. O'Reilly, and J. M. Rondinelli, Spectral manipulation in Fabry–Perot lasers: Perturbative inverse scattering approach, *Journal of the Optical Society of America B*, vol. 23, no. 6, pp. 1046–1056, Jan. 2006.
53. X. Li and J. Park, Time-domain simulation of channel crosstalk and inter-modulation distortion in gain-clamped semiconductor optical amplifiers, *Optics Communications*, vol. 263, no. 2, pp. 219–228, Feb. 2006.

54. W. Li, W.-P. Huang, and X. Li, Digital filter approach for simulation of a complex integrated laser diode based on the traveling-wave model, *IEEE Journal of Quantum Electronics*, vol. 40, no. 5, pp. 473–480, May 2004.
55. Y.-P. Xi, W.-P. Huang, and X. Li, High-order split-step schemes for time-dependent coupled-wave equations, *IEEE Journal of Quantum Electronics*, vol. 43, no. 5, pp. 419–425, May 2007.
56. L. Dong, R.-K. Zhang, D.-L. Wang, S.-Z. Zhao, S. Jiang, Y.-L. Yu, and S.-H. Liu, Modeling widely tunable sampled-grating DBR lasers using traveling-wave model with digital filter approach, *Journal of Lightwave Technology*, vol. 27, no. 15, pp. 3181–3188, Apr. 2009.
57. L. Zhu, A. Scherer, and A. Yariv, Modal gain analysis of transverse Bragg resonance waveguide lasers with and without transverse defects, *IEEE Journal of Quantum Electronics*, vol. 43, no. 10, pp. 934–940, Oct. 2007.
58. Y. Li, Y.-P. Xi, X. Li, and W.-P. Huang, A single-mode laser based on asymmetric Bragg reflection waveguides, *Optics Express*, vol. 17, no. 13, pp. 11179–11186, June 2009.
59. X. Li, Z.-S. Zhu, Y.-P. Xi, L. Han, C. Ke, Y. Pan, and W.-P. Huang, Single-mode Fabry–Perot laser with deeply etched slanted double trenches, *Applied Physics Letters*, vol. 107, no. 9, pp. 091108, Sept. 2015.
60. C. Ke, X. Li, and Y.-P. Xi, A horn ridge waveguide DFB laser for high single longitudinal mode yield, *Journal of Lightwave Technology*, vol. 33, no. 24, pp. 5032–5037, Nov. 2015.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

31

Traveling Wave Modeling of Nonlinear Dynamics in Multisection Laser Diodes

31.1 Introduction 153
31.2 Basic TW Model in the Solitary Laser 155
31.3 Model of Material Gain Dispersion 158
31.4 Thermal Detuning 161
31.5 Spatially Inhomogeneous Carrier Density 163
31.6 Nonlinear Gain Saturation 165
31.7 Further Modifications of the TW Model 167
31.8 Multisection Lasers and Coupled Laser Systems 170
31.9 Simulations of Nontrivial MSL Device 172
31.10 Beyond Numerical Simulations of the TW Model 174
31.11 Conclusions 179

Mindaugas
Radziunas

31.1 Introduction

Semiconductor lasers, and a narrow waveguide edge-emitting semiconductor lasers, in particular, are attractive devices for different applications. Among others, these are high-speed all-optical signal processing, optical data storage, thermal and xerographic printing, scanning, directional lighting, secure communications, random number generation, frequency conversion, or various interferometric, spectroscopic, instrumentation, and other quantum-optical experiments.

A typical solitary narrow waveguide (single transversal mode) semiconductor laser exhibits a single-wavelength emission required in different applications. In many cases, however, small fluctuations of the operation conditions impose a significant phase noise which, in turn, causes an unwanted broadening of the emission linewidth. Moreover, the stable performance of the laser can be easily violated by optically reinjected light, and there is a huge number of studies devoted to the analysis of the nonlinear dynamics in lasers with a delayed optical feedback.

A properly designed optical feedback, however, can also play a constructive role when seeking to improve an operation of the solitary laser, or create a new dynamical regime. For example, an external cavity with a diffractive grating can be used for emission linewidth reduction or tuning of the lasing wavelength

[1]. Or, on the contrary, specially designed external cavities allow realizing a chaotic emission usable for cryptography [2] or random number generation [3].

Multisection semiconductor lasers (MSLs) in linear or ring configurations and coupled laser devices provide even more possibilities to tailor laser dynamics for certain applications. For example, a variety of important functionalities of the optical data communications [4], such as pulse generation, clock recovery, and fast switching can be realized by specially designed and differently interconnected MSLs. Several examples of theoretically investigated and experimentally verified dynamic performance of MSLs considered in our previous works are excitability [5], high-frequency mode-beating pulsations [6], and modulation band enhancement [7] in distributed feedback (DFB) lasers with an integrated passive phase tuning section (passive feedback lasers); passive [8] or hybrid [9] harmonic and sub-harmonic mode-locking in lasers with saturable absorber, and pulse broadening in quantum dot (QD) mode-locked lasers [10]; tunable high-frequency pulsations in the detuned grating DFB lasers with an integrated phase tuning section (phase controlled mode-beating lasers) [11,12]; stationary, pulsating, and irregular regimes and their bifurcations in DFB lasers with integrated phase tuning and amplifying sections (active feedback laser) [6,13]; Joule heating–induced transitions between steady states in distributed Bragg reflector (DBR) lasers [14] or external cavity diode lasers (ECDLs) [15].

All these examples confirm the practical importance of modeling, simulations, and analysis of MSLs for designing new devices with a particular dynamical behavior. The most precise models usually are given by 2+1 or 3+1 dimensional systems of partial differential equations (PDEs) [16,17]. The numerical simulations in this case, however, are time consuming, whereas application of analytic methods for the analysis of the nonlinear dynamics is very limited. Unfortunately, numerical simulations of such models are time consuming, whereas an application of analytic methods for the model analysis is very limited. For this reason, we prefer to use simpler approaches which, may be, fail to reproduce a quantitative-, but still allow to get a qualitative agreement between theory and experiments.

For some MSLs, already simple ordinary differential equation (ODE) or delay differential equation (DDE) systems (rate equations) admit a reasonable description of the laser dynamics. An advantage of these models is their simplicity allowing fast numerical simulations and application of advanced analytic methods, such as asymptotic analysis, stability analysis, or numerical continuation and bifurcation analysis. These models, however, usually are based on mean-field approximations, i.e., neglect inhomogeneity of laser parameters and dynamical variables along the laser cavity, take into account only a few fundamental characteristics of the considered lasers, or are suited to describe particular MSL configurations [18–20].

The 1+1-dimensional traveling wave (TW) model considered in this chapter is a compromise between simplicity and precision. It is a first-order PDE system having a single spatial dimension corresponding to the longitudinal (z-) direction along the laser cavity and describing dynamics of the slowly varying optical field amplitudes, polarization functions, and carrier density [21–23]. Comparing to ODE and DDE models mentioned earlier, the TW model is computationally more demanding but still enables an advanced analysis, which is hardly possible in the case of the multidimensional PDE models.

By taking into account or neglecting different physical effects, one can derive a whole hierarchy of TW models of different complexities. The standard part of all such models is a pair of TW equations governing the evolution of the complex forward- and back- propagating field amplitudes, $E^{+}(z,t)$ and $E^{-}(z,t)$. These equations originate from the decomposition of the dominant fundamental transverse electric (TE) component of the electromagnetic wave,

$$E(\mathbf{r},t) = \Phi(x,y)\left[E^{+}(z,t)e^{-ik_0 z} + E^{-}(z,t)e^{ik_0 z}\right]e^{i\omega_0 t}.$$

Here, ω_0 is the central reference frequency and k_0 the corresponding wave vector. Whereas the transversal waveguide mode profile, $\Phi(x,y)$, is an eigensolution of the waveguide equation, the related complex eigenvalue of the same problem defines the propagation factor β [24], which determines the evolution of the field amplitudes $E^{\pm}$. In general, the propagation factor depends on the complex interaction of carriers and photons. In our modeling approach, we apply a phenomenological dependence of this factor on the real carrier

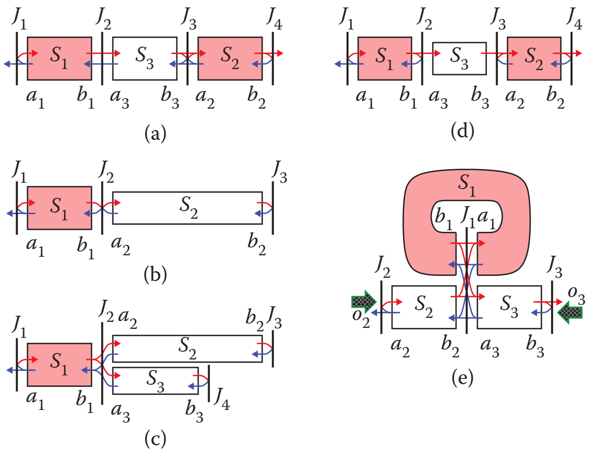


FIGURE 31.1 Schematic representation of five semiconductor laser devices, which can be considered by our modeling approach: (a) three-section laser, (b) laser with a trivial external cavity, (c) laser with a dual external cavity, (d) master–slave laser system, (e) optically injected ring laser with an outcoupling waveguide. Shaded and white frames represent active and passive *sections* (S_*) of the MSL. Thick black segments and thick hatched arrows indicate *junctions* (interfaces of these sections, J_*) and *optical injections* (o_*), respectively. Thin arrows show optical field transmissions and reflections at the interfaces of the laser sections.

density function N, which can represent dynamics of the spatially distributed carrier density, $N(z, t)$, or the section-wise averaged density, $N(t)$. The evolution of N itself is governed by a single or several rate equations.

One can use the TW modeling approach for consideration of various differently interconnected linear and curved, active and passive semiconductor waveguiding parts, taking into account optical injections, field reflections, and transmissions at the interfaces of different laser parts, as well as delayed feedback of the optical fields from the external cavities. For simulation and analysis of the MSLs, we apply our software LDSL-tool [25], which is suited to investigate the longitudinal dynamics of multisection semiconductor lasers. This software allows considering a large variety of MSL devices or coupled laser systems which can be represented by a set of mutually interconnected *sections* and *junctions*, see the schematic representations of several laser devices in Figure 31.1. Besides of numerical integration, LDSL-tool can find longitudinal optical modes and analyze their dynamics [23,26]. In some cases, it locates stable and unstable stationary states of the system [15,27], constructs the reduced ODE models based on a finite number of the optical modes [28], and, together with the software package AUTO [29], performs numerical continuation and bifurcation analysis of these reduced models [7].

In following, we shall introduce a basic TW model for the solitary laser, and present several model extensions allowing to take into account initially neglected physical effects. Next, we shall discuss a possibility to join several laser sections into a single multisection laser or a coupled laser device. For an illustration of the available device complexity, we shall present simulations of a ring laser with four branches of filtered feedback. At the last part of this chapter, we shall briefly introduce the concept of the instantaneous optical modes, discuss the mode analysis, the location and semi analytic continuation of the stationary states, the model reduction, and the numerical bifurcation analysis.

31.2 Basic TW Model in the Solitary Laser

In this section, we formulate the simplest TW model suitable for simulations of a solitary semiconductor laser. Let us consider an edge-emitting narrow-waveguide semiconductor laser (see Figure 31.2a). According to our notations, the "interior" part of this laser is referred as *section* S_1. The longitudinal coordinates

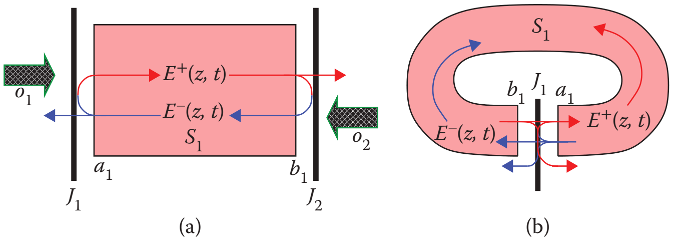


FIGURE 31.2 Schematic representation of the single section lasers in (a) linear and (b) ring configurations. Thin arrows indicate directions of the counter-propagating fields and their reflections/transmissions at the laser facets [J_1 and J_2, panel (a)] or point outcoupling interface [J_1, panel (b)]. Thick hatched arrows represent optically injected fields.

of the section edges and the length of this section are $z = a_1$, $z = b_1$, and $|S_1| = b_1 - a_1$, respectively. The front and the rear laser facets (*junctions* J_1 and J_2), in this case, correspond to the left and the right edges a_1 and b_1 of S_1.

The backbone of the TW model of this laser is the linear system of partial differential (TW) equations describing an evolution of the slowly varying complex amplitudes $E^+(z, t)$ and $E^-(z, t)$ of the counter-propagating optical fields:

$$\begin{cases} \frac{n_g}{c_0}\partial_t E^+ + \partial_z E^+ = -i\beta E^+ - i\kappa E^- + F_{sp}^+ \\ \frac{n_g}{c_0}\partial_t E^- - \partial_z E^- = -i\beta E^- - i\kappa E^+ + F_{sp}^- \end{cases}, \quad z \in S_1. \tag{31.1}$$

Here, c_0 is the speed of light in vacuum, $F_{sp}^{\pm}$ are the Langevin noise source contributions to the optical fields, and n_g is the group velocity index. The real and the imaginary parts of the complex coefficient κ represent the distributed index and gain/loss coupling of the counter-propagating fields, respectively. κ is nonvanishing in the laser sections containing Bragg grating and is set to zero in the straight sections without the grating. Without an additional scaling of the field functions $E^{\pm}$, $|E(z, t)|^2 = (E, E) = |E^+|^2 + |E^-|^2$ represents the photon density and is proportional to the local field power,

$$\mathcal{P}(z, t) = \frac{\sigma c_0}{n_g} \frac{hc_0}{\lambda_0} |E(z, t)|^2.$$

Here, σ is the cross-section area of the active zone, λ_0 the central wavelength, and h the Planck constant.

Active Sections:

The propagation factor β in the TW equations above can be defined as

$$\beta = \delta_0 + \tilde{n}(N) + \frac{i(g(N) - \alpha)}{2}, \tag{31.2}$$

where the peak gain and refractive index change functions $g(N)$ and $\tilde{n}(N)$ are given by the simple linear relations

$$g(N) = \Gamma g' \left(N - N_{tr}\right), \qquad \tilde{n}(N) = \frac{\alpha_H g(N)}{2}. \tag{31.3}$$

Here, N is the carrier density. The parameters δ_0, α, Γ, g', and α_H are the internal field loss, the initial fixed detuning from the central frequency, the confinement factor, the differential gain, and the linewidth enhancement (Henry) factor evaluated at the transparency carrier density N_{tr}, respectively.

To define the evolution of the spatially averaged carrier density $N(t)$, we use a single rate equation

$$\frac{\mathrm{d}}{\mathrm{dt}}N = \frac{I}{q\sigma|S_1|} - \mathcal{R}(N) - \mathcal{S}(N, E^{\pm}). \tag{31.4}$$

Here, q is the electron charge, I is the injected current into the active zone of the section, whereas $\mathcal{R}$ and $\mathcal{S}$ are spontaneous and stimulated recombination functions, respectively. We use a cubic spontaneous recombination function,

$$\mathcal{R}(N) = AN + BN^2 + CN^3, \tag{31.5}$$

which can be simplified by assuming vanishing recombination parameters B and C and defining $A = \tau_N^{-1}$, where τ_N denotes the carrier lifetime. Function $\mathcal{S}$ in the carrier rate equation 31.4 represents the spatially averaged stimulated recombination:

$$\mathcal{S}(N, E^{\pm}) = \frac{c_0}{n_\mathrm{g}} g(N) \|E\|_1^2. \tag{31.6}$$

Here, $\|E\|_1^2$ is the spatial average of the local photon density along the section S_1,

$$\|E\|_1^2 = \langle (E, E) \rangle_1, \qquad \langle \eta \rangle_1 = \frac{1}{|S_1|} \int_{S_1} \eta(z) \mathrm{d}z, \qquad (\zeta, \xi) = \zeta^{+*}\xi^{+} + \zeta^{-*}\xi^{-},$$

and * denotes the complex conjugate.

To complete the system, we still need to define the incident forward- and backward-propagating fields at the section edges $z = a_1$ and $z = b_1$, respectively. For the solitary laser, these incident fields can be defined by the following reflection/transmission conditions:

$$E^{+}(a_1, t) = -r_1^* E^{-}(a_1, t) + o_1(t), \;\; E^{-}(b_1, t) = r_2 E^{+}(b_1, t) + o_2(t). \tag{31.7}$$

Here, r_1 and r_2 are the complex field amplitude reflectivity coefficients at the laser facets (junctions J_1 and J_2), whereas complex functions $o_{1,2}(t)$ represent optical injections at these junctions.

One can also use Equations 31.1 through 31.6 for simulations of narrow-waveguide semiconductor ring lasers with the field in- and out-coupling concentrated in the single point of this laser. Figure 31.2b shows a schematic representation of such single-section ring laser device. According to this scheme, we assume that both, "left" and "right" edges a_1 and b_1 of the section S_1, are connected at the single junction J_1. The boundary conditions (Equation 31.7), in this case, should be replaced by the following field transmission-reflection conditions at J_1:

$$E^{+}(a_1, t) = t_1 E^{+}(b_1, t) - r_1^* E^{-}(a_1, t), \;\; E^{-}(b_1, t) = t_1 E^{-}(a_1, t) + r_1 E^{+}(b_1, t). \tag{31.8}$$

Here, t_1 is the real field amplitude transmission factor back into the ring section S_1 at the outcoupling point J_1, whereas the complex factor r_1 represents the localized field backscattering at J_1.

To perform simulations of the basic TW model determined by Equations 31.1 through 31.6 and 31.7 or 31.8, one still needs to choose some initial conditions $E^{\pm}(z, 0)$ and $N(0)$. For the first run of simulations, one can use any small distribution of the optical fields $E^{\pm}(z, 0)$ and a small positive value of $N(0)$. After some transient, the computed trajectory will be attracted by one of the few regular or irregular attractors of the considered dissipative system. To keep tracing the same attractor during the following parameter continuation calculations, one should better use previously obtained carrier density and field distributions.

Passive Sections:

It is noteworthy that one can also use the TW equations 31.1 for a description of the field propagation in the passive sections, such as gratings, free space between the laser and the external mirror, etc. Here, carriers are absent, do not couple to the emission wavelength (the material gain band of these sections does not support the lasing frequencies), or are just kept at transparency level by an appropriately adjusted bias current. In all such cases, $g(N) = \tilde{n}(N) = 0$, the carrier rate equations 31.4 are decoupled from the field equations 31.1 and, therefore, are irrelevant.

In the case of the passive section S_k containing no grating ($\kappa = 0$), simple analytic relations of the field function values on the both sides of S_k,

$$\begin{gathered} E^+(b_k, t) = \eta e^{i\varphi/2} E^+(a_k, t - \tau_k), \quad E^-(a_k, t) = \eta e^{i\varphi/2} E^-(b_k, t - \tau_k), \quad \text{where} \\ \tau_k = \frac{|S_k| n_g}{c_0}, \quad \eta = e^{-\alpha |S_k|/2}, \quad \varphi' = -2\delta_0 |S_k|, \end{gathered} \tag{31.9}$$

can replace the field equations 31.1.

In the case of the passive grating ($\kappa \neq 0$), the analytic solution of the field equations 31.1 in the frequency domain is given by the 2×2-dimensional transfer matrix M [23,26,28],

$$\hat{E}(z, \omega) = M(\beta, \kappa, \omega; z, a_k) \hat{E}(a_k, \omega). \tag{31.10}$$

Here, $\beta = \delta_0 - i\alpha/2$, ω is the relative frequency, $\hat{E}(z, \omega) = \left(\hat{E}^+, \hat{E}^-\right)^T$, T denotes the transpose vector, whereas $\hat{E}^\pm$ are the frequency domain representations of the fields $E^\pm(z, t)$. Within any interval $[z', z]$ where parameters β and κ are constant, the matrix M is defined by

$$\begin{gathered} M(\beta, \kappa, \omega; z, z') = \begin{pmatrix} \cos \eta(z - z') - \frac{i\mathcal{B}}{\eta} \sin \eta(z - z') & -\frac{i\kappa}{\eta} \sin \eta(z - z') \\ \frac{i\kappa}{\eta} \sin \eta(z - z') & \cos \eta(z - z') + \frac{i\mathcal{B}}{\eta} \sin \eta(z - z') \end{pmatrix}, \\ \mathcal{B}(\omega) = \beta + \frac{\omega n_g}{c_0}, \qquad \eta = \sqrt{\mathcal{B}^2 - \kappa^2}. \end{gathered} \tag{31.11}$$

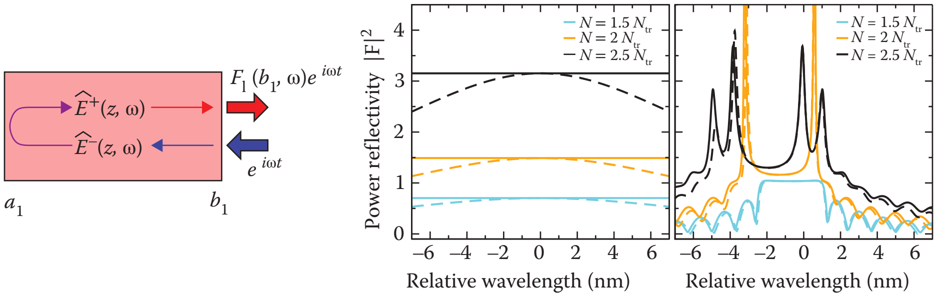
Once the parameters β or κ are peacewise constants, i.e., constant within each small subinterval $[z_s, z_{s-1}]$, $z' = z_0 < z_1 < \cdots < z_n = z$, the transfer matrix M is the superposition of the corresponding transfer matrices over these small subintervals:

$$M(\beta, \kappa, \omega; z, z') = M_n \times \cdots \times M_1, \quad M_s = M(\beta(z_{s-1/2}), \kappa(z_{s-1/2}), \omega; z_s, z_{s-1}). \tag{31.12}$$

31.3 Model of Material Gain Dispersion

The relations (Equation 31.3) introduced in Section 31.2 are simple linear approximations of the gain and refractive index functions G and $\tilde{N}$. In general, these functions depend not only on the carrier density N, but also on the optical frequency ω, field intensities $|E^+|^2$ and $|E^-|^2$, and some other physical effects, such as temperature, not considered in our modeling approach. In this section, we introduce the model of the gain dispersion of the semiconductor material, which restricts the gain band in the frequency domain and is the primary optical frequency selection mechanism in Fabry–Pérot (FP) lasers.

Before switching to the modeling of the gain dispersion, let us find out the expression of the laser response $F_l(b_1, \omega)$ to the incident plane wave $e^{i\omega t}$ applied to the right edge of the device (see Figure 31.3a). For this reason, we freeze the propagation factor β and substitute the ansatz $E^\pm(z, t) = \hat{E}^\pm(z, \omega) e^{i\omega t}$ into



(a) (b) (c)

FIGURE 31.3 Laser response to the incident plane wave $e^{i\omega t}$. (a) Schematic representation: thin and thick arrows show field propagation directions and incident/emitted optical fields. (b) Response intensity of the solitary FP laser: parameters $\lambda_0 = 1.57\,\mu\text{m}$, $|S_1| = 0.25\,\text{mm}$, $n_g = 3.6$, $\delta_0 = 0$, $\alpha = 20\,\text{cm}^{-1}$, $\alpha_H = -4$, $\Gamma = 0.15$, $g' = 4 \cdot 10^{-16}\,\text{cm}^2$, $N_{tr} = 10^{18}\,\text{cm}^{-3}$, $r_1 = r_2 = \sqrt{0.3}$, $\kappa = 0$. (c) Response intensity of the solitary DFB laser: parameters are the same as in (b), only $\kappa = 130\,\text{cm}^{-1}$ and $r_1 = r_2 = 0$. Solid and dashed curves in panels (b) and (c) represent models with ($\bar{g} = 100\,\text{cm}^{-1}$, $\bar{\lambda} = 0$, $\bar{\gamma}_\lambda = 40\,\text{nm}$) and without ($\bar{g} = 0$) gain dispersion.

the field equations 31.1. The solution of the resulting system of ODEs within S_1 can be represented by Equation 31.10, where the transfer matrix M is defined in Equation 31.11. The ratio of the outgoing and incident waves at $z = b_1$ together with the (noninjective) boundary condition (Equation 31.7) at $z = a_1$ define the function $F_l(b_1, \omega)$, which shows the laser response dependence on the optical frequency of the injected field†. In two simple cases of FP and DFB lasers with vanishing facet reflectivity, the response function is given by

$$F_l(b_1, \omega) = \frac{\widehat{E}^+(b_1,\omega)}{\widehat{E}^-(b_1,\omega)} = \begin{cases} -r_1^* e^{-2i\beta|S_1|} e^{-i2\omega n_g |S_1|/c_0}, & \kappa = 0,\; r_1 \neq 0 \;\; \text{(FP laser)} \\ \dfrac{\kappa}{i\eta(\omega)\cot[|S_1|\eta(\omega)] - \mathcal{B}(\omega)}, & \kappa \neq 0,\; r_1 = 0 \;\; \text{(DFB laser)} \end{cases}. \tag{31.13}$$

Figures 31.3b and c show the intensities of these response functions in FP and DFB lasers calculated for different values of carrier densities N. Note also, that the abscissa axis in these figures represents relative wavelengths λ related to the relative frequencies ω by formula $\lambda \approx -\frac{\lambda_0^2}{2\pi c_0}\omega$.

The flat laser response curves in Figure 31.3b indicate an absence of frequency selection mechanisms for FP lasers in our fundamental TW model. Thus, this simple model is not suitable for simulations of FP lasers. In contrast, the wavelength selection in DFB lasers is mainly determined by the Bragg grating, and numerical integration of the TW model can provide reliable information. One should note, however, that the index-coupled DFB laser (characterized by a real coupling factor κ) can emit at one of two resonance wavelengths located at both sides of the stopband, see a solid dark gray curve in Figure 31.3c, and the parameter tuning implied jumping between these two resonances can be expected in simulations. The gain dispersion, in this case, can be exploited for the suppression of one of the resonances [12,21].

Lorentzian Approximation of the Material Gain Function:

There are several methods for introduction of the frequency-selective gain dispersion into the time-domain TW model. Many of these approaches use an additional digital filtering of the numerically calculated optical field time series [30–34]. In some cases, these digital filters are equivalent to the numerical schemes obtained by discretization of some additional integrodifferential operators or differential equations. For the purpose of the analysis of the model equations, it is preferable to introduce the frequency band limiting elements directly into the model equations. For example, the TW model extensions admitting Lorentzian approximation of the material gain dispersion curves can be given by convolution integrals [34] or by an equivalent set of the linear first-order ODEs [21,35]. Another approach to model more sophisticated gain function profiles within the TW modeling frame by including nonlinear polarization equations was used, e.g., in References [22,36,37].

---

† In the same way, one can also define the response function $F_r(a_1, \omega)$ at the left side of the laser.

In this chapter, we follow the strategy proposed in References [12,21]. For this reason, we approximate the gain profile in the frequency domain by a Lorentzian with the amplitude $\bar{g}$, the full width at the half maximum $\bar{\gamma} = \frac{2\pi c_0}{\lambda_0^2}\bar{\gamma}_\lambda$, and the detuning of the peak frequency $\bar{\omega} = -\frac{2\pi c_0}{\lambda_0^2}\bar{\lambda}$. Here, $\bar{\gamma}_\lambda$ and $\bar{\lambda}$ are the wavelength representations of the Lorentzian width and its peak position. In the time domain, this approximation is represented by the additional linear operator $\mathcal{D}$ in the TW field equations, and a pair of linear differential equations for polarization functions $P^{\pm}(z,t)$:

$$\begin{aligned} \frac{n_g}{c_0}\partial_t E^{\pm} &= \mp\partial_z E^{\pm} - i\left(\beta - i\mathcal{D}\right)E^{\pm} - i\kappa E^{\mp} + F_{sp}^{\pm}, \quad z \in S_1, \\ \mathcal{D}E^{\pm} &= \frac{\bar{g}}{2}\left(E^{\pm} - P^{\pm}\right), \quad \partial_t P^{\pm} = \frac{\bar{\gamma}}{2}\left(E^{\pm} - P^{\pm}\right) + i\bar{\omega}P^{\pm}, \quad z \in S_1. \end{aligned} \tag{31.14}$$

The introduction of operator $\mathcal{D}$ also implies the following modification of the stimulated recombination function $\mathcal{S}$ entering the carrier rate equation 31.4:

$$\mathcal{S}(N,E^{\pm}) = \frac{c_0}{n_g}\mathfrak{R}\langle\left(E,[g(N) - 2\mathcal{D}]E\right)\rangle_1. \tag{31.15}$$

To understand the impact of the operator $\mathcal{D}$, we consider the laser response function $F_l(b_1,\omega)$ again according to the modified TW equations 31.14. When repeating the procedure described at the beginning of this section, the factor $\mathcal{B}$ entering Equation 31.11 takes the form

$$\mathcal{B}(\omega) = \beta + \frac{\omega n_g}{c_0} + \chi(\omega), \ \text{ where } \ \chi(\omega) = \frac{\bar{g}}{2}\frac{(\omega-\bar{\omega})}{\bar{\gamma}/2 + i(\omega-\bar{\omega})}, \quad -i\mathcal{D}\widehat{E}(z,\omega) = \chi(\omega)\widehat{E}(z,\omega). \tag{31.16}$$

Thus, an introduction of the linear dispersion operator $\mathcal{D}$ implies modifications of both, gain and refractive index change functions. The total gain (twice the imaginary part of $\beta-(\delta_0 - i\alpha/2)+\chi(\omega)$) and the refractive index change function (real part of the same factor), in this case, are given by the expressions

$$G(N,\omega) = g(N) - \frac{\bar{g}(\omega-\bar{\omega})^2}{(\bar{\gamma}/2)^2 + (\omega-\bar{\omega})^2}, \quad \tilde{N}(N,\omega) = \tilde{n}(N) + \frac{\bar{g}}{4}\frac{\bar{\gamma}(\omega-\bar{\omega})}{(\bar{\gamma}/2)^2 + (\omega-\bar{\omega})^2}. \tag{31.17}$$

The dashed curves in Figures 31.3b and c illustrate the impact of the introduced gain dispersion. Whereas these corrections in the case of DFB lasers (panel [c]) are small, for the FP lasers they provide an efficient wavelength selection mechanism.

It is noteworthy that vanishing factor κ in the field equations 31.2 and 31.14 implies the following simple expression of the monochromatic field *transmission* through the laser section:

$$\widehat{E}^{+}(b_1,\omega) = e^{-i\mathcal{B}(\omega)|S_1|}\widehat{E}^{+}(a_1,\omega), \quad \widehat{E}^{-}(a_1,\omega) = e^{-i\mathcal{B}(\omega)|S_1|}\widehat{E}^{-}(b_1,\omega),$$

where $\mathcal{B}(\omega)$ is defined in Equation 31.16. Thus, the TW equations 31.2 and 31.14 with vanishing functions $g(N)$ and $\tilde{n}(N)$, large Lorentzian amplitude $\bar{g}$, and small Lorentzian width $\bar{\gamma}$ can be effectively used for modeling of the optical filters, i.e., for extracting field frequency components located close to the relative frequency $\bar{\omega}$.

Further Modifications of the Model for Material Gain Dispersion:

The gain peak $g(N)$ and the simple Lorentzian dependence on the optical frequency determined by three fixed parameters $\bar{g}$, $\bar{\gamma}$, and $\bar{\omega}$ define the material gain profile $G(N, \omega)$ in Equation 31.17. To improve fitting of the gain profiles obtained by calculations of microscopic models for various values of N, one can replace these three factors by appropriately selected carrier-dependent functions $\bar{g}(N)$, $\bar{\gamma}(N)$, and $\bar{\omega}(N)$. In the cases, when the gain spectrum has two and more peaks or the asymmetry of the single peak is important, one can also introduce an additional set or several sets of polarization functions $P^{j\pm}(z, t)$. The gain dispersion operator $\mathcal{D}$ and corresponding total gain and refractive index functions, in this case, read as

$$\mathcal{D}^{(s)}E^{\pm} = \sum_{j=1}^{s} \frac{\bar{g}_j}{2}\left(E^{\pm} - P^{j\pm}\right), \qquad \partial_t P^{j\pm} = \frac{\bar{\gamma}_j}{2}\left(E^{\pm} - P^{j\pm}\right) + i\bar{\omega}_j P^{j\pm}, \quad z \in S_1,$$

$$G(N, \omega) = g(N) + 2\Im\chi^{(s)}(\omega), \;\; \tilde{N}(N, \omega) = \tilde{n}(N) + \Re\chi^{(s)}(\omega), \;\; \chi^{(s)}(\omega) = \sum_{j=1}^{s} \frac{\bar{g}_j}{2} \frac{(\omega - \bar{\omega}_j)}{\bar{\gamma}_j/2 + i(\omega - \bar{\omega}_j)},$$

where s is the number of polarization function sets. Since the maximal value of $2\Im\chi(\omega)$ is, in general, smaller than zero, one should also correct the function $g(N)$.

Concluding the discussion of this section, we note that a proper numerical resolution of the gain and refractive index functions (Equation 31.17) for the broad frequency band when simulating the time-domain TW mode requires a careful selection of the numerical algorithm and temporal discretization steps. The size of the frequency band that can be represented by calculated discrete time series is inversely proportional to the time step, whereas the precision of the numerical simulations when approaching borders of this band are rapidly degrading. Thus, a suitable time discretization step should ensure that all important frequency regions (Bragg resonances, surrounding of a gain peak frequency, a frequency of optically injected beams, if present) are within the central part of the allowed frequency band.

31.4 Thermal Detuning

Let us switch now to the consideration of the thermal effects. An increase of the bias current implies changes of the device temperature and, consequently, changes in the refractive index and the lasing wavelength. To model these thermal tuning effects in our device, we supplement the propagation factor β from Equation 31.2 with an additional thermal detuning term $\tilde{n}_T$ [14,38]:

$$\beta = \delta_0 + \tilde{n}(N) + \tilde{n}_T(I) + \frac{i(g(N) - \alpha)}{2}, \qquad \tilde{n}_T = \frac{2\pi n_g}{\lambda_0^2}\nu_1^1 I. \tag{31.18}$$

The linear thermal detuning function $\tilde{n}_T(I)$ determines the impact of the injection current I to the refractive index change. The factor ν_1^1 in solitary lasers determines an approximate red shift of the lasing wavelength due to increased bias current:

$$\nu_1^1 \approx \frac{\Delta_\lambda}{\Delta_I},$$

Here, Δ_I is the bias current tuning interval, whereas Δ_λ is the (continuous) lasing wavelength change during this current tuning.

Figures 31.4a and c show the simulated wavelength change with the increased bias current in the solitary FP and DFB lasers, respectively. Here, besides the dominant optical modes shown in white, one can see other slightly excited optical modes which are (almost) equidistant in the FP case (a) or indicate the DFB laser resonance located on the other side of the stopband. The estimated wavelength shift $\frac{\Delta_\lambda}{\Delta_I} \approx 3.14$ nm/A obtained for the FP laser and 3.09 nm/A for the DFB laser slightly differs from the factor $\nu_1^1 = 3.2$ nm/A

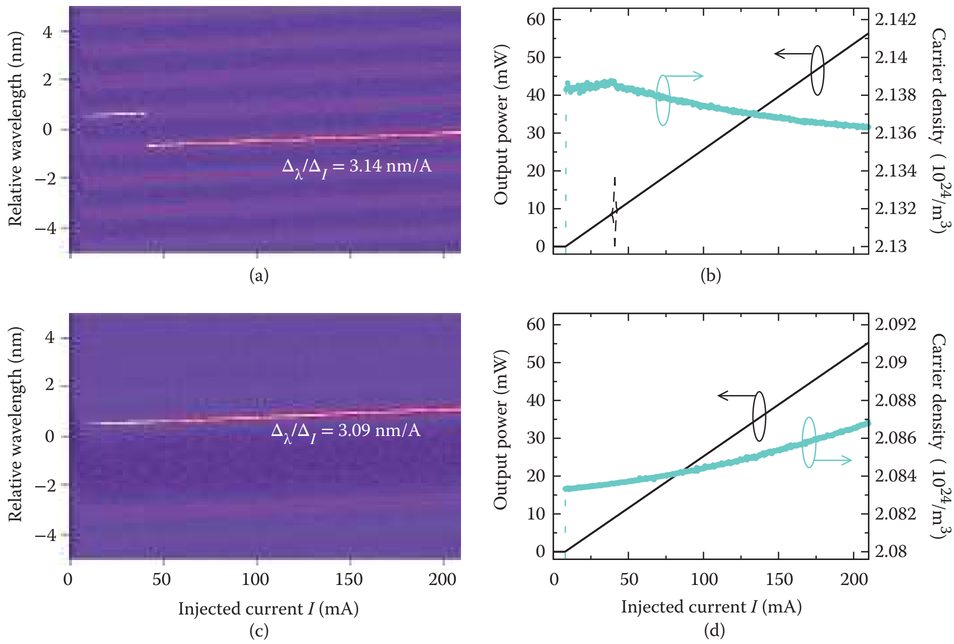


FIGURE 31.4 Mapping of the optical spectra (a), (c), and mean emitted power and carrier density (b), (d) as functions of the increased injected current I in solitary FP (a), (b), and DFB (c), (d) lasers. $\sigma = 2 \cdot 10^{-13}\ \mathrm{m}^2$, $A = 2 \cdot 10^8/\mathrm{s}$, $B = 1 \cdot 10^{-16}\ \mathrm{m}^3/\mathrm{s}$, $C = 1.3 \cdot 10^{-41}\ \mathrm{m}^6/\mathrm{s}$, $\nu_1^1 = 3.2 \cdot 10^{-9}\ \mathrm{m/A}$, whereas other parameters are the same as in Figures 31.3b and c in the case of nonvanishing gain dispersion. White lines and light shading in (a) and (c) represent main and side peaks of the calculated optical spectra. Solid black and gray curves in (b) and (d) show time-averaged emission intensity and carrier density, respectively. Dashed black lines indicate minima and maxima of the emission intensity.

used in our simulations. We attribute this slight discrepancy to the additional contribution of the full refractive index change function $\tilde{N}(N, \omega)$ defined in Equation 31.17. Namely, the dependence of the carrier density N and the relative lasing wavelength λ on the bias current I (see gray curves in Figures 31.4b and d and wavelength shifts in Figures 31.4a and c, respectively) implies nonvanishing changes of the function $\tilde{N}(N, \omega) = \tilde{N}(N, -\lambda c_0/\lambda_0^2)$ that counteracts the thermal detuning term $\tilde{n}_T(I)$ and slightly reduces the redshift of the lasing wavelength.

It is noteworthy that an introduction of the thermal detuning term $\tilde{n}_T$ in our still simple TW model of the solitary laser implies, in general, only a continuous tuning of the lasing frequency. Once achieving threshold, the carrier density changes only slightly (gray curves in Figures 31.4b and d), whereas the emitted field intensity increases linearly without a visible saturation (black curves in the same panels), which is still not taken into account in our model. This linear growth of the lasing wavelength can be correctly understood when analyzing TW field Equation 31.1 with the propagation factor β defined by Equation 31.18 and neglected gain dispersion. Due to the transfer matrix formalism (Equations 31.10 and 31.11) and the expression of $\mathcal{B}$ in Equation 31.11, the extension of β by the nonvanishing real term $\tilde{n}_T$ is equivalent to the change of the relative frequency ω by $-\tilde{n}_T c_0/n_g$, or, alternatively, the change of λ by $\frac{-\lambda_0^2}{2\pi c_0}\frac{-\tilde{n}_T c_0}{n_g} = \nu_1^1 I$.

The unique, more complicated feature in Figure 31.4 is the transition between two states in the FP laser at $I \approx 40\ \mathrm{mA}$, see panels (a) and (b). Figure 31.4a shows that in the vicinity of the transition, these two states are determined by two optical modes belonging to the opposite slopes of the wavelength-dependent gain profile with the peak wavelength at $\bar{\lambda} = 0$. Due to the redshift, all optical modes located on the falling (increasing) slope of this gain profile undergo an increase (decrease) of the detuning from the gain peak wavelength and, consequently, a slight rise (fall) in the mode threshold N; see the gray curves in panel (b) for $I < 40$ mA ($I > 40$ mA). A similar increase of the dominant falling-gain-slope mode threshold can also be seen in Figure 31.4d. At the position of the state transition, the wavelengths of two involved modes are symmetric with respect to the gain peak wavelength and, what is more important, their thresholds become equal. Due to a further tuning of the bias current, the previously suppressed mode at the increasing gain profile slope becomes the minimal threshold mode, is amplified and, finally, turns to be the dominant one. See References [14,15,26] for more details on similar and more complex mode transitions.

Modeling of Cross-Talk Heating Effects in Multisection Devices:

In MSLs devices, one can also use a more advanced model for thermal detuning function $\tilde{n}_T$ which takes into account local and nonlocal cross-talk heating effects [38]:

$$\tilde{n}_T|_{z\in S_k} = \tilde{n}_{T,k} = \tfrac{2\pi n_{g,k}}{\lambda_0^2} \textstyle\sum_{r=1}^{m} \nu_k^r I_r. \tag{31.19}$$

Here, m is a number of sections in the considered MSL. The coefficients ν_k^r of the linear thermal detuning function $\tilde{n}_T(I)$ determine the impact of the injection currents I_r attributed to the sections S_r on the refractive index change within each laser section S_k.

The effect of the thermal detuning in MSLs is much more complicated than that one of the solitary laser. Besides of the red shift of the lasing wavelength, the MSLs can also exhibit periodically or almost periodically reappearing transitions between different states. The change of mean carrier density in various sections during each such period between state changes can be significant and cannot be explained by simple gain saturation or detuning from the gain peak effects. In some cases, a measured variation of the lasing wavelength with an increase or decrease of the injection current in different laser sections, together with the analysis of the field equations provide good estimates of thermal detuning coefficients including cross-talk effects [14,15,38].

Another well-known effect occurring with the heating of the semiconductor laser is the red shift of the gain peak wavelength [39]. If required, these changes can be accounted by the relation [38]

$$\bar{\lambda}|_{z\in S_k} = \bar{\lambda}_k = \bar{\lambda}_k^0 + \textstyle\sum_{r=1}^{m} \bar{\nu}_k^r I_r, \tag{31.20}$$

which is quite similar to the thermal detuning relation (Equation 31.19). Here, $\bar{\lambda}_k^0$ denotes an injection-independent part of the gain peak wavelength in the section S_k, and $\bar{\nu}_k^r$ are linear thermal gain peak detuning coefficients. When applying these expressions, one should be aware that a proper numerical time-domain resolution of a significant (tens or even hundreds of nanometers) gain peak shift requires very small time and, consequently, space discretization steps.

31.5 Spatially Inhomogeneous Carrier Density

Another important extension of the basic TW model takes into account sectionally inhomogeneous distributions of carrier density N. Namely, in this case, the sectionally averaged carrier density function $N(t)$ is replaced by the spatially distributed function $N(z, t)$, $z \in S_1$. This model extension can be especially important in the situations admitting localization of the high-intensity fields within the laser cavity, which takes place, e.g., during propagation of ultrashort optical pulses in mode-locked lasers, or DBR lasers with a high coupling factor κ. Due to stimulated recombination, the high-intensity fields at these localized regions can significantly deplete the carrier distribution causing a spatial hole burning (SHB) of the carriers [12,40].

To achieve a quantitative description of SHB, we replace simple carrier rate Equation 31.4 by the following equation for spatially distributed carrier density:

$$\begin{aligned}\partial_t N(z,t) &= \mathcal{J}(I,N) - \mathcal{R}(N) - \mathcal{S}(N,E^{\pm}), \quad z \in S_1,\\ \mathcal{J}(I,N(z,t)) &= \tfrac{1}{q\sigma|S_1|}\left[I + \tfrac{U_F'}{r_S}\left(\langle N\rangle_1 - N(z,t)\right)\right].\end{aligned} \tag{31.21}$$

Here, $\mathcal{J}$ is the inhomogeneous injection current density [12,41,42], N and $\langle N\rangle_1$ are spatially distributed and sectionally averaged carrier densities, U_F' and r_s denote the derivative of the Fermi level separation with respect to N and the series resistivity, whereas $\mathcal{S}$ is the spatially distributed stimulated recombination function,

$$\mathcal{S}(N,E^{\pm}) = \tfrac{c_0}{n_g}\Re\left(E, [g(N) - 2\mathcal{D}]E\right). \tag{31.22}$$

In the case of the limit $r_s \to 0$, the spatially distributed carrier density, $N(z,t)$, at each position z converges to the sectional average, $N(t) = \langle N\rangle_1$. Since the sectional averaging of the relations 31.21 and 31.22 yields expressions 31.4 and 31.15, the TW models with and without spatial distribution of carriers in this limit case are equivalent.

Figure 31.5 shows some effects occurring due to the SHB of carriers in solitary DFB lasers. The impact of the SHB depends on the injection level. Just above the lasing threshold, the field intensity is small, and the carrier density remains nearly homogeneous, having only a small dip in the center of the laser. With raising injection, this dip increases, but, due to a simultaneous increase of the carrier density at the facets, the mean density remains nearly constant: see only slightly increasing $\langle N\rangle_1$ for $I \leq 90\,\text{mA}$ in panel (b) of the same figure. At these small-to-moderate bias current levels, the carrier density remains symmetric with respect to the laser centrum (dashed curve in panel [d]), and emission at the both facets is the same

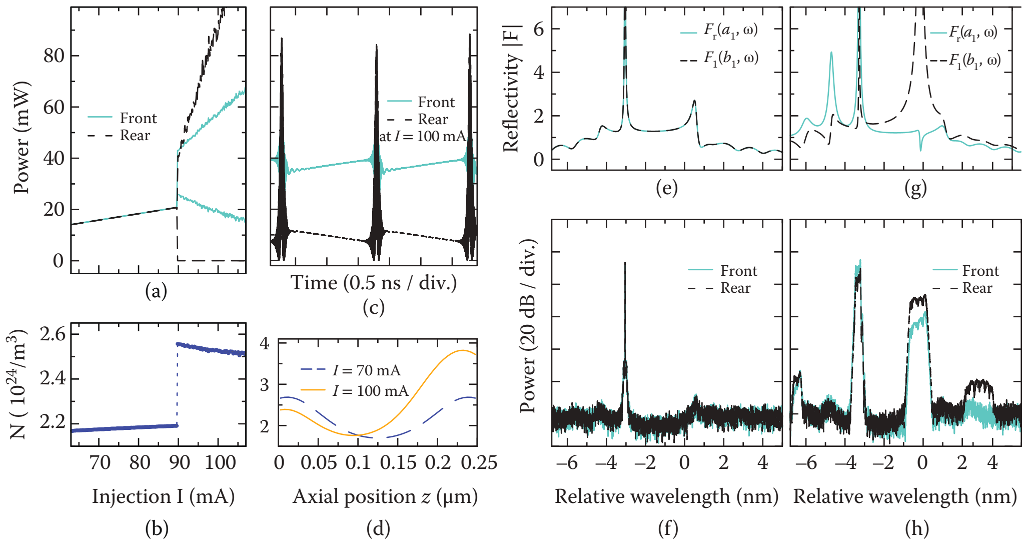


FIGURE 31.5 SHB in solitary DFB laser. (a) Minimal and maximal output power at both facets and (b) sectionally averaged carrier density $\langle N\rangle_1$ as functions of increased bias current. (c) Time trace of the emitted field intensity at both facets for fixed $I = 100\,\text{mA}$. (d) Time-averaged carrier densities for $I = 70\,\text{mA}$ and $I = 100\,\text{mA}$. DFB reflectivity spectra (e), (g), and calculated optical spectra (f), (h) for $I = 70\,\text{mA}$ (e), (f), and $I = 100\,\text{mA}$ (g), (h). Solid gray and black dashed curves in panels (a), (c), (e) through (h) represent optical fields at the front (J_1) and rear (J_2) facets of the laser. $U_F' = 10^{-25}\,\text{Vm}^3$, $r_S = 5\,\Omega$, whereas other parameters are the same as in Figures 31.4c and d except $\nu_1^1 = 0$.

(see coinciding solid gray and black dashed curves in panel (a) for $I \leq 90$ mA and panel (f) at $I = 70$ mA). The inhomogeneous carrier density (dashed curve in Figure 31.5d) causes a corresponding longitudinal variation of the index of refraction. The Bragg resonance of the grating thus has not the identical spectral position along the section but varies over a considerable portion of the stopband [12]. As a consequence, the symmetry of the stopband is lost, and the laser preferably operates on the short wavelength side of its stopband (see Figure 31.5f). The preference of the short-wavelength mode is also shown by the laser response functions $F_l(b_1, \omega)$ and $F_r(a_1, \omega)$ (see Figure 31.5e) calculated for spatially distributed carrier density profile (dashed curve in Figure 31.5d) at $I = 70$ mA according to the formulas 31.10 through 31.12 and the algorithm explained in Section 31.3.

At $I \approx 90$ mA, the symmetric solution loses its stability in symmetry breaking pitchfork bifurcation [43]. According to References [12,43], the supercritical pitchfork bifurcation of the stable symmetric steady state in DFB lasers generates a pair of new stable steady states with asymmetric density profiles (each one the mirror of the other). In our case, the pitchfork bifurcation seems to be of the subcritical type. Instead of finding two asymmetric stable steady states, we immediately jump to the pulsating state with different emission at both laser facets and larger mean carrier density (see panels [a] and (b) for $I > 90$ mA, respectively). The spatial distribution of the carrier density, in this case, is strongly asymmetric, see a solid curve in panel (d). This asymmetry together with noncommuting intermediate transfer matrices M_s from Equation 31.12 implies differences in DFB laser response functions estimated at the front (F_r) and the rear (F_l) sides of the laser, see Figure 31.5g. The optical spectra of the emission at the both sides of the laser (Figure 31.5h) also reveal these differences. Like the response functions of the panel (g), the left (thick gray curves) and the right (thin dashed curves) facet emissions have more pronounced contributions at the shorter and longer wavelength sides of the stopband, respectively.

31.6 Nonlinear Gain Saturation

Until now, our phenomenological models for peak gain and refractive index change (Equation 31.3) were taking into account their dependence on the sectionally averaged or local carrier density N. It is known, however, that the high-intensity optical fields saturate the gain function. To account for such saturation, one can introduce the following modifications of the gain and refractive index functions, which should be used for the definition of the propagation factor β in Equations 31.2 or 31.18 and stimulated recombination function $\mathcal{S}$ in Equations 31.6, 31.15, or 31.22:

$$g(N, E^{\pm}) = g(N)\rho_G(E^{\pm}), \qquad \tilde{n}(N, E^{\pm}) = \tilde{n}(N)\rho_I(E^{\pm}),$$

$$\text{where} \qquad \rho_j = \begin{cases} \left(1 + \varepsilon_j |E|^2\right)^{-1}, & \text{if } N = N(z,t) \\ \left(1 + \varepsilon_j \|E\|^2\right)^{-1}, & \text{if } N = N(t) \end{cases}, \quad j = G, I.$$

Two different parameters, ε_G and ε_I, separately define the nonlinear gain and refractive index dependence on the local or spatially averaged optical field intensity. A typical assumption $\varepsilon_G = \varepsilon_I$ relates the gain and refractive index functions by the linewidth enhancement factor α_H. Another reasonable assumption $\varepsilon_I = 0$, $\varepsilon_G > 0$ [44] used for modeling of high power amplifiers considers the nonlinear compression of the gain function alone.

The importance of the nonlinear gain compression is best visible in high-power lasers and optical amplifiers showing several Watt emission intensity [44]. Some impact of the gain compression in small-to-moderate ($\leq$100 mW) intensity regimes can also be observed when operating in the vicinity of various bifurcations, where a small change of parameters implies qualitative changes of the operating states. We should note, however, that the gain compression, in this case, implies only small shifts of the bifurcation positions, but has no significant impact on the qualitative description of laser dynamics in a large parameter domain. An analysis of simple TW model Equations 31.1 and 31.4 can explain the little influence of the gain compression in these regimes. A nonvanishing gain compression depletes the gain function $g(N)$ what implies a growth of the carrier density needed to reach threshold gain condition g_{th}. In solitary lasers, this growth is given by factor $g_{\text{th}}\varepsilon_G|E|^2/(g'\Gamma)$, which for typical gain compression coefficients and small-to-moderate field intensities is not exceeding a few percents of threshold carrier density. Consequently, a similar (up to a few percent) decay of the emission intensity can be observed.

A somehow different situation occurs in semiconductor ring lasers [19,22,23,45], where a proper introduction of nonlinear gain compression is crucial when deciding the type of operation states. In this case, one should distinguish the gain compression implied by co- and counter- propagating fields:

$$g^{\pm}(N,E^{\pm}) = g(N)\rho_G^{\pm}(E^{\pm}), \qquad \tilde{n}^{\pm}(N,E^{\pm}) = \tilde{n}(N)\rho_I^{\pm}(E^{\pm}),$$
$$\text{where} \quad \rho_j^{\pm} = \begin{cases} \left(1+\varepsilon_{js}|E^{\pm}|^2+\varepsilon_{jc}|E^{\mp}|^2\right)^{-1}, & \text{if } N=N(z,t) \\ \left(1+\varepsilon_{js}\langle|E^{\pm}|^2\rangle+\varepsilon_{jc}\langle|E^{\mp}|^2\rangle\right)^{-1}, & \text{if } N=N(t) \end{cases}, \quad j\in\{G,I\}, \tag{31.23}$$

whereas parameters ε_{js} and ε_{jc}, $j = G, I$, determine self- and cross-saturation of the gain and refractive index functions. In the ring lasers, usually is assumed that $\varepsilon_{Gc} > \varepsilon_{Gs}$, and $\varepsilon_{Ic} > \varepsilon_{Is}$. A detailed analysis based on the Maxwell–Bloch equations showed that the cross-saturation factor for two resonant modes in the ring cavity is twice larger than the self-saturation one [46].

The generalized functions $g^{\pm}$ and $\tilde{n}^{\pm}$ enter the definition of the propagation factor $\beta = \beta^{\pm}$ and the stimulated recombination function $\mathcal{S}$:

$$\beta^{\pm} = \delta_0 + \tilde{n}^{\pm}(N,E^{\pm}) + \tilde{n}_T(I) + \frac{i(g^{\pm}(N,E^{\pm})-\alpha)}{2},$$
$$\mathcal{S}(N,E^{\pm}) = \begin{cases} \frac{c_0}{n_g}\Re\sum_{\nu=\pm} E^{\nu *}[g^{\nu}(N,E^{\pm})-2\mathcal{D}]E^{\nu}, & \text{if } N=N(z,t) \\ \frac{c_0}{n_g}\Re\sum_{\nu=\pm}\langle E^{\nu *}[g^{\nu}(N,E^{\pm})-2\mathcal{D}]E^{\nu}\rangle_1, & \text{if } N=N(t) \end{cases}. \tag{31.24}$$

It is noteworthy that differences in parameters ε_{js} and ε_{jc}, $j = G, I$, imply differences in the propagation factors β^+ and β^- determining the evolution of the fields E^+ and E^-, respectively. These differences are crucial when determining type and stability of operating states in the ring laser, see References [19,23,45] for more details.

To illustrate how an asymmetry of the self- and cross-gain saturation implies different operation states in the ring laser (see Figure 31.2b), we have simulated the TW model equations 31.14, 31.21, 31.8, 31.5, 31.23, 31.24, 31.3 for vanishing ε_{Is} and ε_{Ic}, fixed nonvanishing sum $\varepsilon_{Gc} + \varepsilon_{Gs} = C > 0$ and tuned difference $\varepsilon_{Gc} - \varepsilon_{Gs}$. Figure 31.6 shows the results of these simulations. Solid gray and dashed black curves in all panels of this figure represent clockwise (CW) and counter-clockwise (CCW) propagating field functions $E^-(b_1, t)$ and $E^+(a_1, t)$ at the point scattering source J_1, respectively (see Figure 31.2b). Panel (a) of this figure gives an overview of all obtained states when tuning $\varepsilon_{Gc} - \varepsilon_{Gs}$ from $-C$ (full self-saturation with vanishing ε_{Gc}) up to $+C$ (full cross-saturation with vanishing ε_{Gs}). Panels (b)–(e) of the same figure represent four different observed dynamic regimes. The first three regimes occurring with a consequent increase of the cross-gain saturation are the bidirectional stable stationary state (b), the alternate oscillations (c), and the unidirectional bistable state (d). These three regimes can be observed experimentally and recovered theoretically using a simple two-mode ODE model [19]. An analysis of the TW model performed in Reference [23] has explained the relation between the asymmetry of the gain compression factors, ε_{Gc} and ε_{Gs}, and stability of the bidirectional steady state (regime b) or unidirectional bistable states (regime d). It was also shown, how the difference $\beta^+ - \beta^-$ and localized backscattering r_1 determine the frequency of alternating oscillations (regime c).

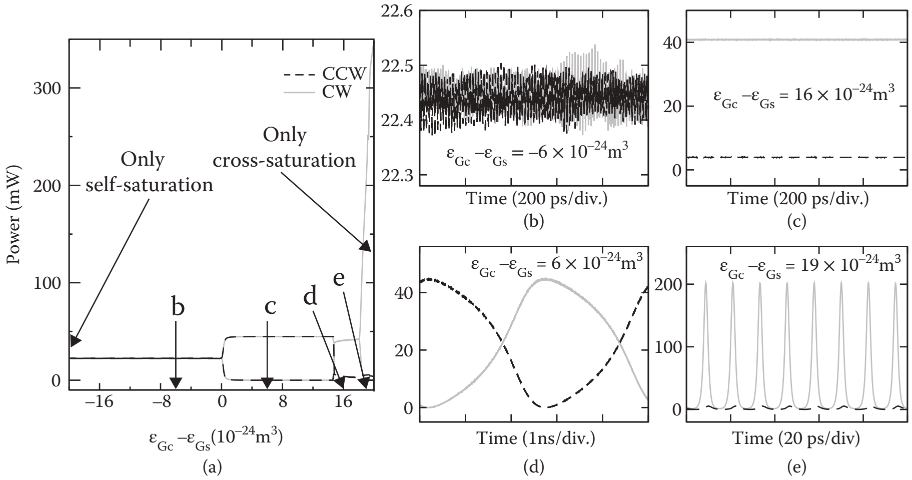


FIGURE 31.6 Dynamic regimes for different contributions of the cross- and self-gain saturations. (a) Maximal and minimal intensities of the optical fields at J_1 for changing values of $\varepsilon_{Gc} - \varepsilon_{Gs}$ but fixed $\varepsilon_{Gs} + \varepsilon_{Gc} = 20 \cdot 10^{-24}$ m^3. (b)–(e) Typical representatives of the observed regimes. Thick gray and dashed black curves indicate clockwise (CW) and counter-clockwise (CCW) propagating fields. Parameters are similar to those of Figures 31.4a and b, only $|S_1| = 1000\,\mu$m, $\alpha_H = -2$, $\alpha = 2\,\text{cm}^{-1}$, $\nu_1^1 = \varepsilon_{Is} = \varepsilon_{Ic} = 0$, $I = 100$ mA. U_F' and r_S are the same as in Figure 31.5, whereas the field transmission and localized backscattering parameters at J_1 are $t_1 = \sqrt{0.7}$ and $r_1 = 0.007$, respectively.

The last simulated regime (e) was observed for the dominant cross-gain saturation. Like usual mode-locking pulsations, this regime is characterized by large short pulses occurring with the round-trip period. However, in contrast to the mode-locking observed in multisection ring lasers [47], this state is unidirectional and does not require any fast saturable absorption. Similar mode-locked pulsations in a single-section ring laser were found and discussed theoretically in Reference [45].

31.7 Further Modifications of the TW Model

There exist a vast number of further possible modifications of the TW model for MSLs. Each of these modifications, however, requires a few new not very well-known parameters and, therefore, should be used with the great care. On the other hand, some of these modifications being crucial when analyzing a particular group of MSLs can be irrelevant for simulations and analysis of different type MSLs. Later we present several modifications of the TW model used for investigation of specific types of MSLs.

Multiple Carrier Rate Equations in QD Lasers:

When modeling QD lasers, one should take into account carrier exchange processes between a carrier reservoir (CR) and discrete levels in QDs.

One of the simplest ways to account for all these transitions within the TW modeling frame is provided by the rate equations for the normalized carrier density $N^{cr}(z, t)$ (scaled by the factor Θ_N) within the CR, and occupation probabilities $N^{gs}(z, t)$, $N^{es}(z, t)$ of the ground state (GS) and the first excited state (ES) of QDs, respectively [10,48,49].

To keep the structure of the TW field equations 31.14 unchanged, we neglect the inhomogeneous spectral broadening effect due to QD nonuniformity and consider a simple single-Lorentzian gain spectrum profile, which limits the material gain bandwidth. Besides, we assume that the laser operates at the GS transition only. In this case, the propagation factor β depends on the ground state occupation probability

$N^{gs}(z,t)$ only. The expression of $\beta(N^{gs})$ is equivalent to that one given by Equations 31.2 and 31.3 with spatially distributed occupation probability $N^{gs}(z) \in [0,1]$ and factor 1/2 instead of the carrier density $N(z,t)$ and transparency carrier density N_{tr}, respectively.

To describe carrier exchange processes between the CR, GS, and ES of the QDs in the active section (see Figure 31.7a), we use the following set of rate equations:

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}N^{gs}(z,t) &= -\frac{N^{gs}}{\tau_{gs}} + 2\left(\frac{N^{es}(1-N^{gs})}{\tau_{es\to gs}} - \frac{N^{gs}(1-N^{es})}{2\tau_{gs\to es}}\right) - \frac{1}{\theta_E}\mathcal{S}(N^{gs},E^{\pm}),\\
\frac{\mathrm{d}}{dt}N^{es}(z,t) &= -\frac{N^{es}}{\tau_{es}} - \left(\frac{N^{es}(1-N^{gs})}{\tau_{es\to gs}} - \frac{N^{gs}(1-N^{es})}{2\tau_{gs\to es}}\right) + \left(\frac{N^{cr}(1-N^{es})}{4\tau_{cr\to es}} - \frac{N^{es}}{\tau_{es\to cr}}\right),\\
\frac{\mathrm{d}}{\mathrm{d}t}N^{cr}(z,t) &= \frac{I}{q|S_1|\theta_I} - \frac{N^{cr}}{\tau_{cr}} - 4\left(\frac{N^{cr}(1-N^{es})}{4\tau_{cr\to es}} - \frac{N^{es}}{\tau_{es\to cr}}\right).
\end{aligned} \tag{31.25}$$

Here, $\mathcal{S}(N^{gs},E^{\pm})$ is defined by Equation 31.22, whereas τ_a^{-1} and $\tau_{a\to b}^{-1}$, $a,b \in \{gs,es,cr\}$, denote spontaneous relaxation and transition rates between GS, ES, and CR, respectively. Factors $(1-N^{gs})$ and $(1-N^{es})$ represent the Pauli blocking, factors 2 and 4 account for the spin degeneracy in the QD energy levels. Note that here we neglect direct transitions between CR and GS. θ_I and $\theta_E = \frac{2hc_0\theta_I}{\lambda_0}$ are scaling factors relating the injection current I, the field intensity $|E|^2$, the CR scaling factor θ_N, the differential gain g', and the QD density in the active zone.

In the saturable absorption sections (see Figure 31.7b), there is no pumping, $I = 0$, so that the transitions from CR to ES can be neglected, $\tau_{cr\to es} \to \infty$, and the last of the equations 31.25 can be ignored. The carrier transition from ES to CR can be added to similar spontaneous recombination term: $\bar{\tau}_{es}^{-1} = \tau_{es}^{-1} + \tau_{es\to cr}^{-1}$. Following Reference [48], one can model the carrier transitions in the negatively driven saturable absorber

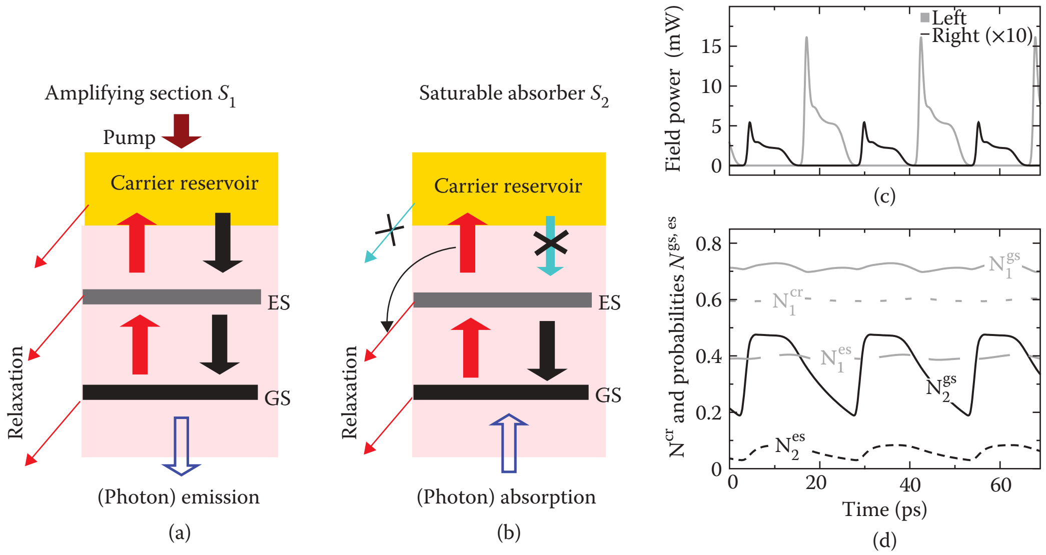

Amplifying section S_1
Pump
Carrier reservoir
ES
GS
Relaxation
(Photon) emission
(a)
Saturable absorber S_2
Carrier reservoir
ES
GS
Relaxation
(Photon) absorption
(b)
Field power (mW)
15
10
5
0
Left
Right (×10)
(c)
N^{cr} and probabilities $N^{gs,\,es}$
0.8
0.6
0.4
0.2
0
N_1^{gs}
N_1^{cr}
N_1^{es}
N_2^{gs}
N_2^{es}
0
20
40
60
Time (ps)
(d)

FIGURE 31.7 Schematic representation of carrier transitions in (a) the gain section and (b) the saturable absorption section of the QD MSL. Crossed arrows in panel (b) indicate the transitions which are neglected in the model equations. A sequence of (c) the emitted mode-locking pulses with an enhanced trailing edge plateaux at both sides of the laser and (d) the sectionally averaged carrier functions $N^{gs,es,cr}$ in both laser sections. All parameters and a detail description of the MSL device can be found in References [10,49].

by assuming an exponential decay of $\bar{\tau}_{es}$ with growing negative voltage U, whereas all other relaxation rates remain unchanged.

Figures 31.7c and d presents an example of simulated mode-locked QD laser containing an amplifying and a saturable absorber sections S_1 and S_2. Panel (c) of this figure gives an evidence of strongly asymmetric pulses with a broad trailing edge plateau. Our theoretical analysis has shown that such pulses arise mainly due to noninstant carrier transitions between the CR, ES, and GS of the QD laser shown in panel (d) of the same figure. The presence of these transitions exert a smoothening effect on all spatial/temporal carrier and field intensity distributions and, in turn, imply a broadening of the trailing edge of the pulse. We have also found that an increase of the intradot transition rates leads to a reduction of the filtering effect and, hence, to a growth of the pulse peak intensity and narrowing of the pulse and its trailing edge. More details on our analysis as well as experimental demonstration of such asymmetric pulses can be found in Reference [10,49].

Further modifications of the TW model can be used for more precise simulations of QD lasers. For example, one can improve the model of carrier transitions (Equation 31.25) by separate consideration of electrons and hole densities [50]. To allow a simultaneous radiation on the spectrally well-separated ground and ES, one can introduce another pair of TW equations for optical fields [51]. An inhomogeneous spectral broadening and an accompanying description of the radiation at GS and ES can also be modeled by an introduction of multiple sets of carrier rate and polarization equations representing carrier transitions within the QDs of different size and their impact on the laser emission at different wavelengths [52].

Nonlinear Gain and Refractive Index Functions:

In the earlier discussion, the gain and the refractive index dependence on the carrier density N was modeled by linear functions related to each other by the linewidth enhancement factor α_H. This modeling approach is reasonable for small and slow variation of carrier density N, but can fail once N exhibits some significant changes, see, e.g., Figure 31.5d, where a variation of the spatially distributed $N(z,t)$ was of the order of the mean value of the carrier density. In such situations, one should better use nonlinear peak gain functions, $g(N)$, which can better represent measured or precalculated gain spectra profiles. For this reason, the following logarithmic gain peak function dependence on the carrier density is frequently used:

$$g(N) = \Gamma g' N_{\mathrm{tr}} \ln\left(\frac{\max\{N, N^*\}}{N_{\mathrm{tr}}}\right), \qquad \tilde{n}(N) = \frac{\alpha_H g(N)}{2}. \tag{31.26}$$

Here, N^* indicates a cutoff carrier density value, which prevents the convergence $g(N) \to -\infty$ with carriers $N \to 0$. These expressions for the gain and index change functions replace the relations 31.3 used in the TW models discussed earlier.

Another issue is related to the linewidth enhancement factor α_H. Initially, this factor was used to relate gain and refractive index functions at a fixed value of N. Such approach implies a rather simple model for propagation factor, β, and can be quite useful when performing an advanced analysis of model equations. In reality, however, the ratio between the gain and refractive index is not a constant, but a function depending on carrier density, temperature, and several other factors not discussed in this chapter. Thus, an experimental estimation of this factor in the semiconductor laser operating at different conditions or using different methods can lead to rather different values of α_H. For this reason, it can be preferable to use separately defined nonlinear peak gain and index change functions $g(N)$ and $\tilde{n}(N)$. These functions depend on the properties of the semiconductor material and the design of the device, and, therefore, should be adjusted individually for each considered laser.

A satisfactory description of these functions for a broad class of semiconductor lasers is given by the logarithmic, and the square-root-like expressions [38]

$$g(N) = \Gamma g' N_{\mathrm{tr}} \ln\left(\frac{\max\{N, N^*\}}{N_{\mathrm{tr}}}\right), \qquad \tilde{n}(N) = \tilde{n} + \alpha_H \Gamma g' \sqrt{N \cdot N_{\mathrm{tr}}}. \tag{31.27}$$

Here, $\tilde{n}$ represents the offset of the refractive index change function, $\Gamma g' = \partial_N g(N_{\text{tr}})$, and $\alpha_H = 2\partial_N \tilde{n}(N_{\text{tr}})/\partial_N g(N_{\text{tr}})$ is the linewidth enhancement factor evaluated at the transparency carrier density N_{tr}.

We should admit, however, that linear formulas 31.3 with slightly corrected factors g', α_{H}, N_{tr}, and a proper selection of δ_0 can be used for approximation of nonlinear functions 31.27 not only in the vicinity of N_{tr} but also over a larger range of densities N including the threshold density N_{th}. In many cases, the simplifications of the gain and refractive index functions still imply qualitatively the same results when performing simulations of MSLs with varying parameters [53].

31.8 Multisection Lasers and Coupled Laser Systems

A vast variety of MSLs and coupled laser systems can be represented as a set of differently interconnected laser sections, each characterized by its material and geometry parameters. To distinguish these parameters or functions attributed to different laser sections, we shall use the lower indices. For example, α_k, $g_k(N)$, and a_k denote the field losses, the gain function and the left-edge coordinate of the section S_k. Note also that for unique identification of longitudinal coordinate z within all laser sections, different sections of our device are represented by nonoverlapping intervals (a_k, b_k).

According to our laser device construction, for any edge of all sections S_k, we can attribute a unique junction J_l. On the other hand, each junction has, at least, one section joining it from one or another side, see, e.g., Figure 31.1, where MSLs are represented as sets of laser *sections* mutually interconnected through different *junctions*. To explain the relations between section edges, corresponding junctions, and applied optical injections, we use the following notations in the sequel of this chapter. By l we denote the index of the junction J_l, as well as optical fields and the section edges attributed to this junction. l' (l'') is the vector of length $|l'|$ $(|l''|)$ containing indices of the sections connected to J_l by their left (right) edge $a_{l'_j}$ $(b_{l''_j})$; see Figure 31.8a. $l^e = |l'| + |l''| \geq 1$ is a total number of such section edges connected to J_l. By $o_l(t)$ and $E_l^{\text{out}}(t)$ we denote the applied optical injection and the recorded emission at the same junction, see solid and dashed thick light gray arrows in Figure 31.8a. When the injection or emission at J_l is absent, the corresponding function is simply set to zero.

General field scattering conditions at the arbitrary junction J_l are defined by the $l^e \times l^e$ dimensional complex field scattering matrix $\mathcal{T}_l$, $l^e \times 1$ dimensional injection distribution matrix $\mathcal{T}_l^i$, and $1 \times (l^e + 1)$

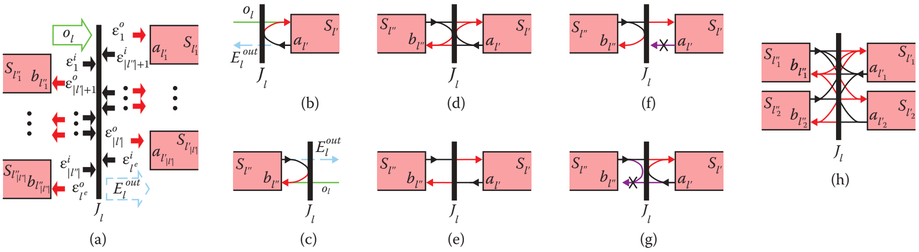


FIGURE 31.8 Schematic representation of different interfaces between the sections at the junction J_l. (a) general configuration, (b) and (c) field reflection and optical injection at laser facets, (d) and (e) reflecting (transmission and reflection) and trivial (transmission only) interface of two section edges, (f) and (g) directionally absorbing interfaces of two section edges for modeling of a master–slave laser system, and (h) transmitting/reflecting interface of four section edges for modeling ring lasers with an outcoupling waveguide. Black segments and gray frames represent junction J_l and all section edges connected to this junction, respectively. Black and dark gray arrows in all diagrams represent all components of the vector fields $\mathcal{E}_l^i$ and $\mathcal{E}_l^o$, respectively. Optical injections o_l and emitted fields E_l^{out} are shown by solid and dashed light gray arrows. Crossed arrows in panels (f) and (g) represent a full absorption of the corresponding fields.

dimensional outcoupling matrix $\mathcal{T}_l^o$:

$$\begin{aligned}
&\mathcal{E}_l^o = \mathcal{T}_l\mathcal{E}_l^i + \mathcal{T}_l^i o_l, \quad E_l^{\text{out}} = \mathcal{T}_l^o \begin{pmatrix} \mathcal{E}_l^i \\ o_l \end{pmatrix}, \qquad \text{where} \\
&\mathcal{E}_l^o = \left(E^+_{l'_1}(a_{l'_1},t), \dots, E^+_{l'_{|l'|}}(a_{l'_{|l'|}},t), E^-_{l''_1}(b_{l''_1},t), \dots, E^-_{l''_{|l''|}}(b_{l''_{|l''|}},t)\right)^T, \\
&\mathcal{E}_l^i = \left(E^+_{l''_1}(b_{l''_1},t), \dots, E^+_{l''_{|l''|}}(b_{l''_{|l''|}},t), E^-_{l'_1}(a_{l'_1},t), \dots, E^-_{l'_{|l'|}}(a_{l'_{|l'|}},t)\right)^T.
\end{aligned} \tag{31.28}$$

The vector functions $\mathcal{E}^i$ and $\mathcal{E}^o$ (see black and dark gray arrows in Figure 31.8a) denote the internal optical fields, which are incident into the junction from all adjacent sections and are scattered from the junction back into these sections, respectively.

In most cases, the interfaces between the sections are much simpler. For example, the scattering matrices at the facets of the solitary laser (Figures 31.8b and c as well as junctions J_1 and J_2 in Figure 31.2a) are determined by the boundary conditions (Equation 31.7), i.e.,

$$\begin{cases} \mathcal{T}_l = -r_l^*, \ \mathcal{T}_l^i = 1, \ \mathcal{T}_l^o = (t_l, 0), & \text{single "left" edge } a_{l'}, \ |l'| = 1, \ |l''| = 0 \\ \mathcal{T}_l = r_l, \ \mathcal{T}_l^i = 1, \ \mathcal{T}_l^o = (t_l, 0), & \text{single "right" edge } b_{l''}, \ |l''| = 1, \ |l'| = 0 \end{cases}, \tag{31.29}$$

where r_l and t_l are field reflection and transmission coefficients,

$$|r_l| \le 1, \qquad t_l \le \sqrt{1 - |r_l|^2}. \tag{31.30}$$

Another frequently used case in MSLs is the interface of two adjacent sections (Figures 31.8d through g). At such interfaces, we have no optical injections and field emission, so that we can set $\mathcal{T}_l^i = (0,0)^T$ and $\mathcal{T}_l^o = (0,0,0)$. The scattering of the field at J_l, in this case, is entirely defined by the 2×2 dimensional matrix

$$\mathcal{T}_l = \begin{pmatrix} t_l & -r_l^* \\ r_l & t_l \end{pmatrix}, \tag{31.31}$$

where t_l and r_l satisfy the conditions 31.30 (see Figure 31.8d and, e.g, J_3 in Figure 31.1a). Here, the nonvanishing reflections r_l can appear, e.g., due to different heterostructure of the adjacent sections. In the simplest case of $r_l = 0$ and $t_l = 1$, $\mathcal{T}_l$ is an identity matrix, and the interface admits a full transmission of the optical fields (see Figure 31.8e and J_2 in Figure 31.1a).

When modeling the master–slave laser system (S_1 and S_2 in Figure 31.1d), only one-directional field propagation should be allowed in the air gap between two lasers (section S_3 in the same figure). This effect can be achieved by modification of otherwise standard scattering matrices $\mathcal{T}_l$ (Equation 31.31) at one of the gap section edges. One can model a full absorption of the backward incident beam at the interface of the master laser and the gap section (see Figure 31.8f and J_2 of Figure 31.1d), or prohibit the field backscattering into the air gap at the interface of the slave laser and the gap section (Figure 31.8g and J_3 of Figure 31.1d). Formally, both these situations can be defined by the scattering matrices

$$\mathcal{T}_j = \begin{pmatrix} t_j & 0 \\ r_j & 0 \end{pmatrix} \ \text{(master-gap interface)}, \qquad \mathcal{T}_j = \begin{pmatrix} t_j & -r_j^* \\ 0 & 0 \end{pmatrix} \ \text{(gap-slave interface)}.$$

More complicated situations occur at the junctions connecting more than two section edges of the MSL. For example, Figure 31.8h and J_1 of Figure 31.1e represent an interface connecting two "left" and two "right" section edges. This situation is used for modeling of a localized coupling of the ring laser (section S_1 in Figure 31.1e) and the outcoupling waveguide (S_2 and S_3 in the same figure).

Similarly to the previously discussed case, the optical injection- and field emission-relevant matrices can be defined by $\mathcal{T}_l^i = (0,0,0,0)^T$ and $\mathcal{T}_l^o = (0,0,0,0,0)$. By assuming a nonvanishing field reflection r_l at the ring laser part of this junction (section edges $b_{l_1''}$, $a_{l_1'}$ in Figure 31.8h or b_1, a_1 in Figure 31.1e), we model a *localized* linear backscattering of the fields [23,37]. The 4×4 dimensional scattering matrix $\mathcal{T}_l$, in this case, can be defined as

$$\mathcal{T}_l = \begin{pmatrix} t_l & i\tilde{t}_l & -r_l^* & 0 \\ i\tilde{t}_l & t_l & 0 & 0 \\ r_l & 0 & t_l & i\tilde{t}_l \\ 0 & 0 & i\tilde{t}_l & t_l \end{pmatrix}, \qquad t_l^2 + \tilde{t}_l^2 + |r_l|^2 \le 1. \tag{31.32}$$

Here, t_l is a real field amplitude transmission factor within the same (ring or outcoupling) waveguide and $i\tilde{t}_l$ is an imaginary coefficient representing part of the field amplitude, which is outcoupled from the ring or transmitted into the ring from the external waveguide. It is noteworthy that a proper estimation of the transmission–reflection–outcoupling matrix $\mathcal{T}_l$ in the ring laser case requires some appropriate measurements or an advanced modeling. Such modeling should take into account the curvature of the ring cavity, the length of the coupling regions, the field diffraction, and the overlapping of the lateral modes in the coupling region [54]. Moreover, the coefficients of the scattering matrix are, in general, frequency dependent. In our TW modeling approach, we use constant coefficients describing field scattering at the central reference frequency.

31.9 Simulations of Nontrivial MSL Device

The concept of differently interconnected *sections* and *junctions* allows modeling rather complicated MSLs. One of such nontrivial configurations is a semiconductor ring laser with four separate branches of the filtered optical feedback, see Figure 31.9a. The multichannel feedback scheme of this laser admits a fast switching between steady states determined by the resonances of the ring laser and the wavelengths of the activated filtering channels [55].

The gray-shaded frames in Figure 31.9a represent device sections of different types. Namely, we distinguish here the amplifying sections $S_{A\cdot}$, where the field and carrier dynamics is governed by the full TW model (Equations 31.5, 31.14, 31.21, 31.23, 31.24, 31.26), and two kinds of passive sections, $S_{P\cdot}$ and $S_{F\cdot}$, where gain and refractive index functions are set to zero, allowing to ignore the carrier rate equations at all. The notations of all sections in the section indexes are made according to the cardinal directions "n," "e," "s," and "w".

Almost all parameters of the TW model in all sections of our MSL are the same as in Figure 31.6. A few exceptions are parameters $\alpha_H = -4$ and $\bar{\gamma}_\lambda = 100\,\text{nm}$. In the passive waveguiding sections, $S_{P\cdot}$ (medium gray), we assume $|S_{Pe}| = |S_{Pw}| = 330\,\mu\text{m}$, $|S_{Pne}| = |S_{Pnw}| = 50\,\mu\text{m}$, $|S_{Pse}| = |S_{Psw}| = 2500\,\mu\text{m}$, and neglect the gain dispersion, $\bar{g} = 0$. In the passive filtering sections, $S_{F\cdot}$ (light gray), we assume $|S_{F\cdot}| = 530\,\mu\text{m}$ and significantly modify the profile of Lorentzian gain dispersion by setting $\bar{g} = 5 \cdot 10^4\,\text{m}^{-1}$ and $\bar{\gamma}_\lambda = 4\,\text{nm}$. The relative peak wavelengths of four filtering branches (sections S_{Fwj} and S_{Fej}, $j = 1, \ldots, 4$) are $\bar{\lambda} = -2, -0.67, 0.67$, and $2\,\text{nm}$, respectively. Finally, in the amplifying sections (dark gray) within the primary ring laser, S_{Ajk}, $j = n, s$ and $k = e, w$, we use $|S_{A\cdot}| = 380\,\mu\text{m}$, $\varepsilon_{Gs} = 6 \cdot 10^{-24}\,\text{m}^3$, $\varepsilon_{Gc} = 2\varepsilon_{Gs}$, and the bias currents $I = 26\,\text{mA}$, which is 1.5 times higher than the lasing threshold in the laser without feedback. In the amplifying sections belonging to the four filtering branches, S_{Asj}, $j = 1, \ldots, 4$, we assume $|S_{A\cdot}| = 190\,\mu\text{m}$ and $\varepsilon_{Gc} = \varepsilon_{Gs} = 9 \cdot 10^{-24}\,\text{m}^3$. Once the bias current in these sections is zero, $I = 0\,\text{mA}$,

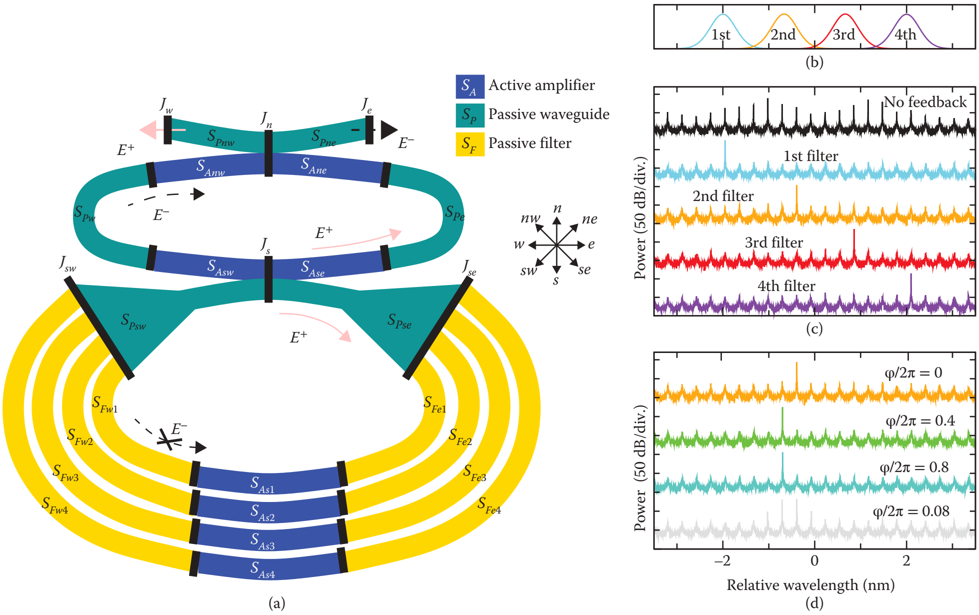

1st
2nd
3rd
4th
S_A Active amplifier
S_P Passive waveguide
S_F Passive filter
J_w
J_n
J_e
S_{Pnw}
S_{Pne}
E^+
E^-
S_{Anw}
S_{Ane}
S_{Pw}
S_{Pe}
E^-
E^+
J_s
J_{sw}
J_{se}
S_{Asw}
S_{Ase}
S_{Psw}
S_{Pse}
E^+
n
nw
ne
w
e
sw
se
s
S_{Fw1}
S_{Fw2}
S_{Fw3}
S_{Fw4}
E^-
S_{Fe1}
S_{Fe2}
S_{Fe3}
S_{Fe4}
S_{As1}
S_{As2}
S_{As3}
S_{As4}
No feedback
1st filter
2nd filter
3rd filter
4th filter
Power (50 dB/div.)
φ/2π = 0
φ/2π = 0.4
φ/2π = 0.8
φ/2π = 0.08
Power (50 dB/div.)
−2
0
2
Relative wavelength (nm)
(a) (b) (c) (d)

FIGURE 31.9 (a) Scheme of the semiconductor ring laser with four branches of filtered and amplified unidirectional optical feedback. Black segments and gray-shaded frames indicate junctions and different sections of the MSL. Solid light-gray and dashed black arrows show propagation directions and the emission of the fields E^+ and E^-, respectively. (b) Transmission spectra of four filtering branches. Maximal transmission at $I_{Asj} = 10\,\text{mA}, j = 1, \ldots, 4$ is approximately 1.6. (c) Stabilization of the multimode behavior of the ring laser (black) by the single-branch filtered feedback (gray). (d) Dependence of the lasing wavelength on the feedback phase once the second filtering branch is activated.

the feedback branches are efficiently absorbing the optical fields. To activate one of the feedback branches, we set the corresponding injection $I = 10\,\text{mA}$.

The field transmission and reflection conditions at J_n and J_s are given by the conditions 31.32 with $r = 0$, $t = \sqrt{0.8}$, and $\tilde{t} = \sqrt{0.2}$. At J_{se} and J_{sw}, we neglect all possible reflectivity, admit full-field transmission from filtering branches to the passive waveguide sections S_{Psw} or S_{Pse}, equally distribute the intensity of the optical field E^+ propagating from S_{Pse} to the filtering branches, and fully absorb E^- at J_{sw}:

$$E^-(b_{Pse}, t) = \textstyle\sum_{j=1}^{4} E^-(a_{Fej}, t), \qquad E^+(a_{Fej}, t) = \sqrt{\tfrac{1}{4}}\, E^+(b_{Pse}, t), \quad j = 1, \ldots, 4;$$

$$E^+(a_{Psw}, t) = \textstyle\sum_{j=1}^{4} E^+(b_{Fwj}, t), \qquad E^-(b_{Fwj}, t) = 0, \quad j = 1, \ldots, 4.$$

At J_w and J_e, the fields E^+ and E^- are emitted from our MSL. Here, the field reflection–transmission conditions are given by (Equation 31.29) with the reflectivity factors $r_w = r_e = 0.1$. All other junctions of this MSL are trivial, i.e., the optical fields cross the interfaces according to the relations 31.31 with $r_l = 0$ and $t_l = 1$.

A series of simulations represented in the remaining panels of Figure 31.9 are in good agreement with the experimental results reported in Reference [55]. First of all, panel (b) shows the transmission spectra (modulus of the wavelength-dependent complex transmission function) of the optical fields E^+ propagating through each of four optical feedback branches activated by the injected current into the corresponding amplifying section. For 10 mA injections used in these simulations, the peak amplitude transmission is around 1.6.

Panels (c) and (d) of Figure 31.9 show simulated optical spectra of the emitted field at the “west” facet J_w of the MSL for different operation conditions. An upper black curve in panel (c) represents the optical spectrum in the case of deactivated feedback branches. Multiple significant spectral peaks with the mode separation corresponding to the field round-trip time in the ring laser indicate a multimode lasing of the laser. An optical field E^+ propagating along the filtering branches, however, is not entirely absorbed. For higher ring laser injections, we observed the steady states determined by a single ring resonance mode. In these cases, the amplifier within corresponding filtering branch was optically pumped, and the related peak amplitude transmission was around 0.2. The competition between nominally equivalent unpumped filtering branches, however, does not allow predicting the lasing wavelength of such a steady state.

The four lower spectra in Figure 31.9c represent switching between different optical modes by activation of the corresponding filter and deactivation of the remaining ones. A close inspection of these spectra shows, that whereas the first and the fourth filters select the resonance modes which are closest to the filter peak position, the third, and, especially, the second filter prefers modes admitting smaller optical feedback. We have found, that this mode selection is related to the phase of the optical field within the filtering branch. Figure 31.9d demonstrates, how tuning of the feedback phase within the second filtering branch (realized by variation of the detuning factor $\delta_{0,Fw2}$) implies changes between the resonant modes located within the filtering band.

In conclusion, we have simulated the MSL consisting of 22 sections interconnected at 18 junctions. Our theoretical findings were in a good qualitative agreement with experimental observations of similar ring laser device reported in Reference [55].

31.10 Beyond Numerical Simulations of the TW Model

In the previous sections, we have introduced different modifications of 1+1 dimensional TW model suited for simulations of various MSL devices and coupled laser systems. In the remaining part of this chapter, we introduce the concept of instantaneous optical modes and present several applications of these modes for an advanced analysis of MSLs. In all these cases, we consider MSLs without optical injection and neglect a contribution of Langevin noise term $F_{sp}^{\pm}$, which is of minor importance in the lasers operating well above threshold.

Instantaneous Optical Modes:

The concept of optical modes plays a significant role in understanding laser dynamics in general. They represent the natural oscillations of the electromagnetic field and determine the optical frequency and the lifetime of the photons contained in the given laser cavity. The *instantaneous* optical modes correspond to a fixed *instant* distribution of the propagation factor β [26].

In general, compared to a variation of the optical fields, the changes of the carrier density N are slow. The change of N is mainly determined by the carrier relaxation time which, typically, is measured in nanoseconds (or tens of picoseconds when considering saturable absorbers). On the other hand, picosecond or sub-picosecond time windows are sufficient for significant changes of the photon densities. Since the gain compression for small and moderate field intensities is also small, the propagation factor β experiences only minor modifications in the picosecond range. For this reason, in the remaining part of this chapter we analyze the field equations for the frozen distribution of the propagation factor $\beta(z, t_0)$ at the time instant t_0.

The instantaneous optical modes of MSLs are pairs $(\Omega(\beta), \Theta(\beta, z))$ of complex frequencies Ω and vector-functions $\Theta = (\Theta_E^+, \Theta_E^-, \Theta_P^+, \Theta_P^-)^T$, where imaginary and real parts of $\Omega(\beta)$ are mainly defining the angular frequency and the damping of the mode, whereas $\Theta(\beta, z)$ determines the spatial distribution of the mode.

Complex frequencies Ω and vector-functions $\Theta(\beta, z)$ solve the linear system of algebro-differential equations

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}z}\Theta_E^+ = -i\mathcal{B}(\Omega)\Theta_E^+ - i\kappa\Theta_E^- \\ \frac{\mathrm{d}}{\mathrm{d}z}\Theta_E^- = i\mathcal{B}(\Omega)\Theta_E^- + i\kappa\Theta_E^+ \\ \Theta_P^\pm(\beta, z) = \frac{\bar{\gamma}/2}{\bar{\gamma}/2 + i(\Omega - \bar{\omega})}\Theta_E^\pm(\beta, z) \end{cases}, \quad z \in S_k\Big|_{k=1}^{n}, \qquad \theta_l^o(\beta) = \mathcal{T}_l\theta_l^i(\beta)\Big|_{l=1}^{m}, \tag{31.33}$$

obtained by assuming a nonvarying in time propagation factor, $\partial_t\beta = 0$, and substituting the expressions

$$E^\pm(z, t) = \Theta_E^\pm(\beta; z)e^{i\Omega(\beta)}, \qquad P^\pm(z, t) = \Theta_P^\pm(\beta; z)e^{i\Omega(\beta)}$$

into the field equations 31.14 within each of n sections S_k, and boundary conditions 31.28 at each of m junctions J_l. Similarly to the vector functions $\mathcal{E}_l^o$ and $\mathcal{E}_l^i$ in Equation 31.28, complex vectors $\theta_l^i(\beta)$ and $\theta_l^o(\beta)$ in Equation 31.33 represent functions $\Theta_E^\pm(\beta, z)$ at the section edges $z = a_{l'_j}$ or $z = b_{l''_j}$ connected by the junction J_l. The function $\mathcal{B}(\Omega)$ entering Equations 31.33 is defined in 31.16.

Each pair of linear ODEs in Equation 31.33 can be solved by the transfer matrix† 31.11 with the coefficients nonlinearly depending on still unknown complex frequency Ω. These matrices define $2n$ homogeneous linear equations relating $4n$ components of the complex vector $\mathcal{S} = (s_1, \ldots, s_{4n})^T$ representing field functions $\Theta_E^\pm(\beta, z)$ at both edges of all sections S_k. Another $2n$ homogeneous linear equations relating the same complex numbers are given by the field scattering matrices $\mathcal{T}_l$ at all junctions J_l. In such a manner, we build a linear $4n$ dimensional algebraic system

$$\mathcal{M}(\beta, \kappa; \Omega)\mathcal{S} = 0,$$

determined by a sparse $4n \times 4n$ dimensional matrix $\mathcal{M}$. Nontrivial solutions $\mathcal{S}$ (i.e., nontrivial functions Θ of the problem 31.33) are available only for those Ω which are the complex roots of the complex characteristic equation

$$\det \mathcal{M}(\beta, \kappa; \Omega) = 0. \tag{31.34}$$

The finite number of these roots can be found using Newton iterations and the homotopy method; see Reference [26] for more details.

It is noteworthy that linear configurations of MSLs admit rather simple expressions of the characteristic equations 31.34 involving the response functions $F_l(z, \Omega)$ and $F_r(z, \Omega)$ defined at some longitudinal position z of the MSL:

$$\det \mathcal{M}(\beta, \kappa; \Omega) = 0 \qquad \Leftrightarrow \qquad F_l^{-1}(z, \Omega) = F_r(z, \Omega). \tag{31.35}$$

For example, for the solitary lasers considered in Figure 31.3b and c, $F_l(b_1, \Omega)$ is defined in Equation 31.13, whereas $F_r(b_1, \Omega) = r_2$. In the general case, functions F_l and F_r are defined by a consequent superposition of the sectional transfer matrices $M(\beta, \kappa, \Omega)$ and the left-to-right or right-to-left junction-transfer matrices

$$\mathcal{T}_{j,22}^{-1}\begin{pmatrix} \det \mathcal{T}_j & \mathcal{T}_{j,12} \\ -\mathcal{T}_{j,21} & 1 \end{pmatrix} \qquad \text{or} \qquad \mathcal{T}_{j,11}^{-1}\begin{pmatrix} 1 & -\mathcal{T}_{j,12} \\ \mathcal{T}_{j,21} & \det \mathcal{T}_j \end{pmatrix};$$

see References [23,26,28] for more details.

---

† In the case of nonvanishing $\Delta_\beta = \frac{\beta^+ - \beta^-}{2}$, the transfer matrix in each section S_k should be constructed for $\bar{\beta} = \frac{\beta^+ + \beta^-}{2}$ and later multiplied by the factor $e^{-i\langle\Delta_\beta\rangle_k|S_k|}$ [23].

The calculated optical field function $\Psi(z,t) = (E^+, E^-, P^+, P^-)^T$ can be represented as a superposition of the suitably normalized vector functions $\Theta(\beta, z)$ which are slowly changing with a variation of the propagation factor $\beta(z,t)$:

$$\Psi(z,t) = \sum_{j=1}^{\infty} f_j(t)\Psi_j(\beta(z,t), z). \tag{31.36}$$

Here, $f_j(t)$ is the complex amplitude of the mode, which can denote the mode contribution to the field emission at the laser facet a_k once normalization of mode functions assumes $\Theta_E^-(\beta, a_k) = 1$. According to our notations, index $_1$ denotes the most significant mode having a largest (instant) amplitude $|f|$ or a lowest damping $\Im\Omega$. An increasing index means a decreasing importance of the mode. This numbering does allow us to achieve good approximations of the field function $\Psi(z,t)$ already by low-dimensional truncated mode expansions 31.36.

Calculation of optical modes and expansion of the field function into the modal components can give a broad understanding of different operating regimes in MSLs and explain parameter change-induced transitions between these states observed in simulations and experiments. We have applied our mode analysis for interpretation of experimental observations in different MSLs. Namely, we have explained a stable operation of ring lasers at alternating oscillation or bi- and unidirectional steady state regimes [23]; almost periodically reappearing state transitions and estimation of thermal tuning parameters in master-oscillator power-amplifier device [38,56], DBR laser [14], or ECDL [15]; and strongly asymmetric pulse shapes in QD mode-locked laser [49]. More theoretical examples of our mode analysis can be found in Reference [26].

Steady States:

Any stationary (rotational wave) state of the MSL is determined by an optical mode with a *real* mode frequency

$$(\Psi(z,t), N(z,t)) = \left(\hat{f}\,\Theta(\hat{\beta}, z)e^{i\hat{\omega}t}, \hat{N}(z)\right), \quad \text{where } \Omega(\hat{\beta}) = \hat{\omega} \in \mathbb{R},$$

and $\hat{\beta}(z)$ is a constant in time spatially distributed propagation factor. Let us consider the TW model with sectionally averaged carrier density and neglected nonlinear gain compression given by Equations 31.4, 31.5, 31.14, 31.18, 31.15, and 31.28. In this case, all steady states are fully defined by a set of n_a+2 real numbers $\left(\hat{\omega}, |\hat{f}|^2, \hat{N}_1, \ldots, \hat{N}_{n_a}\right)$, which are a mode frequency, a mode intensity, and sectionally averaged carrier densities within all n_a "active" sections having nonvanishing functions g and $\tilde{n}$. The rotational invariance of the TW model implies freedom in selection of the phase of the complex mode amplitude $\hat{f}$. The set of these real numbers is a root of a nonlinear algebraic system of one complex characteristic equation and n_a real steady-state carrier rate equations:

$$\begin{aligned} &\det \mathcal{M}\left(\beta(\hat{N}), \kappa; \hat{\omega}\right) = 0, \\ &\frac{I_r}{q\sigma_r|S_r|} - \mathcal{R}(\hat{N}_r) - |\hat{f}|^2 \frac{c_0}{n_g} G(\hat{N}_r, \hat{\omega})\langle\left(\Theta_E, \Theta_E\right)\rangle_r = 0, \quad r = 1, \ldots, n_a. \end{aligned} \tag{31.37}$$

Here, the frequency-dependent gain function G is defined in Equation 31.17, whereas the sectional average $\langle\left(\Theta_E, \Theta_E\right)\rangle_r$ can be expressed as a function of $\hat{\omega}$ and $\hat{N}$ [28].

In the case of the single active section, $n_a = 1$, the steady state frequency $\hat{\omega}$ and threshold carrier density $\hat{N}_1$ can be directly found from the characteristic equation, whereas the remaining equation determines the value of $|\hat{f}|^2$. Assume that the single active section S_1 of linear MSL is located on the left side of the device (see Figure 31.1b), and the optical fields within the adjacent passive section S_2 are governed by the simple relations 31.9. For an illustration of this situation, we have considered a three section passive dispersive reflector laser consisting of the active DFB, passive DBR, and another passive phase tuning section in between. Due to the relation 31.35, we can replace the complex characteristic equation by a couple of real-valued equations,

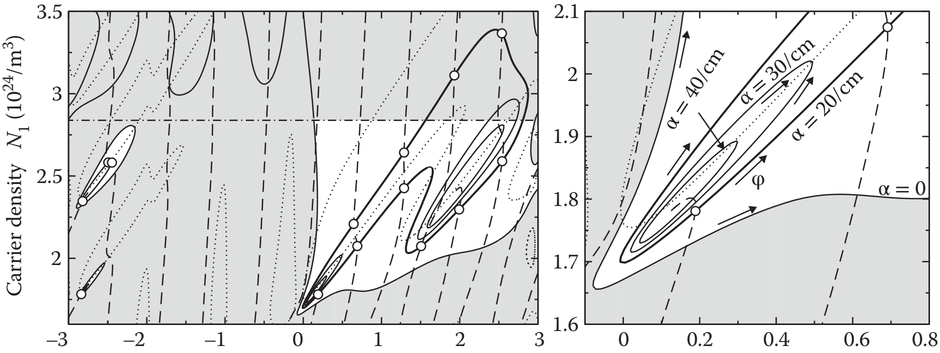
$$F_l^{-1}(a_2,\hat{\omega}) = F_r(a_2,\hat{\omega}) = e^{-\alpha_2|S_2|}e^{i(\varphi-2\hat{\omega}\tau_2)}F_r(b_2,\hat{\omega}) \qquad \Leftrightarrow$$
$$\begin{cases}\tilde{\mathcal{M}}_\alpha(\hat{N}_1,\hat{\omega}) = e^{\alpha_2|S_2|} \\ 2\hat{\omega}\tau_2 - \tilde{\mathcal{M}}_\varphi(\hat{N}_1,\hat{\omega}) = \varphi\end{cases}, \quad \text{where } \tilde{\mathcal{M}}(\hat{N}_1,\hat{\omega}) = \tilde{\mathcal{M}}_\alpha e^{i\tilde{\mathcal{M}}_\varphi} = F_l(a_2,\omega)F_r(b_2,\hat{\omega}) \quad . \tag{31.38}$$

This formulation suggests a simple way for finding the steady states. Namely, each of these equations for fixed parameters α_2 and φ defines one or several curves in frequency ω – carrier threshold N_1 domain, see solid and dashed curves in Figure 31.10. The intersections of these lines determine the steady state pairs $\hat{\omega}, \hat{N}_1$ (hollow bullets in the same figures). It is noteworthy that to any point in the $\omega - N_1$ domain one can attribute a unique triple of loss, phase, and mode power parameters α_2, φ, and $|\hat{f}|^2$. ω and N_1 within the gray shading regions of Figure 31.10 represent the unphysical steady states corresponding to negative damping in the passive section ($\alpha_2 < 0$) and negative mode intensity ($|\hat{f}|^2 < 0$) due to insufficient pumping of the active section.

The fixed level lines of $\tilde{\mathcal{M}}_\alpha$ determined by larger losses $\alpha_2 = 30$ and 40/cm (thin solid-line ellipses located inside of thick solid curves in Figure 31.10) are shrinking toward central points, which are resonances of the solitary DFB laser. Accordingly, the (odd) number of steady states on each ellipse is also reduced. The saddle-node bifurcation that is responsible for creation or annihilation of the steady state pair, occurs at those φ and α_2, where corresponding fixed level lines of $\tilde{\mathcal{M}}_\varphi$ and $\tilde{\mathcal{M}}_\alpha$ become tangent to each other. The last condition formally given by

$$\mathcal{M}_{SN}(\hat{\omega},\hat{N}_1) \overset{def}{=} \partial_\omega\tilde{\mathcal{M}}_\alpha\partial_{N_1}\tilde{\mathcal{M}}_\varphi - \partial_\omega\tilde{\mathcal{M}}_\varphi\partial_{N_1}\tilde{\mathcal{M}}_\alpha = 0$$

is satisfied on the dotted lines of Figure 31.10.



(a) (b)

FIGURE 31.10 Stationary states (frequencies and threshold carrier densities) of a three section DFB laser. Panels (a) and (b) show a global overview and a zoomed-in region close to the minimal threshold mode with $\hat{\omega} \approx 0$. Solid curves represent steady states for an arbitrary φ but only a few fixed α_2. Dashed curves indicate steady states for arbitrary α_2 and fixed $\varphi = 0$. Dotted curve shows positions where saddle-node bifurcations hold. Empty bullets: steady states for $\alpha_2 = 20$/cm and $\varphi = 0$. Dash-dotted line indicates maximal N_1 which can be achieved for considered bias current I. Gray shading corresponds to unphysical states. Small arrows in panel (b) indicate directions of the steady state shift along the fixed loss lines for growing φ. All parameters as in Reference [26].

In general, the interpretation of the steady states for large α_2 (small feedback) is in good agreement with the analysis of the external cavity modes in the Lang–Kobayashi (LK) model of lasers with delayed feedback [18,27]. A decrease of α_2 leads to blowing up and collision of different ellipses. This scenario involves multiple modes of the solitary DFB laser and can be no more explained by the LK model.

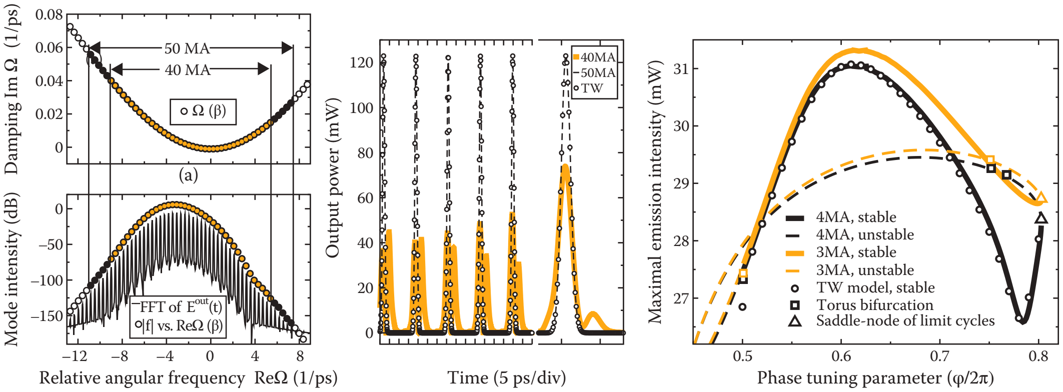
Mode Approximation Systems:

For some MSL devices, the TW model (Equations 31.2 through 31.5, and 31.14, 31.15, 31.28) with sectionally averaged carrier densities, linear gain and index change functions, and neglected gain compression terms can be reduced to the finite-dimensional system of ODEs describing an evolution of q complex mode amplitudes f and real sectionally averaged carrier densities N within n_a active sections of MSL:

$$\begin{aligned} \dot{f}_k &= i\Omega_k(N)f_k + \sum_{l=1}^{q}\left(\sum_{r=1}^{n_a} K^r_{k,l}(N)\dot{N}_r\right)f_l, \quad k = 1,\dots,q; \\ \dot{N}_r &= \frac{I_r}{q|S_r|\sigma_r} - \mathcal{R}(N_r) - \Re\sum_{k,l=1}^{q} L^r_{k,l}(N)f_k^{*}f_l, \quad r = 1,\dots,n_a. \end{aligned} \tag{31.39}$$

This mode approximation (MA) system follows from the substitution of the truncated field expansion (Equation 31.36) into the TW model equations and projection of the resulting field equations onto the linear subspace defined by each of q modes. The nonadjoint nature of the field evolution operator and small but nonvanishing time derivatives of propagation factor $\beta(N)$ imply the appearance of the mode coupling terms $K^r_{k,l}\dot{N}_r$. For the derivation of the MA equations and analytic expressions for carrier and mode frequency Ω-dependent mode-coupling functions, $K^r_{k,l}$ and $L^r_{k,l}$, see Reference [28].

To check the precision of our MA system, we have performed simulations of the TW model and two related MA systems describing the evolution of a mode-locked laser consisting of a saturable absorber and an amplifying section (case of $n_a = 2$). The solid black curve and hollow bullets in Figures 31.11b and c show typical optical spectrum and time trace of the mode-locking pulsations obtained by numerical integrations of the TW model. To determine the most relevant complex mode frequencies $\Omega(N)$



(b) (c) (d)

FIGURE 31.11 Calculated (a) complex frequencies Ω, (b) optical spectra as a Fourier-transformed field $E^{-}(a_1, t)$ (black solid curve) and as a discrete set of mode intensities $|f|^2$ versus $\Re\Omega$ (bullets), and (c) a comparison of the calculated transients of the TW model (bullets) and the reduced ODE systems determined by 40 (solid gray) or 50 (dashed black) optical modes in a two-section mode-locked laser. Full black and small gray bullets in panels (a) and (b) indicate the modes used for the construction of 40MA and 50MA systems. (d) Numerical path following of the stable periodic solution of the TW model (bullets) and periodic orbits of 3MA (gray) and 4MA (black) systems in a three-section phase-controlled mode-beating DFB laser [12,28]. Solid and dashed curves indicate stable and unstable orbits. Empty boxes and triangles denote torus and fold bifurcations, respectively.

(Figure 31.11a) and the coefficients f in the field expansion 31.36 (bullets in panel [b]), we have used the carrier densities $N = (N_1, N_2)$, optical fields E, and polarization functions P obtained as a result of the numerical integration of the TW model. For the construction of the 40MA and 50MA systems, we have used the modes indicated by full black and a bit smaller gray bullets in Figures 31.11a and b. Solid gray and black dashed curves in Figure 31.11c represent the numerical integration of these MA systems. One can see that whereas 40MA system fails to reproduce the stable periodic regime, the 50MA system provides a perfect approximation of the TW model. We note that a significant number of excited optical modes in the example considered above does not allow achieving a low dimensional approximation of the TW model. The number of active modes usually is much smaller in MSLs containing one or more DFB sections. In this case, already three or four appropriately selected optical modes are sufficient for a good approximation of the TW model [7,28].

An integration of the MA system (Equation 31.39) remains a nontrivial task, because for each actual set of carrier densities $N = (N_1, \dots, N_r)$, one should find the corresponding mode frequencies $\Omega_k(N)$, $k = 1, \dots, q$ by solving the characteristic Equation 31.34 numerically. Since the required computation time of the MA systems grows quadratically with the increasing number of modes, one can integrate the TW model faster than the 50MA system. The usefulness of the MA approach starts to be visible when combining our model reduction technique with the numerical continuation and bifurcation analysis tools [29] suited for investigation of nearly arbitrary systems of ODEs. Figure 31.11d presents an example of numerical bifurcation analysis of 3MA (gray) and 4MA (black) systems describing dynamics of the three-section laser consisting of two active DFB sections and a passive phase tuning section in between ($n = 3$ and $n_a = 2$ in this case). Here, solid and dashed curves represent stable and unstable branches of the periodic orbit implied by beating of two closely located resonances, supported by each DFB section. Empty bullets in the same figure represent the continuation of the stable periodic state by direct integration of the TW model. By comparing the bullets and curves, one can see that both MA systems were able to reproduce the stable branch of the periodic orbit, and identify torus and saddle-node bifurcations where this state have lost its stability. The deviation of the solid gray curve from the bullet positions in Figure 31.11d, however, indicates the insufficiency of the 3MA system to reproduce the orbit shape. More detailed analysis of this laser including a continuation of bifurcations in two parameter domain can be found in Reference [28].

31.11 Conclusions

In this chapter, we introduce a hierarchy of TW models describing nonlinear dynamics in individual semiconductor lasers, various MSLs, and coupled laser systems. To simulate these laser devices, we use our software package LDSL-tool, which treats MSLs as a set of differently interconnected laser sections. At the end of the chapter, we introduce several advanced techniques allowing detailed analysis of the model equations. These methods include computation of optical modes, a study of the mode spectra, expansion of electric fields into modal components, a semianalytic location of all steady states of the MSLs, model reduction, numerical continuation, and bifurcation analysis of the reduced system. Altogether, these advanced possibilities of our software tool allow to achieve a thorough understanding of the processes observed both, in the direct integration of model equations and experiments.

References

1. M.W. Fleming and A. Mooradian. Spectral characteristics of external-cavity controlled semiconductor lasers. *IEEE Journal of Quantum Electronics*, 17:44–59, 1981.
2. A. Argyris, D. Syvridis, L. Larger, V. Annovazzi-Lodi, P. Colet, I. Fischer, J. García-Ojalvo, C.R. Mirasso, L. Pesquera, and K.A. Shore. Chaos-based communications at high bit rates using commercial fibre-optic links. *Nature*, 438:343–346, 2005.

3. A. Uchida, K. Amano, M. Inoue, K. Hirano, S. Naito, H. Someya, I. Oowada, T. Kurashige, M. Shiki, S. Yoshimori, K. Yoshimura, and P. Davis. Fast physical random bit generation with chaotic semiconductor lasers. *Nature Photonics*, 2(12):728–732, 2008.
4. E. Murphy. The semiconductor laser: Enabling optical communication. *Nature Photonics*, 4(5):287, 2010.
5. H.-J. Wünsche, O. Brox, M. Radziunas, and F. Henneberger. Excitability of a semiconductor laser by a two-mode homoclinic bifurcation. *Physical Review Letters*, 88(2):023901, 2002.
6. O. Brox, S. Bauer, M. Radziunas, M. Wolfrum, J. Sieber, J. Kreissl, B. Sartorius, and H.-J. Wünsche. High-frequency pulsations in DFB-lasers with amplified feedback. *IEEE Journal of Quantum Electronics*, 39(11):1381–1387, 2003.
7. M. Radziunas, A. Glitzky, U. Bandelow, M. Wolfrum, U. Troppenz, J. Kreissl, and W. Rehbein. Improving the modulation bandwidth in semiconductor lasers by passive feedback. *IEEE Journal of Selected Topics in Quantum Electronics*, 13(1):136–142, 2007.
8. U. Bandelow, M. Radziunas, A. Vladimirov, B. Hüttl, and R. Kaiser. Harmonic mode-locking in monolithic semiconductor lasers: Theory, simulations and experiment. *Optical and Quantum Electronics*, 38:495–512, 2006.
9. R. Arkhipov, A. Pimenov, M. Radziunas, D. Rachinskii, A.G. Vladimirov, D. Arsenijevic, H. Schmeckebier, and D. Bimberg. Hybrid mode locking in semiconductor lasers: Simulations, analysis and experiments. *IEEE Journal of Selected Topics in Quantum Electronics*, 19(4):1100208, 2013.
10. M. Radziunas, A.G. Vladimirov, E.A. Viktorov, G. Fiol, H. Schmeckebier, and D. Bimberg. Pulse broadening in quantum-dot mode-locked semiconductor lasers: Simulation, analysis and experiments. *IEEE Journal of Quantum Electronics*, 47(7):935–943, 2011.
11. M. Möhrle, B. Sartorius, C. Bornholdt, S. Bauer, O. Brox, A. Sigmund, R. Steingrüber, M. Radziunas, and H.-J. Wünsche. Detuned grating multisection-RW-DFB lasers for high-speed optical signal processing. *IEEE Journal of Selected Topics in Quantum Electronics*, 7(2):217–223, 2001.
12. H.-J. Wünsche, M. Radziunas, S. Bauer, O. Brox, and B. Sartorius. Simulation of phase-controlled mode-beating lasers. *IEEE Journal of Selected Topics in Quantum Electronics*, 9(3):857–864, 2003.
13. S. Bauer, O. Brox, J. Kreissl, B. Sartorius, M. Radziunas, J. Sieber, H.-J. Wünsche, and F. Henneberger. Nonlinear dynamics of semiconductor lasers with active optical feedback. *Physical Review E*, 69:016206, 2004.
14. M. Radziunas, K.-H. Hasler, B. Sumpf, T. Quoc Tien, and H. Wenzel. Mode transitions in DBR semiconductor lasers: Experiments, simulations and analysis. *Journal of Physics B: Atomic, Molecular and Optical Physics*, 44:105401, 2011.
15. M. Radziunas, V.Z. Tronciu, E. Luvsandamdin, Ch. Kürbis, A. Wicht, and H. Wenzel. Study of micro-integrated external-cavity diode lasers: Simulations, analysis and experiments. *IEEE Journal of Quantum Electronics*, 51(2):2000408, 2015.
16. O. Hess and T. Kuhn. Spatio-temporal dynamics of semiconductor lasers: Theory, modelling and analysis. *Progress in Quantum Electronics*, 20(2):85–179, 1996.
17. E. Gehrig, O. Hess, and R. Walenstein. Modeling of the performance of high power diode amplifier systems with an optothermal microscopic spatio-temporal theory. *IEEE Journal of Quantum Electronics*, 35:320–331, 2004.
18. R. Lang and K. Kobayashi. External optical feedback effects on semiconductor injection laser properties. *IEEE Journal of Quantum Electronics*, 16:347–355, 1980.
19. M. Sorel, G. Giuliani, A. Scire, R. Miglierina, J.P.R. Laybourn, and S. Donati. Operating regimes of GaAs-AlGaAs semiconductor ring lasers: Experiment and model. *IEEE Journal of Quantum Electronics*, 39:1187–1195, 2003.
20. A.G. Vladimirov and D. Turaev. Model for passive mode-locking in semiconductor lasers. *Physical Review A*, 72:033808, 2005.

21. U. Bandelow, M. Radziunas, J. Sieber, and M. Wolfrum. Impact of gain dispersion on the spatio-temporal dynamics of multisection lasers. *IEEE Journal of Quantum Electronics*, 37:183–188, 2001.
22. J. Javaloyes and S. Balle. Emission directionality of semiconductor ring lasers: A traveling-wave description. *IEEE Journal of Quantum Electronics*, 45:431–438, 2009.
23. M. Radziunas. Longitudinal modes of multisection edge-emitting and ring semiconductor lasers. *Optical and Quantum Electronics*, 47(6):1319–1325, 2015.
24. U. Bandelow, R. Hünlich, and T. Koprucki. Simulation of static and dynamic proper-ties of edge-emitting multiple-quantum-well lasers. *IEEE Journal of Selected Topics in Quantum Electronics*, 9(3):798–806, 2003.
25. LDSL-tool: A software package for simulation and analysis of longitudinal dynamics of multisection semiconductor lasers. Available at www.wias-berlin.de/software/ldsl.
26. M. Radziunas and H.-J. Wünsche. Multisection lasers: Longitudinal modes and their dynamics. In J. Piprek, editor, *Optoelectronic Devices—Advanced Simulation and Analysis*, chapter 5, pp. 121–150. New York, NY: Springer, 2005.
27. M. Radziunas, H.-J. Wünsche, B. Krauskopf, and M. Wolfrum. External cavity modes in Lang-Kobayashi and traveling wave models. In *SPIE Proceeding Series*, 6184:61840X, 2006.
28. M. Radziunas. Numerical bifurcation analysis of the traveling wave model of multisection semiconductor lasers. *Physica D*, 213:98–112, 2006.
29. E.J. Doedel, A.R. Champneys, T.F. Fairgrieve, Y.A. Kuznetsov, B. Sandstede, and X. Wang. AUTO97: Continuation and bifurcation software for ordinary differential equations (with HomCont). Technical Report TW-330, Department of Computer Science, K.U. Leuven, Leuven, Belgium, 2001.
30. S. Bischoff, J. Mørk, T. Franck, S.D. Brorson, M. Hofmann, K. Frojdh, and L. Prip. Monolithic colliding pulse mode-locked semiconductor lasers. *Quantum and Semiclassical Optics*, 9(5):655–674, 1997.
31. M. Kolesik and J. Moloney. A spatial digital filter method for broad-band simulation of semiconductor lasers. *IEEE Journal of Quantum Electronics*, 37(7):936–944, 2001.
32. A.J. Lowery. New dynamic semiconductor laser model based on the transmission-line modelling method. *IEE Proceedings J – Optoelectronics*, 134(5):281–289, 1987.
33. D.J. Jones, L. Zhang, J. Carroll, and D. Marcenac. Dynamics of monolithic passively mode-locked semiconductor lasers. *IEEE Journal of Quantum Electronics*, 31(6):1051–1058, 1995.
34. E.A. Avrutin, J.H. Marsh, and E.L. Portnoi. Monolitic and multi-gigaHertz mode-locked semiconductor lasers: Constructions, experiments, models and applications. *IEE Proceedings – Optoelectronics*, 147(4):251–278, 2000.
35. C.Z. Ning, R.A. Indik, and J.V. Moloney. Effective Bloch equations for semiconductor lasers and amplifiers. *IEEE Journal of Quantum Electronics*, 33(9):1543–1550, 1997.
36. M. Homar, S. Balle, and M. San Miguel. Mode competition in a Fabry-Perot semi-conductor laser: Travelling wave model with asymmetric dynamical gain. *Optics Communications*, 131:380–390, 1996.
37. A. Perez-Serrano, J. Javaloyes, and S. Balle. Longitudinal mode multistability in ring and Fabry-Perot lasers: The effect of spatial hole burning. *Optics Express*, 19:3284–3289, 2011.
38. M. Spreemann, M. Lichtner, M. Radziunas, U. Bandelow, and H. Wenzel. Measurement and simulation of distributed-feedback tapered master-oscillators power-amplifiers. *IEEE Journal of Quantum Electronics*, 45(6):609–616, 2009.
39. B. Grote, E.K. Heller, R. Scarmozzino, J. Hader, J.V. Moloney, and S.W. Koch. Fabry-Perot lasers: Temperature and many-body effects. In J. Piprek, editor, *Optoelectronic Devices—Advanced Simulation and Analysis*, pp. 27–61. New York, NY: Springer, 2005.
40. H.-J. Wünsche, U. Bandelow, and H. Wenzel. Calculation of combined lateral and longitudinal spatial hole burning in $\lambda/4$ shifted DFB lasers. *IEEE Journal of Quantum Electronics*, 29(6):1751–1760, 1993.

41. U. Bandelow, H. Wenzel, and H.-J. Wünsche. Influence of inhomogeneous injection on side-mode suppression in strongly coupled DFB semiconductor lasers. *Electronics Letters*, 28:1324–1325, 1992.
42. P.G. Eliseev, A.G. Glebov, and M. Osinski. Current self-distribution effect in diode lasers: Analytic criterion and numerical study. *IEEE Journal of Selected Topics in Quantum Electronics*, 3:499–506, 1997.
43. R. Schatz. Longitudinal spatial instability in symmetric semiconductor lasers due to spatial hole burning. *IEEE Journal of Quantum Electronics*, 28(6):1443–1449, 1992.
44. V. Tronciu, S. Schwertfeger, M. Radziunas, A. Klehr, U. Bandelow, and H. Wenzel. Amplifications of picosecond laser pulses in tapered semiconductor amplifiers: Numerical simulations versus experiments. *Optics Communications*, 285:2897–2904, 2012.
45. A. Perez-Serrano, J. Javaloyes, and S. Balle. Bichromatic emission and multimode dynamics in bidirectional ring lasers. *Physical Review A*, 81:043817, 2010.
46. E.J. D'Angelo, E. Izaguirre, G.B. Mindlin, L. Gil, and J.R. Tredicce. Spatiotemporal dynamics of lasers in the presence of an imperfect O(2) symmetry. *Physical Review Letters*, 68:3702–3704, 1992.
47. Y. Barbarin, E.A.J.M. Bente, M.J.R. Heck, Y.S. Oei, R. Nötzel, and M.K. Smit. Characterization of a 15 GHz integrated bulk InGaAsP passively modelocked ring laser at 1.53 μm. *Optics Express*, 14(21):9716–9727, 2006.
48. E.A. Viktorov, T. Erneux, P. Mandel, T. Piwonski, G. Madden, J. Pulka, G. Huyet, and J. Houlihan. Recovery time scales in a reversed-biased quantum dot absorber. *Applied Physics Letters*, 94:263502, 2009.
49. M. Radziunas, A.G. Vladimirov, E.A. Viktorov, G. Fiol, H. Schmeckebier, and D. Bimberg. Strong pulse asymmetry in quantum-dot mode-locked semiconductor lasers. *Applied Physics Letters*, 98:031104, 2011.
50. K. Lüdge and E. Schöll. Quantum-dot lasers—desynchronized nonlinear dynamics of electrons and holes. *IEEE Journal of Quantum Electronics*, 45(11):1396–1403, 2009.
51. A. Markus, M. Rossetti, V. Calligari, D. Chek-Al-Kar, J.X. Chen, A. Fiore, and R. Scollo. Two-state switching and dynamics in quantum dot two-section lasers. *Journal of Applied Physics*, 100:113104, 2006.
52. M. Rossetti, P. Bardella, and I. Montrosset. Time-domain travelling-wave model for quantum dot passively mode-locked lasers. *IEEE Journal of Quantum Electronics*, 47(2):139–150, 2011.
53. V.Z. Tronciu, M. Radziunas, Ch. Kürbis, H. Wenzel, and A. Wicht. Numerical and experimental investigations of micro-integrated external cavity diode lasers. *Optical and Quantum Electronics*, 47(6):1459–1464, 2015.
54. T. Krauss and P.J.R. Laybourn. Very low threshold current operation of semiconductor ring lasers. *IEE Proceedings J – Optoelectronics*, 139(6):383–388, 1992.
55. I.V. Ermakov, S. Beri, M. Ashour, J. Danckaert, B. Docter, J. Bolk, X.J.M. Leijtens, and G. Verschaffelt. Semiconductor ring laser with on-chip filtered optical feedback for discrete wavelength tuning. *IEEE Journal of Quantum Electronics*, 48(2):129–136, 2012.
56. M. Radziunas, V.Z. Tronciu, U. Bandelow, M. Lichtner, M. Spreemann, and H. Wenzel. Mode transitions in distributed-feedback tapered master-oscillator power-amplifier: Theory and experiments. *Optical and Quantum Electronics*, 40:1103–1109, 2008.

32

Mode-Locked Semiconductor Lasers

32.1 General Principles of Mode Locking, the Important Features of Mode-Locked Semiconductor Lasers, and the Role of Theory and Modeling 183
32.2 ML Techniques in Laser Diodes: The Main Features 185
32.3 Theoretical Models of ML in Semiconductor Lasers 187
Small-Signal Time-Domain Models and Self-Consistent Pulse Profile • Frequency and Time-Frequency Treatment of ML and Dynamic Modal Analysis • Large-Signal Time-Domain Approach and Delay-Differential Equation Model • TWMs: The General Considerations
32.4 The Main Predictions of Mode-Locked Laser Theory 210
Operating Regime Depending on the Operating Point • The Main Parameters That Affect Mode-Locked Laser Behavior
32.5 Microscopic and Semimicroscopic Approaches in Mode-Locked Laser Modeling 217
The Basics of Microscopic Input in Mode-Locked Laser Simulations • Improving the Numerical Efficiency of TWMs: Decimation/Space–Time Folding
32.6 Some Novel Problems and Challenges in Mode-Locked Laser Modeling 225
Coherent Population Effects as a Possible Saturable Absorption Mechanism • Spontaneous ML in Single-Section Lasers
32.7 Concluding Remarks 227

Eugene Avrutin
and
Julien Javaloyes

32.1 General Principles of Mode Locking, the Important Features of Mode-Locked Semiconductor Lasers, and the Role of Theory and Modeling

In the most general sense, *mode locking* (ML) is a regime of laser operation that involves emitting light in several modes with a time-independent relation between them, i.e., with constant and precisely equidistant frequencies. Usually, the term is used more specifically, referring to what is, rigorously speaking, *amplitude-modulation* (AM) ML, meaning that the phase differences between of adjacent modes are approximately equal. In time domain, this corresponds to the laser's emitting a train of ultrashort optical pulses at a repetition frequency F near the cavity round-trip frequency or its harmonic:

$$F \approx M_{\mathrm{h}} v_{\mathrm{g}}/(2L). \tag{32.1}$$

Here, v_g denotes the group velocity of light in the laser resonator and L is the Fabry–Pérot resonator cavity length. In the case of the ring resonator, $2L$ in Equation 32.1 is substituted by the ring cavity length. The harmonic number M_h corresponds to the number of pulses coexisting in the cavity; in the simplest and most usual case, $M_h = 1$. The pulse duration is then of the order of $2L/(N_M v_g)$, with N_M being the number of lasing modes in the spectrum.

In most cases (some important exceptions will be mentioned below), ML does not occur spontaneously and requires a special laser construction and/or operating conditions. Namely, it is usually achieved either by modulation of the laser net gain at a frequency F or its (sub)harmonic (known as *active* ML) or by exploiting nonlinear properties of the medium to shorten the propagating pulse (known as *passive* ML, PML); in both cases, the pulse shortening mechanism needs to be strong enough to counter the broadening effects of gain saturation and dispersion effectively. PML, in turn, is usually achieved by introducing a saturable absorber (SA) into the laser cavity. The SA both facilitates a self-starting mechanism for ML and, most importantly, plays a crucial role in shortening the duration of the circulating pulses. As a variation of this principle, *refractive index nonlinearities* approximately equivalent in their action to saturable absorption have been intensely studied in the last decades; salient examples are *additive pulse ML* and *Kerr lens ML* in solid-state lasers, see, e.g., Haus (2000) for more detail.

A combination of active and passive methods of ML is known as *hybrid* ML; if the external modulation is in the form of short pulses, the corresponding regime is referred to as *synchronous* ML.

Mode-locked solid-state lasers, often diode pumped, have allowed sub-100 femtosecond pulses to be generated (Brown et al., 2004; Ell et al., 2001; Innerhofer et al., 2003), with peak powers in the range of many kilowatts (partly due to the relatively low repetition rates, typically in the megahertz range or below).

In *semiconductor diode lasers*, the most basic physical mechanisms underlying the generation of short pulses are fundamentally similar to those of other types of lasers, but a number of features are very different, as regards both technology and physics.

From the practical and technological point of view, diode lasers have a number of advantages: They represent the most compact and efficient sources of picosecond and subpicosecond pulses. They are directly electrically pumped, and the bias current can be easily adjusted to determine the pulse duration and the optical power, thus offering, to some extent, electrical control of the characteristics of the output pulses. These lasers also offer the best option for the generation of high-repetition rate trains of pulses, owing to their small cavity size L in Equation 32.1 and hence the large values of F, well into multigigahertz range. Being much cheaper to fabricate and operate than most other types of lasers, ultrafast semiconductor lasers also offer the potential for dramatic cost savings in a number of applications that traditionally use solid-state lasers. The deployment of high-performance ultrafast diode lasers could therefore have a significant economic impact by enabling ultrafast applications to become more profitable and even facilitate the emergence of new applications. At the moment, actual and potential applications of mode-locked lasers include time- and wavelength-multiplexed communications, metrology, biomedical applications, etc.; see Avrutin and Rafailov (2012) for an overview.

From the point of view of physics, which underlies the technology, most of the distinct features of semiconductor lasers, including mode-locked ones, lie in the energy spectrum of semiconductors that consists in continuous bands of energy with relatively high density of states, as opposed to discrete levels in solid-state lasers. Most mode-locked semiconductor lasers operate on fundamental interband transitions (though there has been some work on active ML of intersubband quantum cascade lasers; see, e.g., Revin et al., 2016 and references therein). Semiconductors thus have both a higher gain per unit length (which is one of the reasons of the short cavity length being possible) and a higher nonlinear refractive index than other gain media (the relation between carrier density dependences of gain and refractive index in semiconductor lasers is often quantified via Henry's linewidth enhancement factor α_H that, as discussed below, plays an important role in the theory of ML). The interaction of the pulse with the gain and saturable absorption and the resulting large changes in the nonlinear refractive index lead to significant self-phase modulation, imparting a noticeable *chirp* to the ML pulses, usually up-chirp in the case of passively mode-locked lasers and down-chirp in actively mode-locked ones. This increases the *time-bandwidth product* of the pulse and has been one of the important limitations in obtaining pulse durations of the order of 100 fs *directly* from the diode lasers, with picosecond pulses being the norm. Furthermore, a strong saturation of the gain also results in stabilization of the pulse energy, which limits the average and peak power to substantially lower levels than in vibronic/solid-state lasers. Output average power levels for mode-locked laser diodes are usually between 0.1 and 100 mW, while peak power levels remain between 10 mW and 1 W. Only with additional amplification/compression setups, can the peak power reach the kW level (Kim et al., 2005). Another distinction of semiconductor lasers is that the typical scales of carrier recombination times in semiconductor materials are of the order of hundreds of picoseconds, comparable to the ML repetition time, leading to a rich variety of dynamic instabilities in the laser behavior, some of which, as will be discussed later, determine the *lower* limit for stable ML frequency (and therefore the longest possible laser cavity) for given material parameters.

There has been a large variety of semiconductor laser designs used for ML, from external cavity ones operating at sub-GHz rates to monolithic ones reaching terahertz repetition frequencies (Avrutin et al., 2000), with laser design strongly affecting, not just the laser performance, but the relative importance of the underlying physical effects in determining this performance.

This combination of practical promise, versatility, and scientific challenge has made mode-locked semiconductor lasers an important topic of research for more than two decades; they have arguably attracted considerably more attention than all other methods of ultrashort pulse generation taken together. The most recent years have seen considerable progress in both improving the theoretical understanding of ML in semiconductor lasers and using this understanding to improve their performance in terms of power, pulse duration/chirp, stability, repetition rates accessible, and integrability issues.

This progress has been partly summarized in previous reviews on ML in semiconductor lasers, with some of them (Avrutin et al., 2005) explicitly concentrating on modeling and simulation and others (Avrutin et al., 2000; Avrutin and Rafailov, 2012) paying significant attention to it. Here, we shall partly follow the logic and layout of the previous paper (Avrutin and Rafailov, 2012), but will attempt to present a more modern perspective and cover the recent results by ourselves and other researchers.

32.2 ML Techniques in Laser Diodes: The Main Features

The main advantages of ML over other methods of generating ultrafast pulses by laser diodes are the higher repetition rate pulses and shorter pulse durations. To realize these advantages to the fullest, a variety of ML techniques and device structures have been investigated and optimized (Vasil'ev, 1995). All three main forms of ML—active, passive, and hybrid—have been extensively studied for semiconductor lasers.

Purely *active* ML in a semiconductor laser can be achieved by direct modulation of the gain section current with a frequency very close to the pulse repetition frequency in the cavity or to a subharmonic of this frequency. Alternatively, an electroabsorption segment of a multielement device can be modulated to produce the same effect, or a separate modulation section introduced. The main advantages of active ML techniques are the resultant low jitter (essentially determined by the electrical generator imposing the modulation) and the ability to synchronize the laser output with the modulating electrical signal, which is a fundamental attribute for optical transmission and signal processing applications. However, high repetition frequencies are not readily obtained through directly driven modulation of lasers because fast microwave modulation, particularly of current, becomes progressively more difficult with increase in frequency.

PML of semiconductor lasers typically utilizes an SA region in the laser diode. Upon start-up of laser emission, the laser modes initially oscillate with relative phases that are random; in other words, the temporal radiation pattern consists of irregular bursts. If one of these bursts is energetic enough to provide energy, or more accurately fluence (energy per unit area) of the order of the saturation fluence of the absorber, it will partly bleach the absorption. This means that around the peak of the burst where the intensity is higher, the loss will be smaller, while the low-intensity wings become more attenuated. The pulse generation process is thus initiated by this family of intensity spikes that experience lower losses within the absorber carrier lifetime. The dynamics of absorption and gain play a crucial role in pulse shaping. In steady state, the unsaturated losses are higher than the gain. When the leading edge of the pulse reaches the absorber, the loss saturates more quickly than the gain, which results in a net gain window, as depicted in Figure 32.1. The absorber then recovers from this state of saturation to the initial state of high loss, thus attenuating the trailing edge of the pulse. It is thus easy to understand why the saturation fluence and the recovery time of the absorber are of primary importance in the formation of mode-locked pulses.

In practical terms, the SA can be monolithically integrated into a semiconductor laser by electrically isolating one section of the device (Figure 32.2a). By applying a reverse bias to this section, the carriers photogenerated by the pulses can be more efficiently swept out of the absorber, thus enabling the SA to recover more quickly to its initial state of high loss. An increase in the reverse bias serves to decrease the

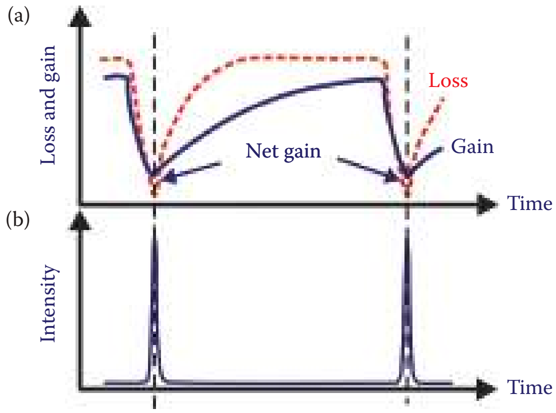


FIGURE 32.1 Schematic illustration of the mechanism of passive mode locking: (a) the loss and gain dynamics that lead to (b) pulse generation.

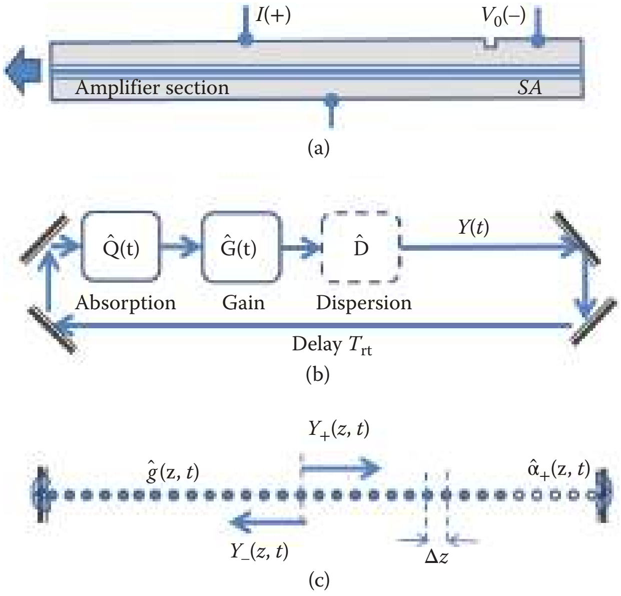


FIGURE 32.2 Schematic of (a) the simplest design of an edge-emitting passively mode locked laser, and its representation in (b) the lumped model of Sections 32.3.1 and 32.3.3, and (c) the traveling-wave model (TWM) of Section 32.3.4.

absorber recovery time, which will have the effect of further shortening the pulses (a faster absorber may act to shorten the trailing as well as the leading edge of the pulse). Alternatively, an SA can also be implemented through ion implantation on one of the facets of the laser, thus increasing the nonradiative recombination (Delpon et al., 1998).

PML provides the shortest pulses achievable by all three techniques, albeit at the expense of somewhat larger pulse jitter and radio frequency (RF) linewidth than in active or hybrid ML. It can be intuitively understood in the following way: Active ML generates a fixed AM of the gain while PML induces a modulation that is proportional to the pulse energy, meaning potentially more efficient mode coupling. Besides, the absence of an RF source simplifies the fabrication and operation considerably. PML also allows for higher pulse repetition rates than those determined solely by the cavity length, by means of harmonic ML ($M_h > 1$ in Equation 32.1; some means of achieving this are considered in more detail below).

Hybrid ML can be achieved by applying RF modulation either to gain or to the SA section. It has been shown, however, that the more efficient method is the latter one, in which case the SA doubles as an electroabsorption modulator. In this case, the pulse generation may be seen as initiated by a modulation provided by the RF signal, while further shaping and shortening are assisted by the SA. This process results in high-quality pulses, synchronized with an external source.

32.3 Theoretical Models of ML in Semiconductor Lasers

Any model of mode-locked laser dynamics should account for pulse shortening by modulation (active/hybrid ML) and/or saturable absorption (passive/hybrid ML) and for pulse broadening by saturable gain and cavity dispersion (including gain/loss dispersion and group velocity/phase dispersion), as discussed above. In addition, if spectral properties are to be accounted for accurately, self-phase modulation needs to be included in the model. In this section, we shall cover general principles of the possible approaches, concentrating on the relatively recent advances in ML theory that have underpinned the significant progress in understanding the details of ML dynamics.

32.3.1 Small-Signal Time-Domain Models and Self-Consistent Pulse Profile

Conceptually the simplest, and historically the oldest, models of mode-locked lasers are *time-domain lumped models* (Figure 32.2b), based on the approximation that the pulsewidth is much smaller than the repetition period, and treating a hypothetic ring laser with unidirectional propagation. The amplification and gain/group velocity dispersion (GVD), which in reality are experienced by the pulse simultaneously, may then be approximately treated in two independent stages. This allows the representation of the distributed amplifier in the model by a lumped *gain element* performing the functions of amplification and self-phase modulation. Mathematically, this element can be described by a nonlinear integral or integro-differential operator acting on the complex pulse shape function (complex slow amplitude) $Y(t)$, t being the *local* time of the pulse. The model was originally designed for solid-state and gas lasers, whose long lengths make for a round-trip time many orders of magnitude longer than the pulse duration, so separate timescales are introduced explicitly for the pulse (the short timescale) and relaxation period between pulses (the long timescale).

The gain operator takes the form

$$\hat{G}Y(t) = \exp\left(\frac{1}{2}(1 - i\alpha_{Hg})G(t)\right) Y(t), \tag{32.2}$$

with α_{Hg} the Henry linewidth enhancement factor in the amplifier and $G(t) = \Gamma \int g(z - v_g t, t)dz$ the total gain integrated over the length of the amplifying region (v_g being the group velocity of light). Further analytical progress can be made by using spatially resolved rate equations for the carrier density N_g in the gain region $\frac{\mathrm{d}N_{\mathrm{g}}(z,t)}{\mathrm{d}t} = -g(N_{\mathrm{g}})\frac{P(z,t)}{\hbar\omega A_x} - \frac{N_{\mathrm{g}}}{\tau_{\mathrm{g}}}$, where P is the power of light, $\hbar\omega$ the photon energy, A_x the cross section of the optical beam (mode) in the gain section, and τ_{g} the gain recovery time. A number of approximations are then made. Those involve neglecting dispersion and fast nonlinearities, assuming (which is a safe assumption in semiconductor active media) that $\tau_{\mathrm{g}} \gg \tau_{\mathrm{p}}(\tau_{\mathrm{p}}$ being the pulse duration), so gain relaxation during the pulse can be neglected, and assuming a linear dependence of gain on the carrier density: $\mathrm{g}(N_{\mathrm{g}}) \approx \sigma_{\mathrm{g}}(N_{\mathrm{g}} - N_{\mathrm{tr}}), \sigma_{\mathrm{g}} = \mathrm{dg}/\mathrm{d}N$ being the gain cross section (the derivative of the gain on the carrier density) and N_{tr} the transparency carrier density. With these assumptions, an approximate explicit expression for $G(t)$ on the short timescale commensurate with the pulse duration can be obtained. In the case of $G \ll 1$, it takes the form

$$G(t) = G_{-} \exp(-U(t)/U_{\mathrm{g}}). \tag{32.3}$$

Here, $U(t) = \int_{-\infty}^{t} P(t')\mathrm{d}t' = v_{\mathrm{g}}\hbar\omega A_X \int_{-\infty}^{t} |Y(t')^2\mathrm{d}t'$ is the pulse energy up to the time t, $G_{-} = \Gamma \int g(z, t \to -\infty)\mathrm{d}z$ is the total amplification in the gain element at the time before the arrival of the pulse, and

$$U_{\mathrm{g}} = \frac{\hbar\omega A_X}{\sigma_{\mathrm{g}}} \tag{32.4}$$

is the saturation energy of the amplifier.

The SA, if any, is also considered as a lumped element, described, in the simplest case, by an operator similar to Equation 32.3:

$$\hat{\mathrm{Q}}_{\mathrm{S}} Y(t) = \exp\left(-\frac{1}{2}(1 - i\alpha_{\mathrm{H}\alpha})Q_{\mathrm{S}}(t)\right) Y(t), \tag{32.5}$$

with the absorber linewidth enhancement factor $\alpha_{\mathrm{H}\alpha}$, and the dimensionless slow saturable absorption $Q_{\mathrm{S}} = \Gamma \int \alpha(z - v_{\mathrm{g}}t, t)\mathrm{d}z$, in the ideal slow absorber approximation and with $Q \ll 1$, given by

$$Q_{\mathrm{S}}(t) = Q_{-} \exp(-U(t)/U_{\alpha}). \tag{32.6}$$

Here, again, the total initial absorption $Q_{-} = \Gamma \int \alpha(z, t \to -\infty)\mathrm{d}z$, and the absorber saturation energy is

$$U_{\alpha} = \frac{\hbar\omega A_{X\alpha}}{\sigma_{\alpha}}, \tag{32.7}$$

where $\sigma_{\alpha} = |\mathrm{d}\alpha/\mathrm{d}N|$ is the SA cross section, and, depending on the construction, the cross section of the beam $A_{X\alpha}$ in the SA may be different from that of the amplifying section; as shown below, it is usually advantageous to have $A_{X\alpha} < A_X$.

Equation 32.6 is obtained in the same way as Equation 32.3 by using the spatially resolved rate equation for the carrier density in the absorber region $\frac{\mathrm{d}N_{\alpha}(z,t)}{\mathrm{d}t} = \alpha(N_{\alpha})\frac{P(z,t)}{\hbar\omega A_{X\alpha}} - \frac{N_{\alpha}}{\tau_{\alpha}}$, with τ_{α} the absorber relaxation time and, and making the same assumptions: assuming a linear dependence of the absorption on the SA carrier density $\alpha(N_{\alpha}) \approx \alpha_0 - \sigma_{\alpha}N_{\alpha}$, with α_0 the unsaturated absorption, and neglecting the SA relaxation during the pulse. The latter is known in the theory of ML as *the ideal slow absorber* approximation, meaning that the SA recovery time τ_{α}, like that of gain τ_{g}, needs to be much longer than the pulse duration τ_{p}. However, while the assumption $\tau_{\mathrm{g}} \gg \tau_{\mathrm{p}}$ is readily fulfilled in semiconductor lasers ($\tau_{\mathrm{g}} \sim 1$ ns), the assumption $\tau_{\alpha} \gg \tau_{\mathrm{p}}$ may be strained ($\tau_{\alpha} \sim 10$ ps), which may necessitate some modifications to the model, described below, to improve its accuracy.

Absorption in semiconductor SAs tends, in addition, to contain a subpicosecond component that acts as a "fast" SA (recovery time $\ll \tau_p$) even for short pulses ($\tau_p \sim 1$ ps), typically generated by semiconductor lasers. Some lumped time-domain models (Haus and Silberberg, 1985) also include "fast" effects in the SA (SA nonlinearities) as an equivalent fast absorber characterized by an operator $\hat{Q}_F Y(t) = \exp\left(-\frac{1}{2}Q_F(t)\right) Y(t)$ and with an equivalent absorption:

$$Q_F(t) = Q_{iF}(1 - \varepsilon_\alpha |Y(t)|^2). \tag{32.8}$$

Then, the total absorption is

$$\hat{Q}Y(t) = \hat{Q}_S\hat{Q}_F Y(t). \tag{32.9}$$

Gain nonlinearities may, in principle, be included in the same way, although a more accurate account of dynamics may be preferable, particularly in the case of quantum dot (QD) materials.

Finally, in the traditional form of a lumped model, the dispersion of material gain and refractive index, together with any artificial dispersive elements present in the cavity, such as a distributed Bragg reflector (DBR), are combined in a lumped *dispersive element*. In the frequency domain, its effect on the pulse may be written as

$$\hat{D}Y^T(\omega) = e^{i\varphi_0}\left[\frac{1}{1 - i(\omega - \omega_p)/\gamma} + D(\omega - \omega_0)^2\right] Y^T, \tag{32.10}$$

where Y^T is the Fourier transform of the complex pulse shape $Y(t)$, ω_p and $\gamma \ll \omega_p$ are the peak frequency and the bandwidth of the dispersive element (defined by the gain curve of the amplifier and the frequency selectivity of a grating element, if it is present in the cavity), and ω_0 is the reference frequency as in the analysis of amplifiers. The value of ω_p may change during the pulse (due to gain curve variation with carrier density, most importantly the gain peak shift); this modifies the dispersive operator (Leegwater, 1996), although in the majority of papers on the subject, the effect is not included. ϕ_0 denotes the phase shift introduced by the element and D is the equivalent dispersion (including the GVD of the passive waveguide and the effective dispersion of the external grating element, if any). To rewrite the operator (Equation 32.10) in the time domain, one may expand the first term around the reference frequency ω_0 noting that $|\omega_p - \omega_0| \ll \omega_0$. Then, after a standard transformation, $(\omega - \omega_0)Y^T \div id/dt\ Y$ (Equation 3.7) becomes a differential operator; if the exponential is expanded keeping the first two terms, the operator is reduced to second order.

The dynamics of ML process are then described by cascading the operators and setting:

$$Y_{i+1}(t) = \left(\sqrt{\kappa}\hat{G}\hat{Q}\hat{D}\right) Y_i(t), \tag{32.11}$$

where i is the number of the pulse round-trip (determining the "slow" evolution of the ML pulse), the time t is on the fast timescale commensurate with the pulse duration, and the dimensionless parameter $\kappa < 1$ introduces the total (integrated) unsaturable intensity losses in the cavity, both distributed and due to outcoupling. The model reflects the balance of the main processes affecting the pulse in a mode-locked laser in that the saturable absorption operator $\hat{Q}$ acts to narrow the pulse down, whereas the gain saturation $\hat{G}$ and the dispersion operator $\hat{D}$ act to broaden it. The *stationary* ML equation is thus obtained by writing out the condition that the broadening and narrowing cancel each other, and the shape of the pulse is conserved from one repetition period to the next. In the operator notation introduced above, this means

$$\left(\sqrt{\kappa}\hat{G}\hat{Q}\hat{D}\right) Y(t) = e^{i\delta\psi} Y(t + \delta T), \tag{32.12}$$

where δT is the shift of the pulse or detuning between the repetition period and the round-trip of the "cold" cavity (or its fraction in case of locking at harmonics of the fundamental frequency), and $\delta\psi$ is the optical phase shift induced by the round-trip. In between the pulses, on the slow timescale commensurate with the round-trip time, gain and SA are allowed to recover with their characteristic relaxation times, according to the rate equation for carrier density with $S = 0$. This allows one to calculate the values of gain and saturable absorption at the onset of the pulse, given the pulse energy and repetition period (the only point at which this latter parameter enters a lumped time-domain model).

In the approximation of no dispersion ($\widehat{D} = 1$), the broadening of the pulse by gain saturation alone in the lumped model cannot compensate for the shortening by the absorption. The model in this approximation thus predicts the steady output in the form of a series of *infinitely short* (delta function-like) pulses; neither the pulse shape nor the duration can be analyzed in this approximation. However, it is possible to determine the total pulse *energy* and also analyze the stability of the solutions by requiring that net gain both immediately before the pulse and immediately after the pulse is smaller than one that translates into

$$\begin{aligned} G_- - Q_- - \ln\kappa &< 0 \\ G_+ - Q_+ - \ln\kappa &< 0. \end{aligned} \tag{32.13}$$

Here G_-, Q_- are the total (integrated) gain and absorption immediately before the pulse, and G_+, Q_+ are the values immediately after the pulse. This is known as New's theory of ML (strictly speaking, G. New's original 1970s paper (New, 1974) related to nonsemiconductor lasers in which $\ln(1/\kappa) \ll 1$; however, Equation 32.13 is also applicable in the generalized version of the theory proposed by Vladimirov et al. (2004) and Vladimirov and Turaev (2005) and covered in more detail below).

Analytical approximations for the *pulse shape and duration* have been originally obtained in the case of weakly nonlinear analysis, that is to say if the pulse energy is smaller than $U_{\mathrm{G,A}}$ and the gain and loss (saturable and unsaturable) during one round-trip are small ($\ln(1/\kappa), G, Q \ll 1$). Then, the exponentials in the formulas for the gain and loss operators may be expanded in Taylor series keeping terms up to the second order in Equations 32.3 and 32.6 (weak to moderate saturation of gain during the pulse):

$$\exp\left(-\frac{U(t)}{U_{\mathrm{g},\alpha}}\right) \approx 1 - \frac{U(t)}{U_{\mathrm{g},\alpha}} + \frac{1}{2}\left(\frac{U(t)}{U_{\mathrm{g},\alpha}}\right)^2, \tag{32.14}$$

and to the first order in Equations 32.2 and 32.5 (small gain and loss):

$$\exp\left(\frac{1}{2}(1 - i\alpha_{\mathrm{Hg}})G(t)\right) \approx 1 + \frac{1}{2}(1 - i\alpha_{\mathrm{Hg}})G(t); \exp\left(-\frac{1}{2}(1 - i\alpha_{\mathrm{H}\alpha})Q(t)\right) \approx 1 - \frac{1}{2}(1 - i\alpha_{\mathrm{H}\alpha})Q(t) \tag{32.15}$$

(the accuracy of the model can be improved by expanding these equations, too, to the second rather than first order).

Then, following the route pioneered by H. Haus in the first papers on ML in lasers of an arbitrary type (Haus, 1975) and later adapted specifically to diode lasers (Koumans and vanRoijen, 1996; Leegwater, 1996), the ML Equation 32.12 is rewritten as a complex second-order integro-differential equation known as the master equation of ML, which permits an analytical solution of the form

$$Y(t) = Y_0 \exp(i\Delta\omega t)\left(\cosh\frac{t}{\tau_{\mathrm{p}}}\right)^{-1+i\beta}, \tag{32.16}$$

known as the *self-consistent profile* (SCP). The corresponding theoretical approach is known as the SCP, or Haus's ML theory, as applied to semiconductor lasers (in lasers of other kind, for instance, the account for both slow and fast absorbers is typically not necessary). Assembling the terms proportional to the zeroth, first, and second power of $\tanh(t/\tau_{\mathrm{p}})$ in the ML equation, one obtains three complex, or six real, transcendental algebraic equations (Koumans and vanRoijen, 1996; Leegwater, 1996) for six real variables: pulse amplitude $|Y_0|$, duration measure τ_p, chirp parameter β, optical frequency shift $\Delta\omega = \omega - \omega_0$, repetition period detuning δT, and phase shift arg (Y_0) (which is not a measurable parameter, so in reality there are five meaningful equations). These equations, being nonlinear and transcendental, generally speaking, cannot be solved analytically, but still allow for some insight into the interrelation of pulse parameters. For example, it can be deduced (Leegwater, 1996) that the pulse duration may be considerably shortened by the presence of a fast (instantaneous) component in the saturable absorption and the achievable pulse durations are estimated about 10 times the inverse gain bandwidth, decreasing with increased pulse energy.

By requiring the net small-signal gain before and after the pulse to be negative (Equation 32.13) so that noise oscillations are not amplified, the SCP approach also allows the parameter range of the stable ML regime to be estimated.

Some conclusions from the SCP approach are borne out by more precise models (see below). In particular, it highlights the role of *the gain-to-absorber saturation energy ratio*:

$$s = \frac{U_g}{U_\alpha} = \frac{\sigma_\alpha A_{Xg}}{\sigma_g A_{X\alpha}} \tag{32.17}$$

in the ML laser performance. A minimum value of $s > 1$ is needed to achieve ML at any range of parameters at all and the range of stable ML operation broadens with an increased s. Colliding pulse mode-locked configurations, linear or ring, increase the pulse stability and also lead to shorter pulses by increasing the parameter s.

The SCP model also predicts, correctly, that increasing the dispersion parameter D also increases the parameter range for ML, at the expense of broadening the pulses, and that the slight variation of the frequency F around the estimate (Equation 32.1) shows a minimum in its dependence on current or unsaturated gain.

When applied more quantitatively, however, the SCP model is not too accurate and cannot adequately describe details of pulse shape and spectral features. Indeed, the pulse shape given by the expression (Equation 32.16) is always symmetric, which, in general, needs not, and often is not, the case in practice. Nor are the dynamic regimes of ML faithfully reproduced by the classic SCP model. The reason for this is a large number of approximations involved in the SCP approach, which have been progressively removed by various researchers at the expense of making the model more complex and, in some cases, requiring numerical rather than semianalytical analysis of the pulse profile, even if the model is still lumped.

First, achieving the SCP requires that the relaxation of gain and absorber during the pulse is negligible so that the gain and absorber operators can be written in the form of Equations 32.3 and 32.6. As mentioned above, this is a safe assumption in semiconductor lasers as regards gain media, but not necessarily the SA. The obvious upgrading to the model is then to include the dynamics of the saturable absorption Q by a characteristic recovery time τ_α. If at the same time we abandon the approximation $Q \ll 1$, then the necessary equation will take the form

$$\begin{gathered} \frac{\mathrm{d}Q(t)}{\mathrm{d}t} = -X(Q)Q\frac{P(t)}{U_\alpha} - \frac{Q_0 - Q}{\tau_\alpha}; \\ X(Q) = \frac{1 - \exp(-Q)}{Q}. \end{gathered} \tag{32.18}$$

Here, $Q_0 = \alpha_0 - L_\alpha$ (L_α being the absorption region length) is the unsaturated total absorption (at repetition periods $T_{\mathrm{rep}} \gg \tau_\alpha$, $Q_- = Q_0$), and X is the geometric factor that stems from averaging the absorption over the length of the absorber area for traveling-wave absorption (in the case of small absorption ($Q_0, Q \ll 1$) treated above, we obtain a constant $X = 1$).

Equation 32.18 is then used with Equation 32.5 instead of Equation 32.6, which it obviously reproduces in the limiting case of $X = 1$, and $\tau_\alpha \gg \tau_p$, or $\tau_\alpha \to \infty$ on the short timescale $t \sim \tau_p$. Unfortunately, even this apparently minor modification to the model means that a closed-form solution in the form of Equation 32.16 is no more possible even with $Q_0, Q \ll 1 (X = 1)$, and the iteration-type procedure (Equation 32.11) has to be repeated numerically until a steady-state profile that satisfies Equation 32.12 is found.

Studies with such a modified SCP model found that even with $X = 1$ in Equation 32.18 (small gain/absorption case) and even with absorber recovery times a few times greater than the pulse duration, the finite τ_α makes some difference to the results, noticeably shortening the pulse, making its shape less symmetric, and affecting boundaries of stable ML regime (Dubbeldam et al., 1997).

32.3.2 Frequency and Time-Frequency Treatment of ML and Dynamic Modal Analysis

An approach conceptually alternative to the time-domain analysis of ML, but to an extent sharing the small-signal nature of the model discussed above, is offered by the technique of modal analysis, static or dynamic (as in Avrutin et al., 2003 and references therein and also in Nomura et al., 2002 and Renaudier et al., 2007). In this approach, instead of analyzing the pulse shape dynamics, a modal decomposition is used and the dynamics of mode amplitudes and phases are analyzed. The advantage of the modal expansion is that the time steps can be much longer than in the spatially distributed models of ML (such as the ones described the following section). Indeed, the stiffness of the dynamics in the modal approach is governed by the temporal evolution of the modal amplitudes, and as PML consists in a steady-state regime for those, one foresees that a modal representation of PML may give rise to smooth solutions—this indeed was shown to be the case (Avrutin et al., 2003). In comparison, spatially distributed models naturally need the time step to be much shorter than the pulsewidth, as discussed in more detail in the following section. Besides, the number of variables can be smaller in the modal analysis, particularly in the case of laser designs with a spectrally selective element where only a few modes are excited, making this approach particularly efficient in analyzing, say, long-scale dynamics of external locking of DBR hybridly mode-locked lasers. It also has the logical advantage of describing steady-state ML as a steady-state solution and, conceptually, allows considering the emergence of stable PML as an order–disorder phase transition in a dissipative system, highlighting the fundamental physical features of PML in addition to its technological implications. Frequency domain analysis can be used as supplementary to time-domain models for some specific problems, as for example the analysis of harmonic operation in a coupled-cavity structure (Yanson et al., 2002), where it actually gives some analytical insight into the modal selectivity of the cavity. Frequency, or time-frequency, modal expansion-based approach to ML is also extremely useful (Kim and Lau, 1993) for analyzing the noise and linewidth properties of the ML signal, as the noise can be seen as exciting higher order *supermodes* (combination of modes) in addition to the fundamental order supermode that is actually realized in ML.

However, the modal expansion approach has a major intrinsic limitation in that it relies on the inherent assumption of weak to modest nonlinearity and modulation, meaning that the results obtained using this method agree reasonably well with time-domain simulations only for the case of high ML frequencies, typically above 100 GHz (short or harmonic cavities) and/or at relatively small currents above threshold. Large-signal instabilities, such as the chaotic leading edge instability, are not predicted accurately, and the accuracy of the frequency-domain models at high amplitudes cannot be guaranteed. In addition, finding the modal structures of complex multisection photonic devices can be cumbersome. Therefore, though the frequency and time-domain analysis of ML originally was introduced approximately simultaneously (reflecting the two major representations of ML, the sequence of periodic pulses and a comb of locked modes), in the context of semiconductor lasers the work on frequency-domain models remains limited. Instead, theoretical progress has been mainly associated with the time-domain models, as they permit a large-signal approach (large modulation of population inversion, or alternatively large nonlinearity) that we cover in the next sections.

32.3.3 Large-Signal Time-Domain Approach and Delay-Differential Equation Model

Both the assumption of weak to moderate pulse saturation during the pulse and, even more so, that of small gain and absorption per pass, as used in small-signal time-domain models discussed in Section 32.3.1, may become even more tenuous in semiconductor lasers than the assumption of an ideally slow absorption—in fact, in edge-emitting lasers, the small-gain assumption is almost always completely inapplicable since at least one of the laser facets is usually uncoated (or even AR-coated to reduce the reflectance to 0.05%–0.1%) to increase the output power, so the outcoupling losses are by necessity significant, making for large gain per pass even with small saturable absorption. Then, it makes sense to abandon the expansions (Equations 32.14 and 32.15) in the fully numerical procedure and use the full exponential form of Equations 32.2 through 32.8, as well as the more accurate full expression for X in Equation 32.18, thus moving from a small-signal SCP model to a *large-signal iterative model* (see, e.g., Khalfin et al., 1995). This also means that the fast nonlinearities of gain and absorption, and possibly part of the dispersion, may be included directly into the gain and absorber operators.

Even in its large-signal form and with the finite absorber (and gain, if necessary) relaxation time taken into account, the iterative procedure (Equation 32.11) is still somewhat artificial in that it requires a trial pulse shape to start with, and explicitly separates the timescale into the short timescale of the pulse and the long timescale of the repetition period. Moreover, if the time window of the pulse is taken as much smaller than the repetition period (which is the standard thing to do if the repetition period is much longer than the pulse), any instabilities related with secondary pulses arising far away from the main pulse may be missed by the model. In semiconductor lasers, neither of these assumptions is well justified, as the pulse may be only about an order of magnitude shorter than the repetition period, so that the separation of scales is not as justified as in lasers of other types, and the chaotic instabilities with several competing pulse trains are a very real threat.

An elegant solution to these modeling limitations is offered in the form of the most sophisticated and the most realistic of the lumped models of mode-locked lasers. In this form of the lumped approach, the two different scales for pulse analysis are, in general, abandoned, and the iteration procedure (Equation 32.11) is substituted by a *delay* one. In a general form, this procedure may be written as

$$Y(t) = \left(\sqrt{\kappa}\hat{\mathrm{G}}\hat{\mathrm{Q}}\hat{\mathrm{D}}\right) Y(t - T_{\mathrm{RT}}), \tag{32.19}$$

where T_{RT} is again the round-trip of the cold cavity, and t is still the local time of the pulse.

A particularly useful form of this model is obtained if the dispersion operator $\hat{D}$ is expanded as a differential one. An efficient form of such an expansion has been derived by Vladimirov et al. (Vladimirov et al., 2004; Vladimirov and Turaev, 2005) who showed that for a bandwidth limiting element with a Lorentzian spectrum similar to Equation 32.10:

$$\hat{\mathrm{D}} \cdot Y^{\mathrm{T}}(\omega) = \left[\frac{1}{1 - i(\omega - \omega_{\mathrm{p}})/\gamma}\right] Y^{\mathrm{T}} \tag{32.20}$$

(i.e., neglecting GVD), assuming without much loss of generality that the peak gain frequency ω_{p} coincides with one of the laser resonator modes, and taking it as the reference optical frequency, we can rewrite Equation 32.19 as

$$Y(t) = -\gamma^{-1}\frac{\partial Y(t)}{\partial t} + \left(\sqrt{\kappa}\hat{\mathrm{G}}\hat{\mathrm{Q}}\right) Y(t - T_{\mathrm{RT}}). \tag{32.21}$$

Equation 32.21 is a delay-differential one, and the model thus becomes the *delay-differential equation*, or DDE, *model* of ML in semiconductor lasers. The development of this model has been arguably the greatest advance in the theoretical analysis of mode-locked lasers since the original papers by Haus and New (of which it is a rigorous generalization and which it reproduces in limiting cases). It allows a full self-contained treatment of mode-locked operation, including a possibility of some (if by necessity limited) analytical progress with a platform for a full large-signal numerical analysis, which gives a complete, and qualitatively correct (if not necessarily completely accurate), picture of all possible regimes of ML laser dynamics and allows a number of important trends to be identified. Therefore, we shall present it here in some detail, following the original papers (Vladimirov et al., 2004; Vladimirov and Turaev, 2005).

The operators $\hat{\mathrm{G}}$ and $\hat{\mathrm{Q}}$ can be calculated using Equations 32.2 through 32.6 (in this version of the DDE model, no fast absorption is present); the integrated absorption Q is found in Equation 32.18, and for the integrated gain G, a similar equation is written. Assuming that the pulse in the unidirectional cavity treated by the model passes the absorber before the amplifier, the equation takes the form

$$\frac{\mathrm{d}G(t)}{\mathrm{d}t} = -\left[\exp\left(G(t)\right) - 1\right]\exp(-Q(t))\frac{P(t)}{U_\mathrm{g}} + \frac{G_0 - G(t)}{\tau_\mathrm{g}}. \tag{32.22}$$

Here, G_0 is the unsaturated gain determined by the pumping conditions.

Equations 32.21, 32.18, and 32.22 are a closed system suitable for a detailed numerical simulation of both stationary and dynamic behavior of PML. They can also be fairly easily adapted to allow numerical analysis of *hybrid* ML behavior. As shown in Vladimirov and Turaev (2005), the DDE model also allows for significant *analytical* progress, similar to one achieved with classical New's and Haus's models as described above, but for a more general case of large single-pass gain and absorption, more relevant for most semiconductor laser constructions than the classical SCP. In the analytical procedure, the slow absorption and gain approximation have to be reintroduced, and the slow (relaxation of gain and absorption between pulses) and fast (evolution during the pulse) stages of laser dynamics are, as in the traditional SCP model, treated separately. Considering the slow stage results in equations connecting the gain and absorption before and after the pulse, we get

$$G_- = G_0 - (G_0 - G_+)\exp(-T_\mathrm{RT}/\tau_\mathrm{g}), \tag{32.23}$$

and

$$Q_- = Q_0 - (Q_0 - Q_+)\exp(-T_\mathrm{RT}/\tau_\alpha). \tag{32.24}$$

At the fast stage, as usual in the theory of short pulses in lasers and amplifiers, the relaxation terms are omitted, and so Equations 32.18 and 32.22 take the form

$$\frac{\mathrm{d}G(u)}{\mathrm{d}u} = -\left[\exp\left(G(u)\right) - 1\right]\exp(-Q(u)); \quad \frac{\mathrm{d}Q(u)}{\mathrm{d}u} = -s\left(1 - \exp\left(-Q(u)\right)\right), \tag{32.25}$$

where u is the dimensionless energy within the pulse, $u(t) = U(t)/U_\mathrm{g}$, and $U(t) = \int_{-\infty}^{t} P(t')\mathrm{d}t' = v_\mathrm{g}\hbar\omega A_{X\mathrm{g}}\int_{-\infty}^{t}|Y|^2(t')\mathrm{d}t'$. Introducing $u_\mathrm{p} = U_\mathrm{p}/U_\mathrm{g}$, $U_\mathrm{p} = U(t \to \infty)$ as the total dimensionless pulse energy (the time of minus infinity on the short timescale meaning the time before the pulse, and plus infinity, covering the entire time of substantial pulse energy, i.e., the entire pulse duration), one can integrate Equation 32.25 to get another set of equations connecting the prepulse and postpulse gain and absorption:

$$Q_+ = Q(u_\mathrm{p}) = \ln\left[1 + \exp(-su_\mathrm{p})(\exp(Q_-) - 1)\right], \tag{32.26}$$

$$G_+ = G(u_\mathrm{p}) = -\ln\left[1 - \frac{1 - \exp(-G_-)}{\left[\exp(-Q_-)\left(\exp(su_\mathrm{p}) - 1\right) + 1\right]^{1/s}}\right]. \tag{32.27}$$

The pulse energy itself may be calculated from Equation 32.21 by taking the modulus square of both sides of the equation and integrating over the pulse. The result can be expressed as

$$\gamma^{-2} v_g \sigma_g \int_{-\infty}^{\infty} \left|\frac{\partial Y}{\partial t}\right|^2 dt + u_p = \kappa \ln \frac{\exp(G_-) - 1}{\exp(G_+) - 1}. \tag{32.28}$$

In general, the integral of the left-hand side cannot be calculated analytically. Two particular cases when this is possible have been analyzed in Vladimirov and Turaev (2005).

The first is the case of a model without spectral filtering when the integral can be set to zero. As noted by Vladimirov and Turaev (2005), this is a fairly crude approximation, as in fact the value of the integral does not disappear even in the limit of infinitely wide gain dispersion curve ($\gamma \to \infty$). Indeed, the integral is over the time of the pulse, and as such roughly proportional to the pulse duration. In the theories of ML, this duration scales as γ^{-1} meaning that $\left|\frac{\partial Y}{\partial t}\right|^2 \propto \gamma^2 Y^2$, so that the integral remains finite as $\gamma \to \infty$. In fact, as mentioned above, the theory with $\gamma \to \infty$ cannot predict the pulse shape or duration, leading to pulses collapsing to a delta-function shape. The total pulse *energy*, however, can be estimated *approximately* by neglecting the integral in the left-hand side of Equation 32.28 and thus obtaining an equation for u_p in the form

$$u_p = \kappa \ln \frac{\exp(G_-) - 1}{\exp(G_+) - 1}. \tag{32.29}$$

Equations 32.23, 32.24, and 32.26 through 32.29 form a closed system of five (nonlinear and transcendental) equations for the five unknowns: $G_\pm$, $Q_\pm$, and u_p. Vladimirov and Turaev (2005) identified this system as the *generalized New's model*, as it does not include spectral filtering (as the original New's model) but, unlike this model, does include arbitrarily large gain and absorption per pass, which are both essential features of diode lasers. The (numerical) solution gives the dependence of pulse energy (though neither duration nor peak power) on pulse parameters, represented by the unsaturated gain (which is related to pumping current) and absorption (which is related to the reverse bias applied to the absorber and the bandgap detuning between the gain and absorber sections). The other fundamental absorber parameter also dependent on the reverse bias, the absorber lifetime, only enters the calculations through the relaxation Equation 32.24 and does not influence the results from this model at all if $\tau_\alpha \ll T_{RT}$ (in which case, obviously, $Q_- \approx Q_0$).

The solution to this nonlinear algebraic equation system can then be substituted into the inequalities (32.13) to analyze the stability boundaries of the ML operating range with respect to the leading-edge and trailing-edge instability. The curves, in general, can only be calculated numerically; however, Vladimirov and Turaev (2005) noted that the leading-edge and trailing-edge instability boundaries met at the codimension-2 point lying on the linear threshold line (G_0-Q_0-ln$\kappa = 0$). This point can be calculated explicitly as

$$Q_0 = \ln \frac{\kappa(s-1)}{s\kappa - 1}; \; G_0 = \ln \frac{s-1}{s\kappa - 1}. \tag{32.30}$$

This means that the condition $s > 1$ for any range of successful ML to be present, derived in the traditional SCP approach for the case of small gain and loss per period, needs to be generalized in the case of arbitrary losses in the cavity as

$$s\kappa > 1. \tag{32.31}$$

In a more realistic construction, an extra geometric factor could also be required to take into account the fact that the absorber may be saturated by both the forward and reverse propagating wave simultaneously, which is not taken into account by a unidirectional ring laser model on which the approach above is based.

In the case of $G_0, Q_0, \ln\kappa \ll 1$, Equations 32.26 and 32.27 simplify to Equations 32.3 and 32.6, such that the equation for the pulse energy simplifies to $Q_- \frac{\exp(su_\mathrm{p})-1}{s} - G_- \left(\exp(u_\mathrm{p}) - 1\right) - u_\mathrm{p} \ln\kappa = 0$, which is the equation for u_p featured in the original New's theory of ML.

The second case when full (semi) analytical solution of the DDE model (with relaxation terms during the pulse neglected) is possible is when the dispersion is taken into account, but the saturation of gain and absorption during the pulse is assumed to be small, as in the Haus model of ML (though the gain and absorption themselves are not necessarily small, unlike the case of the traditional Haus model). Vladimirov and Turaev (2005) called this the generalized Haus model. In this case, a steady-state solution is sought in the form similar to Equation 32.12 in our notations, $Y(t + T_\mathrm{RT}) = e^{-i\delta\psi} Y(t - \delta T)$. Then, from Equation 32.21,

$$\gamma^{-1} \frac{\partial Y(t-\delta T)}{\partial t} + Y(t - \delta T) = F(u(t))Y(t), \tag{32.32}$$

where

$$F(u) = \sqrt{\kappa}\exp\left(G(u)(1 - i\alpha_\mathrm{Hg}) - Q(u)(1 - i\alpha_\mathrm{Hq}) - i\delta\psi\right) \tag{32.33}$$

is the "complex net gain," which can be written out explicitly, substituting the expressions (Equations 32.26 and 32.27) (with u instead of u_p) for $G(u)$ and $Q(u)$.

Next, assuming that the single-pass pulse shift is significantly smaller than the pulse duration and that the saturation of both the gain and absorption during the pulse is weak enough ($u(t) < u_\mathrm{p} << 1/s$)—the latter being the underlying assumption of Haus's theory—both sides of Equation 32.32 can be expanded in Taylor series up to the second-order terms in their respective arguments:

$$Y(t - \delta T) \approx Y(t) - \frac{\partial Y}{\partial t}\delta T + \frac{1}{2}\frac{\partial^2 Y}{\partial^2 t}\delta T^2,$$

and, generalizing the expansions (Equations 32.14 and 32.15) of the original Haus's theory:

$$F(u) \approx F_0 + F_o' u + \frac{1}{2}F_o'' u^2;\; F_0 = F|_{u=0};\; F_o' = \left.\frac{\partial F}{\partial u}\right|_{u=0};\; F_o'' = \left.\frac{\partial^2 F}{\partial u^2}\right|_{u=0}. \tag{32.34}$$

Then, the equation governing the pulse evolution is obtained in the form

$$\delta T\left(\frac{\delta T}{2} - \gamma^{-1}\right)\frac{\partial^2 Y}{\partial^2 t} - \left(\gamma^{-1} - \delta T\right)\frac{\partial Y}{\partial t} + \left(F_0 - 1 + F_o' u(t) + \frac{1}{2}F_o'' u^2(t)\right) Y = 0. \tag{32.35}$$

Recalling the definition of $u \propto \int |Y|^2 dt$, one identifies this second-order nonlinear differential equation as the generalization of the master equation of ML in Haus's theory, which admits solutions of the same form (Equation 32.16) as the original master equation. Six equations are then obtained for six real parameters: peak pulse power, duration, time shift δT, optical frequency shift $\Delta\omega$, phase shift per round-trip $\delta\psi$, and the chirp parameter β.

The stability limits in the generalized Haus's form can be obtained by substituting these solutions into the conditions (Equation 32.13). In general, they depend on the linewidth enhancement factors; however, for direct comparison with other models, the case of $\alpha_\mathrm{Hg} = \alpha_\mathrm{H\alpha} = 0$ is useful. Results of such analysis, reproduced from Vladimirov and Turaev (2005), are plotted in Figure 32.3. In the plot, the subscript N refers to results from New's model, generalized (solid lines) or standard (dashed lines); and the subscript

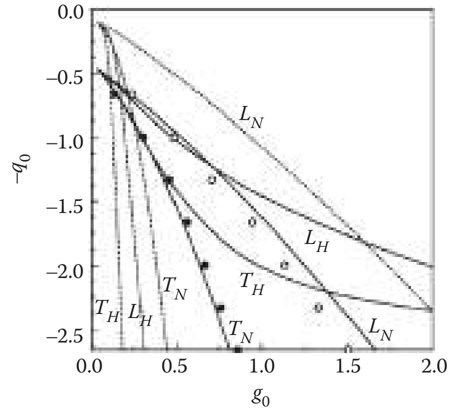


FIGURE 32.3 Stability boundaries of mode locking (ML) with respect to leading (L) and trailing (T) edge instabilities, calculated semianalytically in the DDE approach using traditional and generalized New's (N) and Haus's (H) models. In the calculations, $s = 25$, $T_{RT}/\tau_\alpha = 1.875$, $\tau_\alpha/\tau_g = 0.0133$, $\kappa = 0.1$; $g_0 = (\tau_\alpha/\tau_g)G_0$ as in text; q_0 corresponds to Q_0 in text. (From Vladimirov AG, Turaev D, *Phys Rev A*, 72, 033808, 2005. Reproduced with permission).

H to those from Haus's model (calculated with zero linewidth enhancement factors). The filled/empty dots are the leading/trailing instability boundary calculated by numerical integration of the model. In this numerical integration, the gain and absorber operators are treated on a continuous timescale, without the need to introduce separate timescales for pulse and the free relaxation period as in the iterative procedure. As seen in the figure, standard Haus's and New's models are extremely inaccurate in predicting the instability boundaries of ML in a typical diode laser (with the range predicted by New's model being too wide, and that from Haus's model, too narrow, as noted also in Dubbeldam et al., 1997). The generalized Haus's model gives good agreement within its validity limits at low currents/unsaturated gain values, while the generalized New's model gives very good agreement with numerical simulations at all parameter values (there are some modest deviations that are discussed in more detail below), the reason being that the spectral filtering term neglected in New's approach simply happens to be small in typical diode lasers. Thus, the large-signal nature of the DDE model is proven to be a very important advantage over the classical ML theories.

Apart from allowing some analytical progress in the limiting cases, the DDE model also allows the use of numerical techniques that have been developed for the analysis of DDEs, in particular of numerical packages that allow a full bifurcation analysis of DDEs. Such a study was indeed performed in Vladimirov and Turaev (2005), comprising the full (in)stability analysis of the stationary solution of the DDE. The stationary solution (the steady-state light-current characteristic of the laser) itself is found by seeking the steady-state light output in the form of $Y(t) = Y_{0s}\exp(i\Delta\omega_s t)$. Substituting this into the original Equations 32.21, 32.18, and 32.22 gives the steady-state amplitude and frequency in the parametric form

$$\begin{aligned} &\kappa \exp\left(G_s(Y_0) - Q_s(Y_0)\right) - \Delta\omega_s^2 = 0, \\ &\Delta\omega_s\gamma^{-1} + \tan\left[\Delta\omega_s T_{RT} + \left(\alpha_{Hg}G_s(Y_0) - \alpha_{H\alpha}Q_s(Y_0)\right)/2\right] = 0 \end{aligned} \quad . \tag{32.36}$$

Equation 32.36 is a transcendental trigonometrical equation and thus has an infinite set of formal solutions, corresponding to the cavity modes. The steady-state solution, as usual in the laser theory, is the one with the lowest value of the threshold gain $G_s(Y = 0)$, in other words, the closest to the peak of the gain spectrum. Figure 32.4, after Vladimirov and Turaev (2005), shows the results of a numerical bifurcation analysis of this solution. The line H_1 indicates the *Andronov–Hopf bifurcation* (transition from a steady state to a periodically oscillating solution with an amplitude smoothly increasing from zero as the controlling parameter,

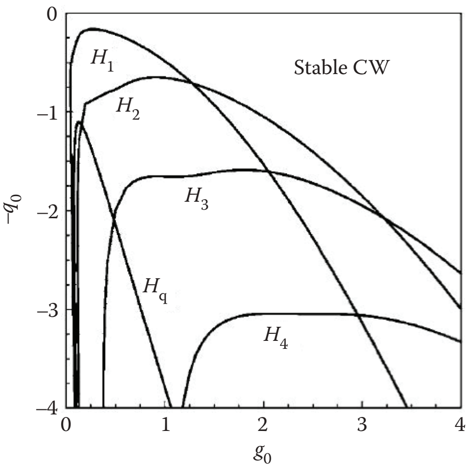


FIGURE 32.4 Bifurcation analysis of the steady-state solutions of the DDE model. Parameters used: $\gamma\tau_\alpha = 33.3$; $\alpha_{Hg,\alpha} = 0$, the rest as in Figure 32.3. The notations H_1, H_2, etc. refer to different harmonic numbers M_h; H_q is the boundary of the Q-switching instability. (From Vladimirov AG, Turaev D, *Phys Rev A*, 72, 033808, 2005. Reproduced with permission).

for example, the unsaturated gain in this case, increases beyond a critical value) corresponding to oscillations at the fundamental ML frequency. ML is predicted for a certain range of conditions regarding the values of the unsaturated gain and absorption, above threshold, whereas at high enough unsaturated gain (or current) and low enough absorption, continuous wave (CW) lasing is expected to be stable. The line H_q indicates the Andronov–Hopf bifurcation corresponding to *passive Q-switching* instability, also known as self-sustained pulsations, which essentially corresponds to the well-known relaxation oscillations in the laser. The positive feedback provided by the SA, which essentially favors pulsed operations, transforms the relaxation oscillations from damped to self-sustained pulsations. The frequency of these oscillations is determined mainly by the unsaturated gain, the gain cross section, the gain relaxation time, and the losses in the cavity, and is typically of the order of 1 GHz, or about an order of magnitude below the ML frequency. Thus, at low frequencies and with high enough amount of saturable absorption in the cavity, the ML pulse train is expected to be modulated by the self-pulsing envelope. The lines H_m, $m > 1$, show the bifurcations corresponding to a solution oscillating at the *m*th *harmonic* of the fundamental ML frequency. At high enough values of unsaturated absorption, there are ranges of G_0 (or current) in which ML at higher harmonics is predicted to be stable, but ML at fundamental harmonic is not.

These predictions are confirmed by a full numerical integration of the DDE model (Figure 32.5), showing the extrema of the laser intensity time dependence calculated for different values of the pumping parameter $g_0 = (\tau_\alpha/\tau_g)G_0$. For each unsaturated gain, the initial transient is omitted before the start of registering signals. At low values of g_0 (and thus current), the laser exhibits a regime when the ML pulse power is modulated by a passive Q-switching envelope, originally with nearly 100% modulation depth (Figure 32.6a). As the pumping parameter increases, the Q-switching modulation gradually decreases in amplitude and eventually the modulation regime undergoes the backward bifurcation, moving to a stable ML regime (this corresponds to the border of the trailing-edge instability in Figure 32.3). Within the area of stable ML, the fundamental round-trip frequency, a train of short pulses, is observed as in Figure 32.7a, whose amplitude increases with G_0. At higher still pumping, the laser dynamics see areas of harmonic ML at the second and third harmonic of the fundamental ML frequency (see Figure 32.7b and c), separated by narrow areas of unstable operation. Finally, the ML breaks up completely with the onset of chaotic modulation of the pulse power, with multiple pulse trains competing in the cavity, as in Figure 32.6b (the regimes separating fundamental frequency ML and harmonic ML areas are similar). Eventually, the system

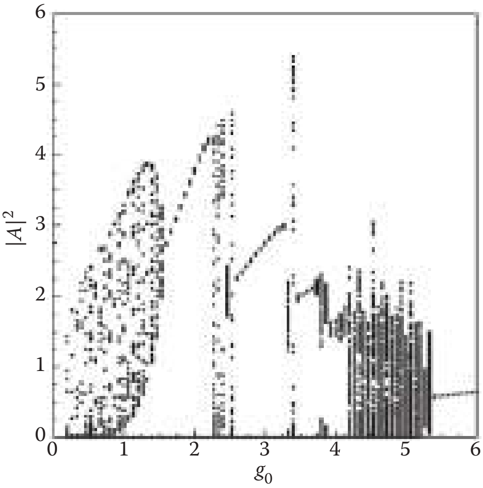


FIGURE 32.5 Bifurcation diagram obtained by direct numerical implementation of a DDE model. $Q_0 = 4$, the other parameters as in Figure 32.3. $|A^2| \propto |Y^2|$ is the normalized output intensity. (From Vladimirov AG, Turaev D, *Phys Rev A*, 72, 033808, 2005. Reproduced with permission).

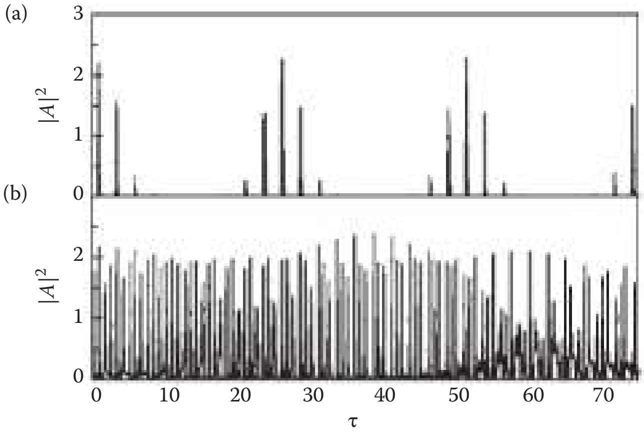


FIGURE 32.6 Illustration of the aperiodic regimes in Figure 32.5: combined mode-locking/Q-switching regime at $G_0 = 50$ (a) and chaotic pulse competition regime at $G_0 = 350$ (b). $|A^2| \propto |Y^2|$ is the normalized output intensity, $\tau = t/\tau_\alpha$ the normalized time. (From Vladimirov AG, Turaev D, *Phys Rev A*, 72, 033808, 2005. Reproduced with permission).

undergoes a transition to CW single-frequency operation in agreement with the bifurcation diagrams of Figure 32.3.

An interesting result obtained in Vladimirov and Turaev (2005) is that, while the conditions (Equation 32.13) of negative net gain before and after the pulse are useful indications of the stability ranges of mode-locked operation, the onset of instabilities in numerical simulations does not coincide with those limits *exactly*. This may be caused in part by the omission of gain dispersion in the analytical study and in part by the neglect of absorber relaxation during the pulse. However, there is also a genuine physical reason for the discrepancy, in that not all small fluctuations in a ML laser were found to grow into full-scale instabilities even if a window of positive gain preceded the ML pulse. Instead, stable ML operation was shown to be possible for a range of parameter (unsaturated gain and absorption) values such that before the pulse,

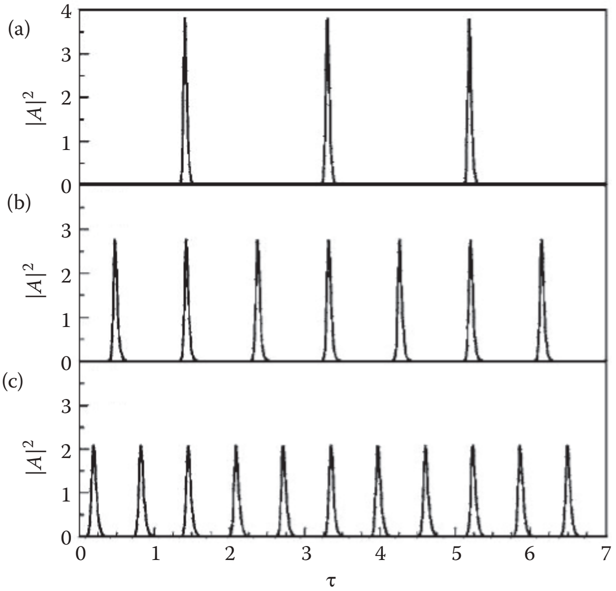


FIGURE 32.7 Illustration of the periodic regimes in Figure 32.5: fundamental frequency ML at $G_0 = 150$ (a) and first and second harmonic ML $G_0 = 225$ (b) and 270 (c). $|A^2| \propto |Y^2|$ is the normalized output intensity, $\tau = t/\tau_\alpha$ the normalized time. (From Vladimirov AG, Turaev D, *Phys Rev A*, 72, 033808, 2005. Reproduced with permission).

the fast absorption had recovered to its unsaturated value, but the slower gain continued recovery, leading to a window of positive net gain preceding the pulse (it may be worth noting that some previous studies, using modifications of Haus's model for semiconductor lasers, indicated the possibility of positive net gain at the *trailing* edge of a stable ML pulse as well; see, e.g., Vladimirov and Turaev, 2005). The possibility of stable ML operation despite a positive net gain window is confirmed by more accurate traveling-wave simulations. One of the consequences of this effect is that the onset of instabilities may be expected to be sensitive to perturbations such as spontaneous noise. The effect of spontaneous emission was indeed studied analytically and numerically in Vladimirov and Turaev (2005), with the noise introduced as a delta-correlated random term in the right-hand side of Equation 32.21. It was concluded that, while the onset of Q-switching oscillations (trailing pulse edge instability) is a dynamic process independent of noise, the onset of the chaotic envelope instability (leading edge instability) is strongly affected by the noise, with an increase in the noise narrowing the window of stable ML. This is fully confirmed by the more complex traveling-wave models (TWMs) described below.

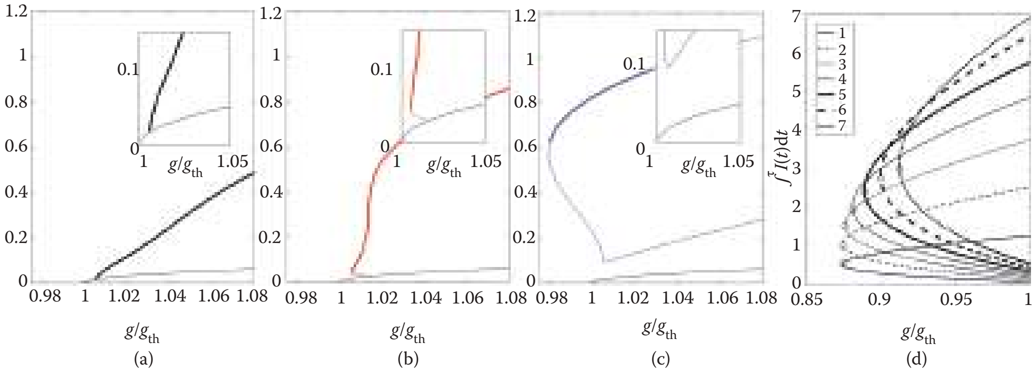
The DDE model, when used as a numerical tool is, not only fully large-signal, but also self-starting: It does not require a trial pulse to start with and can reproduce the emergence of ML pulse train from randomly pulsing light output that is seen as the laser crosses the threshold condition. Thus, the model removes most of the shortcomings traditionally associated with lumped models of mode-locked lasers and presents a relatively simple yet very powerful tool for analyzing the qualitative tendencies of their behavior. As illustrated above, it combines analytical possibilities and numerical methods very naturally within the same framework and, as will be discussed in more detail later, predicts correctly virtually all the dynamic regimes and tendencies observed in a real laser.

An important advantage of the DDE model is that, although strictly speaking derived for the artificial unidirectional ring geometry, it captures enough of the main features of ML to be applicable, with some caution, to predict—at least qualitatively—the phenomena in mode-locked lasers of all types and designs.

An important example is the work presented in Marconi et al. (2014) where the DDE was used to analyze the behavior of mode-locked lasers with *long* delays (cavity round-trip times), comparable to, or even exceeding, the lifetime of carriers in the gain section ($T_{RT} >\sim \tau_g$, unlike the analysis in Vladimirov and Turaev (2005) where the typical situation was $\tau_\alpha < T_{RT} < \tau_g$). Marconi et al. (2014) used a bifurcation analysis similar to that discussed above and presented in Figure 32.3 to analyze the stability of various cavity configurations. The results are shown in Figure 32.8, similar to that presented in Marconi et al. (2014). It was found that the Andronov–Hopf bifurcation mathematically describing the onset of fundamental harmonic ML as discussed above, which is *supercritical* in the case of a short resonator so that ML exists only above its bifurcation point, becomes *subcritical* for a certain value of $T_{RT}/\tau_g > 1$. This means that the (fundamental) ML operation regime can exist below its bifurcation point, coexisting with the CW solution. At longer delays still, the area of stable ML extends below the CW threshold, meaning that ML can coexist with the off solution, thereby implying a bistability between them. Interestingly, during this folding phenomenon, the fundamental ML branch eventually disconnects from the CW solution, meaning that the ML appears for long delays through a saddle-node bifurcation of limit cycles instead of a nascent Andronov–Hopf bifurcation of the CW solution, making this scenario impossible to analyze by any weakly nonlinear analysis such as dynamic modal analysis (Section 32.3.2).

The change in dynamical scenario that occurs in Figure 32.8c has a profound consequence on the mode-locked solutions, as it can be seen in Figure 32.8c and d: The fundamental PML solution becomes stable even in the limit $T_{RT}/\tau_g \gg 1$. Moreover, a very large number of pulsing solutions with different number of pulses per round-trip and different arrangements become stable for the same parameter values. The authors reconstructed analytically some of these solutions, using the generalized New's approximation discussed above, for $T_{RT}/\tau_g = 1$ and restricting the analysis to equally spaced pulses solutions (harmonic PML), as presented in Figure 32.8d. Clearly, all these branches of solutions extend well below the laser threshold, where they stably coexist among them and with the off solution (although the authors noted that New's approximation of neglecting spectral filtering leads to an overestimation of the breadth of the ML region below threshold). This means that the harmonic mode-locked solution of maximal order that exists below threshold becomes fully decomposable, since essentially any pulse of this solution can be set on or off, which the authors confirmed by injecting a digitally modulated sequence of optical pulses into the cavity, which remained stable after a large number of round-trips, meaning that the laser worked as an active version of an optical buffer memory.

The model used in Marconi et al. (2014) was a DDE in its classical form of Vladimirov and Turaev (2005), derived for a hypothetical unidirectional ring laser with large gain and absorption per round-trip (and with the linewidth enhancement factors set to zero as the dynamic effects studied did not significantly depend on them), but the theoretical predictions of the paper were realized experimentally (Marconi et al., 2014)



(a)

(b)

(c)

(d)

FIGURE 32.8 (a) through (c) Bifurcation diagrams similar to that of Figure 32.3 for different cavity round-trip times: $T_{RT}/\tau_g = 1.2$ (a), 2 (b), 4 (c). In the case of a double mode-locked branch, the upper branch is stable and the lower one unstable. Harmonic ML branches are not shown. (d) Folding of harmonic ML branches with a different number for $T_{RT}/\tau_g = 16$. (Reproduced from Marconi M. et al., *Phys Rev Lett*, 112, 223901, 2014).

using a *vertical external cavity surface-emitting* laser (VECSEL). A laser of this type consists of an amplifying (gain) chip and a semiconductor saturable absorber mirror (SESAM) chip, separated by an unguided free-space propagation path (with collimating optics to direct the beam and control the ratio of the spots over the two facets). Various harmonic regimes were realized with this laser, with the number of pulses between 0 and 19 successfully coexisting at high enough currents. The authors then went on (Marconi et al., 2015) to realize the multistability in the long laser to *generate individually addressable* pulses and sequences of pulses, all very well reproduced by the same DDE model.

This work is an important testament to the generality, power, and versatility of the DDE models: The experiments were stimulated by the theoretical predictions which in turn relied on the analytical capabilities (bifurcation analysis using analytical continuation techniques) unique, among semiconductor laser ML theories, to the DDE approach.

It can be pointed out that VECSELs, in fact, are the one class of mode-locked semiconductor lasers for which a delay-differential model, although of a somewhat different form to that used in Vladimirov and Turaev (2005), may be expected to give a fully *quantitative*, as well as qualitative, description of the behavior of a specific and realistic laser construction. As both the gain chip and SESAM are very short asymmetric resonators, with the gain section "length" much shorter than the spatial pulse duration, the lumped-element formalism is a very natural one for their description. The fact that no integration over length is needed for calculating G and Q allows the use of generic nonlinear $g(N)$ and $\alpha(N)$ dependences with no loss of accuracy, as well as introduction of fast gain and absorption saturation omitted in the original DDE of Vladimirov and Turaev (2005). A model based on an approach of this type was successfully used to analyze the dynamics of external-cavity VECSELs in a simple linear cavity (with the laser and SA chips facing each other and the output being from a partially reflecting mirror located between them). The predicted dynamic regimes, pulse durations, and stability ranges matched the ones previously reported in experimental papers, not only qualitatively, but with a reasonable numerical agreement (Mulet and Balle, 2005).

The mathematical distinction of the DDE model usable for quantitative and rigorous description of realistic VECSELs, such as the one used in Mulet and Balle (2005), from the one reported in Vladimirov and Turaev (2005) is, first, that gain and absorption operators, as well as the rate equations for the carriers, need to be modified to take into account the resonator nature of the amplifier and absorber sections and the short length of their active parts. Second, given the Fabry–Pérot rather than ring nature of the resonator, rather than having a single DDE for the light amplitude with the delay time equal to the cavity round-trip, the model for VECSELs in a linear geometry needs separate equations for the dynamics of light amplitudes in the gain and absorber chips, each of them containing a delayed term with a delay equal to *half* of the cavity round-trip.

With further delayed terms introduced, a model of this type can describe different VECSEL geometries, including a folded (rather than linear) cavity one and a colliding pulse operating VECSEL (Avrutin and Panajotov, in preparation).

More advanced constructions known as MIXCELs (standing for mode-locked integrated external cavity surface emitting laser) with the quantum well (QW) gain and QD absorber layers located in one chip, could be described by a similar, possibly even somewhat simpler, model, with the single chip reflectance operator containing the effects of both the gain and the absorption; the model for such a design would therefore be even closer to that of Vladimirov and Turaev (2005).

In the case of *edge-emitting* lasers, the application of a DDE model to a specific, realistic design is somewhat more tenuous. First of all, the DDE model as studied in Vladimirov and Turaev (2005) does not account for fast gain and saturable absorption nonlinearities due to interband processes. Although it could be possible to include them in the QD case, at least in some approximation, the explicit introduction of fast nonlinearities is not necessarily the best strategy. Instead, separate rate equations for dot and reservoir populations can be used. Second, Equations 32.18 and 32.22 for the gain G and absorption Q integrated over the length of the amplifier/absorber element are only accurate if both the gain and absorption have a simple linear dependence on the carrier densities in the corresponding elements, which is in itself an approximation (or if $G, Q \ll 1$). Third, the geometry of the system analyzed in a DDE model in the form presented in Vladimirov and Turaev (2005) is, as in most lumped models, somewhat artificial in that Equations 32.2 and 32.5 are, strictly speaking, valid only in a hypothetical unidirectional ring cavity. In a real laser, the pulse passes through both the amplifier and absorber twice, not once, with the reflected pulse traveling through areas in which gain or absorption has been already partially saturated by the incident pulse, and possibly partially recovered.

An attempt at taking into account a realistic edge-emitting tandem laser geometry in a large-signal *lumped* iterative (not DDE) model, with an end reflector by introducing averaging of gain/absorption over the corresponding sections, with reflections at facets taken into account, was made in a relatively early paper by Khalfin et al. (1995); the model gave good qualitative predictions of the laser performance; however, such an approach has its own inaccuracies as discussed in Avrutin et al. (2000), since it implicitly assumes pulse duration greater than the round-trip time, which by definition is not the case in mode-locked lasers and the results of Khalfin et al. (1995) were never compared to predictions of more accurate models to ascertain their accuracy.

More recently, a study specifically investigating modeling a Fabry–Pérot edge-emitting laser using a DDE model, importantly with a detailed comparison to the more rigorous and accurate model of a travelling wave type (see the next section), has been presented (Rossetti et al., 2011b). The simulations were performed for the special case of QD mode-locked lasers; however, the results appear to be quite generic. Rossetti et al. (2011b) improved the accuracy of the DDE approach by separating the laser into a number F_s of longitudinal sections (e.g., $F_s = 28$ sections were used in the calculations presented in the paper), some of which belong to the gain region and others to the SA, which results in what the authors termed a multisection DDE approach. Essentially, in the notations used here (the formalism of Rossetti et al., 2011b was somewhat more complicated because of account for two-level transitions peculiar to QDs), in a multisection DDE model, Equation 32.21 is rewritten with the single product $\sqrt{\kappa}\hat{G}\hat{Q}$ substituted by a concatenation of gains and losses in individual segments:

$$Y(t) = -\gamma^{-1}\frac{\partial Y(t)}{\partial t} + \left(\prod_{k=1}^{F_s}\sqrt{\kappa_k}\hat{G}_k\right)Y(t - T_{RT}). \tag{32.37}$$

Here, the loss $\sqrt{\kappa_k}$ includes both the distributed losses inside the segment and any lumped scattering/outcoupling loss between the segments k and $k+1$. The complex gain operator $\hat{G}_k$ for each section is calculated in the way similar to Equation 32.2 if the section is within the gain region and similar to Equation 32.5 if it belongs to an SA, with the length used being the length of the section. The gain (or loss) in each segment is calculated from a local rate equation for population inversion. The use of a number of segments gives the model some longitudinal resolution and allows for a more accurate modeling of the outcoupling losses; although not explicitly done in the model of Rossetti et al. (2011b), it can also include fast gain and absorption saturation in addition to the slow processes described by rate equations. The model, however, still contains a single delay term helping maintain the calculation efficiency that is one of the main advantages of the DDE approach. To account for bidirectional propagation, the Fabry–Pérot resonator had to be represented, somewhat artificially, by an equivalent ring resonator twice the length of the Fabry–Pérot one, with the SA length also doubled, and the distributed losses allocated carefully to represent the loss in the realistic cavity. With these approximations, the laser performance simulated by the multisection DDE approach was in good (though not perfect, e.g., the pulse amplitude was accurate to typically within about 10%) agreement with the traveling wave one as regards the pulse shapes, amplitudes, durations, and stability of the results with respect to a chaotic envelope instability. As in most QD mode-locked lasers, the design simulated did not show Q-switching instability, owing to the large damping of the relaxation oscillations, so it is not certain whether the multisection DDE model would accurately predict its limits in a laser in which Q-switching could occur. It can be said that the multisection DDE approach has some similarity to the decimated traveling wave one considered below, though unlike it still retains the unidirectional ring cavity assumption.

The DDE model also has been used recently to investigate the performance of a number of laser cavity designs and operating regimes more complex than purely PML in a simple tandem cavity, whether monolithic or VECSEL type.

In Arkhipov et al. (2013), the DDE model was used to analyze the performance of *hybridly*, rather than passively, mode-locked lasers. As in the experiments with which the calculations were compared, and as in previous simulations using a TWM (Avrutin et al., 1996), the hybrid ML was implemented by voltage modulation at an approximately resonant frequency applied to the SA. Within the DDE approach, this requires the modification of Equation 32.18 in the form

$$\frac{\mathrm{d}Q(t)}{\mathrm{d}t} = -\frac{Q_0 - Q}{\tau_\alpha}\left(1 + a_{\mathrm{mod}}(1 + \mathrm{F}_{\mathrm{mod}}(t)\right) - \left(1 - \exp\left(-Q\right)\right)\frac{P(t)}{U_\alpha}. \tag{32.38}$$

Here, a_{mod} is the amplitude of absorption modulation, $\mathrm{F}_{\mathrm{mod}}(t)$ is a periodic function of time, and $\mathrm{F}_{\mathrm{mod}}$ is a periodic function of time defined so as to vary within the limits $-1 < F_{\mathrm{mod}}(t) < 1$. The authors investigated various modulation profiles, the most straightforward of which was sinusoidal modulation $\mathrm{F}_{\mathrm{mod}}(t) = \cos(2\pi f_{\mathrm{mod}} t)$. The modulation frequency could be near resonance with the fundamental ML frequency, with $f_{\mathrm{mod}} = f_{\mathrm{P}} + \Delta f_{\mathrm{mod}}, \Delta f_{\mathrm{mod}} \ll 1/T_{\mathrm{rt}}$, with $f_{\mathrm{P}} \approx 0.9723/T_{\mathrm{rt}}$ being the free-running PML frequency, or at its second harmonic $f_{\mathrm{mod}} = 2f_{\mathrm{P}} + \Delta f_{\mathrm{mod}}$, or second subharmonic $f_{\mathrm{mod}} = f_{\mathrm{P}}/2 + \Delta f_{\mathrm{mod}}$.

Note that Equation 32.38 captures an important feature of voltage modulation of an SA: Both the unsaturated absorber and the relaxation time are modulated simultaneously.

The efficient DDE model allowed, first and foremost, for very effective numerical search for the *locking range* of hybrid ML (the range of frequency detunings $\delta f = \Delta f_{\mathrm{mod}} T_{\mathrm{rt}}$ within which stable hybrid ML was observed for a given modulation amplitude). This was done by long-time (3000 periods, with the last 200 periods stored) direct numerical simulation of the system (Equations 32.18, 32.21, and 32.22). For improved accuracy, the authors used dual characterization of the locking range, using, first, a straightforward bifurcation diagram with the field maxima and minima plotted, and, second, a stroboscopic diagram where, for each considered δf_{mod}, field intensities separated from each other in time by the interval $1/f_{\mathrm{mod}}$ were collected. Once locking was achieved and the period of the field intensity time trace became equal to $1/f_{\mathrm{mod}}$, all stroboscopic map points had the same value; otherwise, multiple values of the stroboscopic map at a given δf were observed. Figure 32.9 (reproduced from Arkhipov et al., 2013 with permission) illustrates both procedures, as illustrated in Figure 32.9a and b, as well as the calculated *locking tongue* (locking range borders in the coordinates δf, a_{mod}), shown in Figure 32.9c. The latter was in very good qualitative agreement with the observed values. The figure shows the fundamental harmonic locking range (the second harmonic one was very similar as could be expected); however, the simulations also predicted, for the first time and in agreement with the experiments published in the same paper, that a narrower range of *subharmonic* locking was also present due to internal nonlinearities in the laser.

In addition to numerical analysis, the DDE allowed an *analytical* asymptotic model for the calculation of the locking range to be developed, which was in very good qualitative agreement with the numerical simulations up to the modulation amplitude of $a_{\mathrm{mod}} \sim 0.5$–$0.6$. Most importantly, it was shown, in agreement with numerical results, that the magnitude of the locking range was *directly proportional* to the modulation amplitude, a fact also seen in other numerical simulations but only proven analytically in Arkhipov et al. (2013). Explicit, if rather complex, expressions for the proportionality coefficient between the locking range and a_{mod} were derived using perturbative analysis of the periodic ML solution.

Finally, Arkhipov et al. (2013) addressed the issue of the asymmetry of the locking range (it is easier to speed the laser up than to slow down), long known for hybrid ML and shown by both experiments and previous simulations; the authors attributed it to the variation of the absorber relaxation time and hence the average absorption value with modulation.

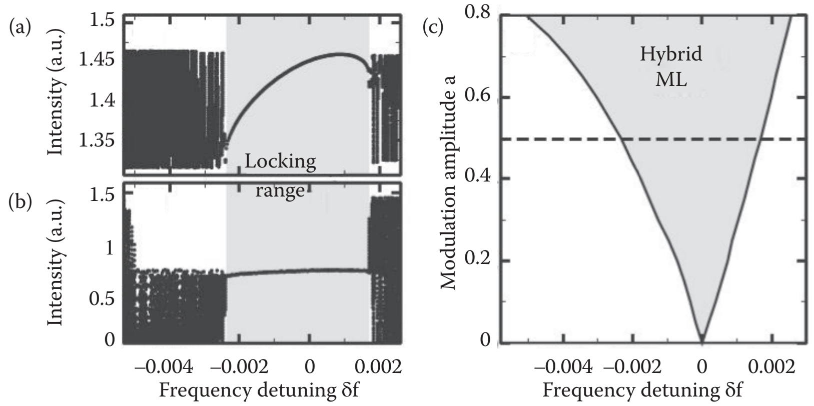


FIGURE 32.9 Numerical DDE estimate of the locking range of a hybrid ML laser. (a) Local maxima of the intensity time trace and (b) 1/fm periodic stroboscopic map points of emitted field intensity time trace at fixed modulation amplitude $a = 0.5$. In the locking range (gray), all ML pulses have the same peak intensity. (c) Locking tongue in the plane of two parameters: frequency detuning δf and modulation amplitude a. (From Arkhipov A. et al., *IEEE J Sel Top Quantum Electron*, 19, 1100208, 2013. Reproduced with permission).

The DDE was also used for modeling the dynamics of mode-locked lasers under the more special conditions of optical injection with a single (Rebrova et al., 2011) or, most recently, dual (Arkhipov et al., 2016c) optical lines, as well as under the condition of external optical feedback (Jaurigue et al., 2015, 2016; Otto et al., 2012). In all of these conditions, the versatility of the DDE allowed analytical insight into operating conditions, as well as direct numerical simulations.

Interestingly, the somewhat abstract nature of the DDE model led to the fact that two very different forms of the model could be used successfully to analyze the situations that in the experiment can be quite similar. In Jaurigue et al. (2015) and Otto et al. (2012), a model with multiple delays (as in the generalized Lang-Kobayashi model of a single-frequency laser with optical feedback from a strong external reflector) was used to analyze ML under external optical feedback, whereas in Arkhipov et al. (2015b), harmonic ML in a compound cavity consisting of an active laser subcavity and a passive one formed by an external reflector (which is very similar to optical feedback) was investigated by modeling both cavities as unidirectional ring ones as in the traditional form of the DDE; mathematically, the compound cavity was represented by two coupled DDEs. Both models showed, and were used to estimate the ranges of, fully or partially rendered harmonic regimes given an integer rational relation between the cavity lengths, and stressed the importance of the subwavelength variations in the cavity length (represented by phase shifts of light amplitude). All of these had been previously independently investigated for Fabry–Pérot laser geometries using frequency-domain and/or traveling wave time-domain models (see the next section), e.g., in Avrutin and Russell (2009). However, the advantage of the DDE model was, first, the possibility of efficient and instructive analysis of the bifurcation diagrams of the laser and, second, the generality as the results apply not just to Fabry–Pérot resonators but to other designs such as ring lasers.

A special version of the DDE model was developed (Viktorov et al., 2006) for analyzing *QD* mode-locked lasers. Detailed description of the properties of QDs as an active medium is beyond the scope of this chapter; the reader is referred to a specialized monograph (Rafailov et al., 2011) or the relevant chapters in the current handbook. In brief, there are two major (interrelated) features that distinguish the QDs from other semiconductor active media, particularly in the context of ML. The first of these is the complex carrier kinetics which in the case of the gain media involves the relatively slow (~5–10 ps) capture of carriers (electrons and holes) into the QDs and subsequent interlevel relaxation, and in the SA sections, the relatively complex nature of carrier escape involving intermediate levels (Viktorov et al., 2009). The second specific feature of QDs as an active medium is the noticeably nonequilibrium distribution of carriers between dots of different sizes and compositions (which therefore have different energy levels resulting in the inhomogeneous broadening of the laser line), and the possibility of dual-wavelength lasing due to the existence of (at least) two electron levels in each dots, the ground and excited state.

The original DDE model of QD mode-locked lasers (Viktorov et al., 2006) ignored inhomogeneous broadening and the excited-level transitions, and thus captured the first of these characteristic features but not the second. The equation for light field in the model was thus essentially the same as in the standard DDE but the equations for gain and absorber dynamics were changed more significantly to reflect the specific features of QDs. Still, even in this simplified form, the model allowed an explanation for a number of features of QD mode-locked lasers, such, first, as the suppressed Q-switching instability (Viktorov et al., 2006) due to the slow carrier capture in the gain section and, second, the enhanced ML at high temperatures due to faster SA relaxation (Cataluna et al., 2006, 2007). Later, more advanced versions of DDE for QD ML lasers have been developed with inhomogeneous broadening and lasing from different levels taken into account, e.g., Cataluna et al. (2010) and also Rossetti et al. (2011b) already mentioned above.

To summarize, DDE models are a very powerful tool capable of predicting, qualitatively describing, and giving unique analytical insight into all the main features and many of the peculiarities of ML in a semiconductor laser. With some caution, models of this type can be used for quantitative description of the performance of a specific laser design, but the reliability and accuracy of such a procedure may be limited in the case of realistic, multisection edge-emitting laser designs, particularly as they often comprise, in addition to the gain and absorber section, elements such as Bragg mirrors, phase-tuning sections, etc.

For these purposes, TWMs are preferable; they will be considered in the next section.

32.3.4 TWMs: The General Considerations

The most accurate and realistic, though usually the most computationally intensive, approach to simulating edge-emitting mode-locked lasers is offered by *distributed time-domain*, or *TWMs* (shown schematically in Figure 32.2c), which treat the propagation of an optical pulse through a waveguide medium with spatial as well as temporal resolution. The model then starts with decomposing the optical field in the laser cavity into components propagating forward (subscript "+") and backward (subscript "–") in the longitudinal direction (say, z):

$$Y(r,t) = \Phi(x,y)\left(Y_+ \exp(i\beta_{ref}z) + Y_- \exp(-i\beta_{ref}z)\right)\exp(-i\omega_{ref}t), \tag{32.39}$$

with Φ being the transverse/lateral waveguide mode profile and ω_{ref} and

$$\beta_{\mathrm{ref}} = n(\omega_{\mathrm{ref}})k_{\mathrm{ref}} = n(\omega_{\mathrm{ref}})\omega_{\mathrm{ref}}/c$$

being the reference optical frequency and the corresponding wave vector, respectively. This results in a reduced equation for slowly varying amplitudes $Y_\pm$, which has the form

$$\pm\frac{\partial Y_\pm}{\partial z} + \frac{1}{v_{\mathrm{g}}}\frac{\partial Y_\pm}{\partial t} = \left(\frac{1}{2}(\widehat{g}_{\mathrm{mod}} - \alpha_{int}) + ik_{\mathrm{ref}}\widehat{\Delta\eta}_{\mathrm{mod}}\right)Y_\pm + iK_{\pm,\mp}Y_\mp + F_{\mathrm{spont}}(z,t). \tag{32.40}$$

The equation is directly solved numerically without the partially analytical integration involved in deriving (Equations 32.2 and 32.5).

The gain and saturable absorption coefficients are most often parametrized as functions of the carrier density and, through the gain and absorption compression coefficients ε_{g} and ε_α, on the photon densities, which is most efficiently implemented using simple relations:

$$g = \frac{g_{\mathrm{lin}}(N)}{1+\varepsilon_{\mathrm{g}}S}; \quad \alpha = \frac{\alpha_{\mathrm{lin}}(N)}{1+\varepsilon_{\alpha}S}, \tag{32.41}$$

where $S = |Y_+|^2 + |Y_-|^2$. Note that taking the total intensity in the denominator of Equation 32.41 for both left- and right-traveling waves, although often used, may be an oversimplification for some problems as it assumes identical cross- and self-saturation coefficients between left- and right-traveling waves. A more accurate analysis may be important, in particular, in ring lasers (see Chapter 31 for more details).

The carrier density dependences of the linear gain and absorption $g_{\mathrm{lin}}(N)$ and $\alpha_{\mathrm{lin}}(N)$ are, in the simplest version of the model, taken in the standard linear $g_{\mathrm{lin}}(N) = \sigma_{\mathrm{g}}(N - N_{\mathrm{tr}})$ or logarithmic $g_{\mathrm{lin}}(N) = \sigma_{\mathrm{g}}N_{\mathrm{tr}} \ln \frac{N+N_1}{N_{\mathrm{tr}}+N_1}$ forms typical for semiconductor laser modeling in general (Coldren et al., 2012) (N_{tr} being the transparency carrier density and σ_{g}characterizing the gain cross section near transparency as shown in Sections 3.1 and 3.2; for the absorption, the linear approximation $\alpha_{\mathrm{lin}}(N) = \alpha_0 - \sigma_{\alpha}N$, also as shown in Sections 32.1 and 32.2, is most often used). In more accurate implementations (e.g., Javaloyes and Balle, 2010b), absorption $g_{\mathrm{lin}}(N)$ and $\alpha_{\mathrm{lin}}(N)$ are calculated microscopically with varying degrees of rigor; some of these implementations will be discussed in more detail in Section 32.5. The compression factors $\varepsilon_{\mathrm{g},\alpha}$ also may be either introduced phenomenologically or calculated microscopically for the two main types of optical nonlinearities in bulk and QW lasers: spectral hole burning and dynamic carrier heating. As pulse duration decreases and particularly for multi-GHz ML, the finite (subpicosecond) relaxation times $\tau_{\mathrm{nl}}^{(\mathrm{g,a})}$ of the nonlinearities become important. To take those into account, Equation (32.41) can be substituted by phenomenological relaxation equations: $\frac{\mathrm{d}\alpha}{\mathrm{d}t} = \frac{1}{\tau_{\mathrm{nl}}^{(a)}}\left(\frac{\alpha_{\mathrm{lin}}}{1+\varepsilon_{\alpha}S} - \alpha\right)$; $\frac{\mathrm{d}g}{\mathrm{d}t} = \frac{1}{\tau_{\mathrm{nl}}^{(\mathrm{g})}}\left(\frac{g_{\mathrm{lin}}}{1+\varepsilon_{\mathrm{g}}S} - g\right)$ (Martins et al., 1995). Some authors choose not to introduce $\varepsilon_{\mathrm{g},\alpha}$ due to carrier heating at all, instead including microscopic analysis of carrier temperature dynamics (Bischoff et al., 1997) and gain-carrier temperature dependence into the model. In QD lasers, with their strong spectral hole-burning effects (nonequilibrium carrier distribution in energy, with the energy levels resonant with the photon energy preferentially depleted), and relatively slow intradot relaxation, kinetic processes are often treated explicitly without introducing $\varepsilon_{\mathrm{g},\alpha}$ (see Section 32.5.1).

The dynamic correction $\Delta\beta = \Delta\eta_{\mathrm{mod}}k_{\mathrm{ref}}$ to the propagation constant, in bulk and QW lasers, is often approximated as related to the gain variation by means of a single parameter, the Henry's linewidth enhancement factor α_{H} (with different values used for the gain and absorber sections), e.g., using a relation $\Delta\beta = \Delta\beta_{\mathrm{SPM}} = -\alpha_{\mathrm{H}}(g - g_{\mathrm{th}})$, the latter parameter being the threshold (or any other reference) value of peak gain. This phenomenological approach, although the simplest and the most traditional, ignores the fact that the spectral dependence of gain and carrier-induced refractive index correction can be different, so the linewidth enhancement factor should be, generally speaking, spectrally and carrier density dependent.

The gain dispersion represented by the operator nature of gain $\hat{g}_{\mathrm{mod}}$ (and to a certain extent GVD, represented by the operator nature of the modal refractive index $\widehat{\Delta\eta}_{\mathrm{mod}}$) is very important in determining the *stability range* of ML (unless there is a dispersive element in the laser construction such as a DBR). As mentioned before, in the lumped model, no stable ML with a finite pulsewidth can be simulated in the absence of dispersion. In the distributed model, most models not including gain dispersion cannot predict stable ML either. Some authors (Bischoff et al., 1997) reported stable ML with finite pulsewidths simulated without the dispersion term, but the model of Bischoff et al. (1997) included finite relaxation times of nonlinearities, which may have had a side effect of introducing *effective* dispersion. As regards the numerical implementation of dispersion, in mode-locked laser constructions realized so far, the spectrum of mode-locked lasers, although quite broad, is still usually significantly narrower than that of gain/absorption, meaning that only the top of the gain curve needs to be represented accurately. Therefore, in most studies reported so far, the dispersion has been approximated in frequency domain, i.e., as function of $\Delta\omega = \omega - \omega_{\mathrm{ref}}$, as a simple, Lorentzian curve in complex numbers (similar to Equation 32.10 with $D = 0$), or Equation 32.20, equivalent to approximating the spectral properties of the material by those of an equivalent two-level medium with homogeneous broadening:

$$P^{\mathrm{T}}_{\pm}(\Delta\omega) = -i\hat{g} \cdot Y^{\mathrm{T}}_{\pm}(\Delta\omega) = -i\frac{g(N,S)}{1 - i\left(\Delta\omega - \Delta\omega_{\mathrm{p}}(N)\right)/\gamma(N)} Y^{\mathrm{T}}_{\pm}(\Delta\omega), \tag{32.42}$$

where the superscript T means Fourier transformed variable in frequency domain; $\gamma(N)$ is the gain spectral width parameter as in Equations 32.10 and 32.20; $\Delta\omega_{\mathrm{p}}(N)$ is the spectral shift of the gain peak from the reference frequency, and $g(N, S)$ can be implemented as in Equation 32.41 (for the absorption, the same method can be used). In time domain, this can be implemented numerically by two alternative but largely equivalent methods. The first one, used in a number of papers (e.g., Bandelow et al., 2001), and traceable to early work on pulse generation in lasers with active media with homogeneously broadened gain spectrum (Fleck, 1968), consists of introducing a separate differential equation for gain polarization, as in the DDE approach. With the gain spectrum centered at the reference frequency ($\Delta\Omega'' = 0$), it takes a particularly simple form, similar to that of Equation 32.21 in the DDE approach:

$$\frac{\partial P_{\pm}(z,t)}{\partial t} = -\gamma\left(P_{\mathrm{f,b}}(z,t) - i\frac{g}{2}(N,S)Y_{\mathrm{f,b}}(z,t)\right). \tag{32.43}$$

The second method of introducing gain dispersion involves using digital filters of varying complexity (Avrutin et al., 2000, 2005; Heck et al., 2006). In the case of a simple Lorentzian gain, the filter is straightforwardly represented as an infinite impulse response (IIR) one of the form of

$$P_{\pm} = -i\hat{g}Y_{\pm} = -i\frac{\gamma}{2}\int_{o}^{\infty} g(z,t-\tau)Y_{\pm}(z,t-\tau)\exp\left(-\tilde{\gamma}_{\mathrm{p}}\tau\right)\mathrm{d}\tau;\ \tilde{\gamma}_{\mathrm{p}} = \gamma - i\Delta\omega_{\mathrm{p}}. \tag{32.44}$$

In practice, the integral requires only storing one iteration in the computer memory: For small integration steps Δt it is easily implemented using a slight generalization of the formula given originally in Schell et al. (1991) as

$$P_{\pm}(t) = \exp\left(-\tilde{\gamma}_{\mathrm{p}}\Delta t\right)P_{\pm}(t-\Delta t) - i\left(\gamma/\tilde{\gamma}_{\mathrm{p}}\right)\left(1 - \exp\left(-\tilde{\gamma}_{\mathrm{p}}\Delta t\right)\right)\frac{g(t)}{2}Y_{\pm}(t).$$

This method of gain dispersion implementation is more tolerant to the simulation time step than the separate differential equation (Equation 32.43), but may be not very accurate if the steps are not small enough (Avrutin et al., 2005; Schell et al., 1991). A more complex, but also somewhat more robust, form of a digital filter implementation of dispersion was described in Carroll et al. (1998) and applied to ML (see, e.g., Jones et al., 1995).

In a recent paper (Javaloyes and Balle, 2010b), a more complex digital filter, representing a more accurate model of the spectra of the complex dielectric permittivity (gain/absorption and refractive index), derived from a microscopic approach and allowing for the realistic fundamental absorption edge spectrum to be modeled, has been implemented; this will be discussed in Section 32.5.

The dispersive nature of the correction $\widehat{\Delta\beta} = k_{\mathrm{ref}}\widehat{\Delta\eta}_{\mathrm{mod}}$ to the (real part of) the propagation constant is usually less important than gain dispersion. In cases when very short (subpicosecond) pulses may be expected in the simulation, the operator $\hat{\Delta\beta}$ may include an additional term describing GVD of the structure (Avrutin et al., 1996)

$$\hat{\Delta}\beta Y = k_{\mathrm{ref}}\widehat{\Delta\eta}_{\mathrm{mod}}Y = \hat{\Delta}\beta_{\mathrm{SPM}}Y - \frac{\beta_2}{2}\frac{\partial^2 Y}{\partial^2 t^2}, \tag{32.45}$$

where the first term describes the self-phase modulation effects, and the second, the GVD, with $\beta_2 = \frac{1}{c}\frac{\mathrm{d}n_{\mathrm{g}}}{\mathrm{d}\omega}$ the first-order GVD coefficient, n_{g} being the group velocity refractive index of the laser waveguide. Numerical simulations (Avrutin et al., 1996) show that GVD affects the parameters of picosecond pulses significantly for the dispersion values of $\mathrm{d}n_{\mathrm{g}}/\mathrm{d}\omega \sim 10^{-14}$ s. Thus, this term is usually negligible in most QW lasers (where the GVD magnitude is estimated as $\mathrm{d}n_{\mathrm{g}}/\mathrm{d}\omega = 10^{-16}$–$10^{-15}$ s) and indeed is omitted in most models of mode-locked laser diodes published to date. In the microscopic or semimicroscopic implementations of TWM (Section 32.5), the dynamic and spectral variation of real, as well as imaginary, part of the dielectric permittivity of the material, and thus the active layer contribution to GVD, is implemented self-consistently. The dispersion (material and waveguide) of the passive waveguide structure is neglected but is believed to be weaker.

The terms containing the forward-back and back-forward propagating coupling constants $K_{\pm,\mp}$ need be included in the model only if a *Bragg grating* is present at the position z (and time t); the constants $K_{\pm}$ and $K_{\mp} = K_{\pm}^{*}$ are, in general, complex due to both refractive index and gain/absorption grating being possible. In the context of a mode-locked laser, accounting for a grating may be needed either if the laser construction contains a DBR section, or to account for *standing wave-induced gratings, or short-scale spatial hole burning* that is important if pulses propagating in opposite directions collide in the active medium (coherent colliding pulse effect). The standing wave-induced grating exists due to the carrier population being increased in the antinodes of the standing wave. Within the SA, where ML pulses are typically engineered to collide, this decreases the local absorption (Martins et al., 1995) and forms an absorption (and possibly refractive index, due to self-phase modulation) grating, to which the fast nonlinearities responding to the standing wave also contribute. The magnitude of the periodic carrier density modulation is then given by the equation (Martins et al., 1995)

$$\frac{\mathrm{d}}{\mathrm{d}t}N_{\mathrm{grat}}(z,t) = -\frac{N_{\mathrm{grat}}(z,t)}{\tau_{\mathrm{grat}}} + \frac{v_{\mathrm{g}}}{2}\left(Y_{+}^{*}\hat{\alpha}Y_{-} + Y_{-}^{*}\hat{\alpha}Y_{+}\right), \tag{32.46}$$

where $\hat{\alpha} = -\hat{g}$ is the saturable absorption operator; the grating relaxation time in QW materials is mainly determined by ambipolar diffusion with the coefficient D_a:

$$\frac{1}{\tau_{\mathrm{grat}}} = \frac{1}{\tau_{\alpha}} + \frac{16\pi^2 D_{\mathrm{a}} n_{\mathrm{g}}^2}{\lambda^2}, \tag{32.47}$$

where λ is the lasing wavelength in vacuum. An estimate gives a value of ~1 ps for the ambipolar diffusion coefficient of 2×10^{-4} m^2/s typical for III–V materials. From the magnitude of carrier density modulation, the coupling in QWs can be estimated as $K_{\mathrm{fb}} \approx i\frac{\partial\alpha}{\partial N}N_{\mathrm{grat}}(z,t)(1 - i\alpha_{\mathrm{H}\alpha})$. Assuming the Henry factor of the absorber is small, as was done in Martins et al. (1995), the grating becomes a purely absorption grating, smoothed down by diffusion. Equation 32.46 is written in terms of SA rather than gain section parameter since in ML lasers designed to utilize the colliding pulse effects, pulses traveling in different directions collide in the SA rather than the gain sections (the former, as will be discussed in the following section, assists ML, the latter impedes it). However, a TWM should contain the grating population in all type of sections; in gain sections, the gain operator $\hat{g} = -\hat{\alpha}$ is used.

The final term in Equation 32.40 is the random noise source that leads to the self-starting of the model and is essential for modeling of noise and pulse jitter. At the laser facets, standard reflection/transmission boundary conditions are imposed on $Y_{\pm}$; thus, unlike the delay-differential model, the traveling wave one accounts accurately for the laser geometry.

The traveling-wave equations are coupled with coordinate-dependent rate equations for the relevant populations. In the context of QW and bulk lasers:

$$\frac{\mathrm{d}}{\mathrm{d}t}N(z,t) = \frac{J(z,t)}{ed_{\mathrm{a}}} - N\left(BN + \frac{1}{\tau_{\mathrm{nr}}} + CN^2\right) - v_{\mathrm{g}}Re\left(Y_{+}^{*}\hat{g}Y_{+} + Y_{=}^{*}\hat{g}Y_{-}\right), \tag{32.48}$$

with J/ed as the pumping term, J being the current density, e the elementary charge, d_{a} the active layer thickness, τ_{nr} the nonradiative recombination rate, and B and C are usually identified as the bimolecular recombination constant and the Auger recombination rate, respectively. Carrier capture dynamics is sometimes taken into account by adding an extra equation for carrier densities in the contact layers, but its significance for most mode-locked lasers (with the exception of QD active media and possibly some specially engineered QW constructions) tends to be modest, except where direct current modulation is involved.

TWMs are very powerful and general and their use is not restricted to ML edge-emitting lasers (see Chapter 31).

The main limitations of TW models are, first, the absence of any analytically solvable cases—the approach is by its very essence numerical, though some analytical insight is given by the modal decomposition of the traveling wave solution (see Chapter 31). Second, there is the fact that, in their traditional form, TWMs pose considerably higher requirements on the computing time and memory compared to delay-differential models. This is mainly due to the fact that, in the traditional implementation of the TWMs, time and space steps are usually related as $\Delta z = v_{\mathrm{g}}\Delta t$, meaning that the computational timescales as $1/\Delta t^2$ and need to be sufficiently short (a typical spatial step being 1–5 μm) to reproduce the pulse characteristics faithfully (the problem is mathematically stiff). However, with the development of computer resources, this limitation has become progressively less important; and the numerical technique of *decimation* has allowed efficient decoupling of time and space steps (see Section 32.5.2). Several commercial or free software simulators of laser diodes include traveling-wave approach of some form as the core of their solver; some of those are directly applicable and have indeed been applied for the analysis of ML lasers; see, e.g., Avrutin and Rafailov (2012), Section 32.5.2 Avrutin and Rafailov (2012), and Chapter 31.

32.4 The Main Predictions of Mode-Locked Laser Theory

In this section, we shall overview some of the main results of ML modeling that can be obtained on a mainly phenomenological level.

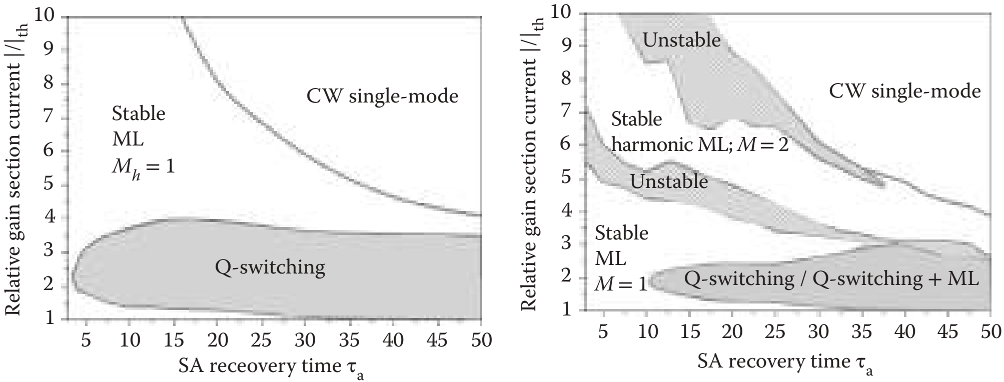
32.4.1 Operating Regime Depending on the Operating Point

The most basic result of all the modern ML theories, confirmed by the experiments, is that the dynamics of semiconductor lasers intended for ML can be quite rich and can show, apart from stable ML near the fundamental round-trip frequency, a number of other dynamic regimes. Here, we shall briefly discuss the general trends in their dependence on the laser parameters.

One of the most important features in the dynamic map of operating regimes of a mode-locked laser is the self-sustained pulsations, or passive Q-switching instability at low currents. As shown in Figure 32.5, produced by the DDE model, the range of currents, or unsaturated gain, values in which this regime is observed increases with the amount of saturable absorption in the laser (which, in a given laser construction, either QW or QD, may be varied to some extent with reverse bias, due to electroabsorption). The other model parameter affected by the reverse bias is the *absorber lifetime* τ_{α}, which is known to decrease approximately exponentially with the reverse bias in QW materials (see Nikolaev and Avrutin, 2003 and references therein) and to some extent in QDs too (Malins et al., 2006). The dependence of the Q-switching range on τ_α is not straightforward; the *Q*-switching range tends to be broadest at a certain absorber recovery time, of the order of the round-trip time though somewhat longer, as can be seen for example in Figure 32.10. At longer τ_α values, the SP range slightly decreases. However, it also decreases as τ_α is *decreased*, and when τ_α reaches a certain value, of the order of a fraction of the round-trip, the Q-switching instability disappears completely, leaving a broad area of stable ML. This is illustrated in Figure 32.10, which is produced using a TWM and shows approximate borders of different dynamic regimes for a representative laser with a short, relatively broadband (length 50 μm, coupling coefficient $\kappa = 120\ \text{cm}^{-1}$) DBR section similar to that realized in Bandelow et al. (2006) and with a cavity length designed for the fundamental ML frequency either near 80 GHz (Figure 32.10a) or near 40 GHz as in the experimental study (Figure 32.10b).

The importance of both τ_α and the unsaturated absorption in determining ML properties means, first, that care needs to be taken when interpreting the bias voltage effects on the performance of either QW or QD mode-locked laser, as the unsaturated absorption, the saturable absorption cross section, and the SA recovery time τ_α are all likely to be affected. The effect on the latter is probably the most significant though, as the dependence of τ_α on voltage is quite strong (exponential), while the effect on the unsaturated absorption appears, from measured threshold currents, to be more modest. Second, it means that for the same absorber parameters, longer lasers with longer repetition periods are less likely to suffer from the Q-switching instability, which needs to be kept in mind when analyzing the dynamics of QD lasers (due to the relatively low gain, these often have to be quite long if stable operation at the ground level wavelength band is desired).

The lower current (or unsaturated gain) limit of the self-pulsing instability may be positioned either below or above the low boundary of ML itself, depending on the gain and absorber saturation energies (*s*-parameter) and the absorber recovery time. If the boundary for ML is below that for self-pulsing (which tends to happen in longer lasers, when τ_α is significantly smaller than T_{RT} but not small enough to completely eliminate self-pulsing), then the stable ML range is split in two by the self-pulsing area, with an area of stable ML seen below the Q-switching limit at currents just above the threshold. The area is narrow, however, and the pulse powers generated in this regime are typically rather low. If, on the other hand, the boundary for ML is above that for self-pulsing (which tends to be the case for shorter lasers or longer absorber relaxation time, when $T_{RT} > \tau_\alpha$), then an area of pure self-pulsing, with noisy/chaotic filling of pulses, is seen at small to modest excess currents above threshold, as in Figure 32.5; as the current is increased, the pulses acquire a regular structure and the combined ML/SP regime develops.



(a)

(b)

FIGURE 32.10 Schematic diagram of regimes in a generic QW mode-locked laser operating at the repetition rate of 80 GHz (a) and 40 GHz (b), calculated using a phenomenological TWM.

Comparing Figure 32.10a and b, we notice that the dynamics of the laser become richer with the increase of the cavity round-trip time, or, equivalently, of the number of modes in the laser spectrum. Indeed, the dynamic behavior of the short-cavity laser in Figure 32.10a displays only regular deterministic regimes (Q-switching, ML, CW) and only the fundamental frequency ML ($M_\mathrm{h} = 1$ in Equation 32.1). In contrast, the dynamics of the longer laser in Figure 32.10b contain a range of currents corresponding to stable *harmonic* ML with $M_\mathrm{h} = 2$, as in the DDE model. It is separated from fundamental frequency ML by a band of currents in which the laser shows the second (after Q-switching) main type of instability of ML: the leading edge, or chaotic, instability. As was shown by the DDE model, the onset/lower border of this instability is pushed somewhat toward higher currents by the increased amount of absorption in the laser. In addition, as illustrated in Figure 32.10b, shortening absorber relaxation time also somewhat decreases the risk of this instability.

At high currents, the unstable operating regime gradually evolves into some type of (irregular) quasi-CW operation (few modes present), which for some parameter values, with higher currents still, gives way to single-frequency, stable CW operation, as shown in Figure 32.10 (for longer τ_α) and predicted also by DDE. Whether or not this true single-frequency CW operation is achieved at a practically feasible current depends on the length of the laser and the gain bandwidth; longer lasers (with a repetition frequency ~10 GHz and below) with broader gain spectrum tend to not reach true CW under any realistic pumping current, instead operating in a chaotic quasi-CW regime with a narrow spectrum including only a few modes.

In shorter lasers, as in Figure 32.10, a direct transition from stable ML, fundamental (Figure 32.10a) or harmonic (Figure 32.10b), to CW operation is also possible. This takes the form of a Hopf bifurcation, with the ML pulses acquiring a constant background with increased current and then their amplitude gradually reducing to zero resulting in CW operation.

32.4.2 The Main Parameters That Affect Mode-Locked Laser Behavior

In addition to determining the operating *regime* of the laser as discussed in the previous section, the theoretical models allow also the effects of the operating point on the main parameters of the optical pulse to be analyzed. Here, we shall concentrate mainly on the PML regime, in which case the most important parameters whose effect on the laser behavior the modeling can allow us to investigate are as follows:

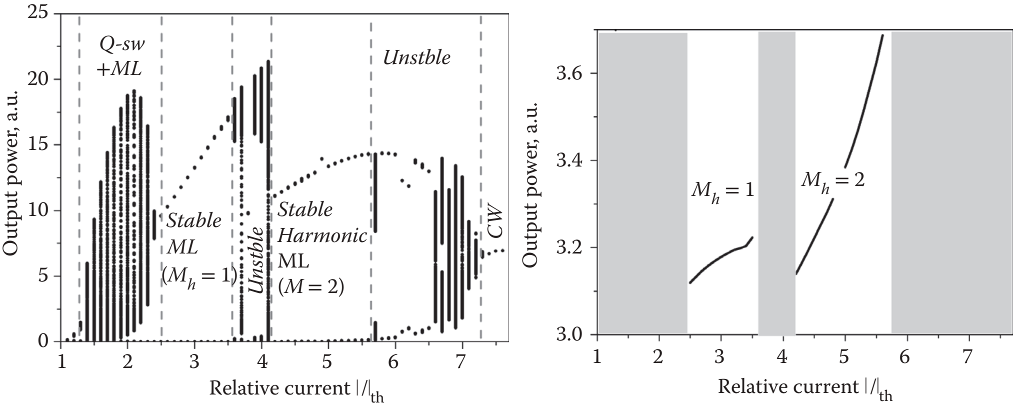
32.4.2.1 The Pumping Current

The pulse *amplitude* grows with pumping current, as illustrated in Figure 32.11, calculated by a traveling-wave simulation of the laser of Figure 32.10b. Notice that qualitatively the behavior of the pulse amplitude shown in Figure 32.11a is quite close to the bifurcation behavior seen from Figure 32.5, except that in the TWM, and with the inevitably different set of parameters, only the second rather than third harmonic operation is predicted (it was also noted by some authors (Vladimirov et al., 2009) that the dynamics of an ML laser in the TWM can predict stable or unstable trailing pulses at different time detuning from the prevalent pulse stream, whereas the DDE simulation gives only harmonic operation even in unstable regimes).

As regards the pulse *duration* shown in Figure 32.11b, it has been predicted by early frequency-domain theories (Lau and Paslaski, 1991) to reach a minimum near the area of Q-switching instability; this has been later confirmed by both experiments and time-domain simulations and can be also seen in Figure 32.11b. The pulse duration thus tends to grow with current within the stable ML range *above* the upper boundary of Q-switching. On the other hand, if an area of stable ML below the Q-switching range is observed (which, as discussed above, is sometimes the case in longer resonators), then a *decrease* of pulse duration with current can be expected within this area.

32.4.2.2 The Absorber Relaxation Time

The dependence of the pulse duration on τ_α (and this on the absorber bias) within the stability range is shown in Figure 32.12. As seen in the figure, to achieve stable ML, the absorber relaxation time needs to be



(a) (b)

FIGURE 32.11 (a) Simulated dependence of the pulse amplitude on the pumping current (bifurcation diagram) for the laser of Figure 32.10b for one value of absorber recovery time, calculated with a phenomenological TWM. (b) Calculated pulse duration for the same laser within the current ranges for stable ML.

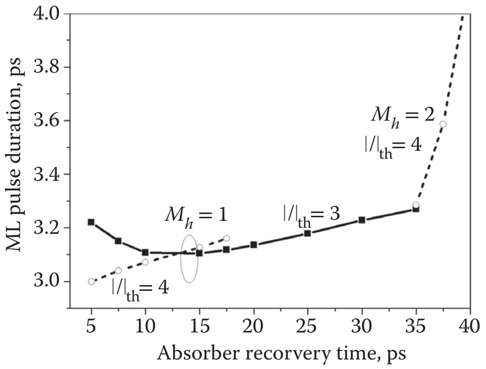


FIGURE 32.12 Typical simulated dependence of pulse duration on the absorber recovery time for two current values: simulations using a phenomenological TMW. Laser design same as in Figure 32.10b.

within a certain range. Values of τ_α above a certain value produce instabilities of either leading or trailing edge type, and may mean switching to harmonic ML. Within the stable ML range, a decrease in τ_α tends to shorten the pulses, due to both the effects of partial absorber relaxation during the pulse and, probably more significantly, to the fact that the slow relaxation of the absorber leads to the absorber being always partially saturated, thus reducing the initial absorption Q_-. This trend can sometimes be reversed for *very* short τ_α, of the order of a few picoseconds, when the absorber may not be saturated efficiently; therefore, there may be an optimum τ_α for shortest pulse generation, though for some laser parameter values this optimum τ_α may be so small as to be technologically unattainable.

The requirement of small τ_α in ML has been recognized from the early days. The methods of reducing τ_α in experiments included, first, ion implantation in early work (Deryagin et al., 1994; Zarrabi et al., 1991), second, choosing QW materials for faster sweepout, e.g., AlGaInAs rather than InGaAsP quaternaries (Green et al., 2011; Hou et al., 2010a,b, 2013) or even QD materials in which SA dynamics is known to be fast (Erneux et al., 2009; Rafailov et al., 2011; Viktorov et al., 2009), third, engineering the QW profile to include steps or oblique rather than vertical walls (Nikolaev and Avrutin, 2004), and fourth, using unitraveling carrier absorbers (Scollo et al., 2005, 2009); see Avrutin and Rafailov (2012) for more detail.

32.4.2.3 The s-Factor, or the Absorber to Gain Saturation Energy Ratio

While it is well known that the increase in s (determined by Equation 32.17) facilitates ML, it is not immediately intuitively clear whether an increased s helps ML stability, as it is known (Kuznetsov, 1985) that the passive Q-switching regime, which is one of the instabilities affecting ML, is also facilitated by an increase in s. However, the results from both DDE and traveling-wave simulations (Bandelow et al., 2006) show that in fact it is the stable ML range that is increased with s at the expense of the Q-switching (or self-pulsing) range.

In VECSELs, the most straightforward way of increasing the s parameter is manipulating the spot ratio of the gain module and the SESAM. For instance, the fast absorption saturation, or the high ratio s, was crucial in achieving the bistability between the off-state and the ML state necessary for low-repetition rate ML and addressable pulse generation studied in Marconi et al. (2014, 2015); this was achieved by placing the SESAM in the Fourier plane of the focusing lens so that the spot area $A_{X\alpha}$ was determined only by the diffraction limit. In *monolithic* diode lasers, a similar strategy can be pursued to some degree by *tapering* the laser waveguide so that the absorber region is narrower than the gain region. Used originally for Q-switching QW lasers in a “bow-tie” construction with the narrow SA in the middle of the cavity and the two amplifier sections tapering outward (Williams et al., 1994), this strategy has later been realized most convincingly in QD mode-locked lasers (Nikitichev et al., 2011a,b; Thompson et al., 2006, 2009); see Avrutin and Rafailov (2012) for an overview of results achieved. More often, monolithic structures use the same waveguide structure in the gain and absorber sections for ease of fabrication and fiber coupling, and so $A_{X\alpha} = A_{Xg}$. The parameter s is then equal to the *absorber-to-gain cross section ratio* $s = \sigma_a/\sigma_g$. QWs, with their sublinear (approximately logarithmic) dependence of gain on carrier density (population inversion) have long been seen as superior to bulk material for ML performance, since the sublinear $g(N)$ helps achieve $\sigma_a/\sigma_g > 1$. Vladimirov et al. (2004) and Vladimirov and Turaev (2005) used their DDE analysis to conclude further that when designing a QW laser for ML purposes, a structure with a *smaller number* of QWs was preferable to one with a larger number—indeed, the smaller number of QWs means a smaller confinement factor, hence a higher threshold carrier density and therefore a smaller dg/dN at threshold due to the sublinear $g(N)$, which in turns gives a higher value of the ratio s. These considerations influenced the choice of structures with just two to three QWs for realizing DBR ML lasers capable of generating very stable pulses about 2-ps long at 40 Gb/s (Bandelow et al., 2006). It may be argued that the same logic also, in part, accounts for the success of QD mode-locked lasers, in which the dependence of gain on the (total) carrier density in the active layer is even more sublinear than in QWs; however, it has to be borne in mind that the concept of total carrier density is somewhat misleading in QDs; a more accurate picture is given by more complex analysis, considering separately the population of the dots themselves and of the reservoir that supplies them with carriers.

32.4.2.4 Gain and Group Velocity Dispersion Parameters

Most models of mode-locked laser operation predict that without *gain* dispersion, stable ML with a finite pulse duration is impossible, and so the gain dispersion, or width of gain curve, represented by the parameter γ in the DDE or ω_L in the TWM, should play an important role in determining the pulsewidth and stability. Within the range of gain dispersion typical in mode-locked semiconductor lasers, which usually corresponds to $\hbar\omega_L$ of the order of tens of meV (or the wavelength range of tens of nm) and does not change too much with operating conditions or construction, gain dispersion is not the most drastic factor limiting the pulsewidth. However, achieving a broad gain spectrum is still desirable. This may be one of the advantages of QD active media, as discussed below.

GVD, like gain dispersion, acts to broaden the pulses in the case of normal dispersion, which is the usual situation in semiconductor lasers. As discussed above, the effect of this parameter is modest in most semiconductor lasers since the pulse durations at which it would become important (~100 fs) are never achieved; however, with stronger GVD possible in QD lasers, some account for this effect may be necessary.

32.4.2.5 The Gain and Absorber Compression Coefficients

Pulses generated by ML lasers tend to be of picosecond duration. This is below the critical pulsewidth at which the fast gain saturation, rather than the average carrier density dynamics, begins to dominate the pulse amplification and shaping, at least in the gain section; this critical pulse duration has been estimated (Mecozzi and Mork, 1997; Mork and Mecozzi, 1997) to be of the order of $\frac{\varepsilon_g}{v_g \frac{dg}{dN}}$. With typical semiconductor parameters, this estimate gives values of the order of 10 ps. Thus, the gain compression (and, similarly, absorber compression) effects and the coefficients that describe them (if introduced) may be expected to play a significant part in ML properties.

In practice, the effect of nonlinearities is twofold. First, gain compression tends to broaden ML pulses, with absorption compression having the opposite effect. Second, and in some regards more important, an increase in gain compression *stabilizes ML operation*, suppressing the *Q*-switching instability (the latter can be easily shown by rate equation analysis of Q-switched lasers (Avrutin et al., 1991). Again, fast absorber saturation has the opposite effect.

32.4.2.6 The Self-Phase Modulation in the Gain and Absorber Sections

Within the simplest phenomenological approach, the linewidth enhancement factors have a relatively modest effect on pulse *energy* for a given current and absorption, but a more noticeable one on amplitude and duration. They do not significantly affect the onset of the *Q*-switching instability (the lower current or unsaturated gain limit of ML stability), but have a stronger effect on the upper limit of ML stability associated with the irregular envelope and pulse competition. Mode-locked behavior is the most stable when the gain and absorber linewidth enhancement factors are not too different from each other. According to the DDE model predictions, the most stable operating point (which also corresponds to the highest pulse amplitude and lowest duration) is for $\alpha_{Hg} = \alpha_{H\alpha}$. Traveling-wave and modal analysis predicts that the best-quality ML is achieved with $\alpha_{Hg} > \alpha_{H\alpha}$, see, e.g., Salvatore et al. (1996); the discrepancy is likely to be caused by the different geometry of the long amplifier and the shorter absorber. The main parameter determined by the linewidth enhancement factors is the *chirp* (dynamic shift of the instantaneous frequency) of the pulse. Passively mode-locked pulses tend to be up-chirped (with the instantaneous optical frequency increasing toward the end of the pulse) when the absorber saturation factor $\alpha_{H\alpha}$ is small and the chirp is mainly caused by α_{Hg}. With a certain combination of α_{Hg} and $\alpha_{H\alpha}$ (typically $\alpha_{Hg} > \alpha_{H\alpha}$), an almost complete compensation of chirp is possible; with $\alpha_{H\alpha} > \alpha_{Hg}$, the pulse is typically down-chirped (Salvatore et al., 1996). As up-chirp is observed more frequently than down-chirp in experiments, one may conclude that typical values of $\alpha_{H\alpha}$ are smaller than α_{Hg}. In active ML, *down-chirp* is typically observed, while hybrid ML allows the chirp to be tuned to some extent, and there is typically a combination of bias and current or voltage modulation amplitude for which the chirp is minimized and close to zero, if only in a very narrow range of operating parameters. In a more detailed semimicroscopic model discussed in Section 32.5, the linewidth enhancement factor is not introduced *explicitly*, with gain and refractive index correction being implemented as imaginary and real part of the same complex dielectric permittivity, respectively, but the tendencies would appear to be general enough to merit being taken into account.

32.4.2.7 The Laser Geometry and Saturable Absorber Location

All the results discussed above could be obtained using either DDE or TWMs, though the latter usually give slightly more realistic predictions (e.g., both types of models predict harmonic operation at high currents, but in the case of multi-GHz PML ($T_{rt} \ll \tau_g$), operation at harmonics above the second one, predicted by the DDE model, is usually not observed experimentally; TWMs tend to predict only fundamental and second harmonic operation, which agrees with the experiment (Bandelow et al., 2006). The effects of a realistic laser cavity geometry (Fabry–Pérot versus ring cavity, absorber position and length, facet reflectances, etc.) on the ML characteristics are one area where the TWMs have an obvious advantage. They are, for example, highly suitable for analyzing the effects of specialized *harmonic* ML designs for high-frequency generation. These fall into two categories.

The first is colliding pulse mode-locking (CPM), including *multiple (MCPM)* (Martins et al., 1995; McDougall et al., 1997) and *asymmetric (ACPM)* (Shimizu et al., 1995, 1997) colliding-pulse ML constructions. These achieve ML at the M_h-th harmonic by positioning one SA (in CPM or ACPM) or several SAs (in MCPM) at fraction(s) M'_h/M_h of the laser cavity length, where $M'_\text{h} < M_\text{h}$ is an integer, and M' (or at least some of the values of M'_h in case of MCPM) and M are mutually prime. The standard CPM corresponds to $M_\text{h} = 2$ (and obviously $M'_\text{h} = 1$) with the SA in the center of the cavity. Constructions of this type have produced ML operation at rates of up to 860 GHz (Shimizu et al., 1997) (in that particular case, $\lambda = 1.55$ μm ACPM construction used $M'_\text{h} = 5, M_\text{h} = 12$). A more detailed review is given in Avrutin et al. (2000). The alternative harmonic ML technique is the use of a *spectrally selective laser cavity*. This may be in the form of a *compound cavity*, which has one or several *intracavity reflectors* (ICRs) positioned at fractions M'_h/M_h of the laser cavity length, rather like the SAs in the MCPM or ACPM technique. The highest ML repetition rates reported to date have been achieved by GaAs/AlGaAs ($\lambda \approx 0.89$ μm) lasers with the ICRs in the shape of deeply etched slots, either single or multiple in a 1-D photonic-bandgap (PBG) mirror arrangement (Yanson et al., 2002). A 608-μm long cavity used ICRs at 1/33 of the cavity lengths, giving $F = 2.1$ THz (similar structures with lower M_h values produced ML at bit rates of the order of hundreds of GHz depending on M_h and L).

TWMs have been long since been used for detailed analysis of the CPM (highlighting the importance of *incoherent* colliding pulse effect, the fact that the SA is saturated simultaneously by pulses propagating in both directions, as described by Equation 32.48, as opposed to coherent CPM effect, the self-induced grating described by Equation 32.46, which in semiconductors is less important than in other active media due to diffusion smoothing of the grating). All studies also predict higher stability and shorter pulses with CPM than with ordinary PML, in good agreement with experiment. A version of TWM was also used to explain and analyze the first realization of MCPM (Martins et al., 1995).

In terms of compound-cavity harmonic ML, TWM can help analyze the effects of the slot reflectances and positions (Yanson et al., 2002; Hou et al., 2010b, 2013) and bias current (Hou et al., 2013), with very good qualitative, and good quantitative, agreement with experiments. A TWM study was also used to propose and analyze the use of an alternative method of harmonic selectivity in a mode-locked laser resonator, a sampled grating reflector (Kim et al., 1999), later successfully realized experimentally, see, e.g., Hou et al. (2014) and references therein.

In principle, qualitative analysis can also be possible with DDE models (in fact, CPM can be analyzed, and ACPM predicted, in VECSEL structures using a DDE model (Avrutin and Panajotov, in preparation)), but this requires a customized model for each design whereas TWMs adapt to new design through a simple change of parameters. Harmonic ML is also one area where the time-frequency domain approaches (Section 32.3.2), despite their limitations, can give some insight into the cavity geometry required for successful harmonic operation (Martins et al., 1995; Yanson et al., 2002); however, TWM is still preferable for full quantitative description of the regime.

In addition to CPM, TWM modeling was also used to highlight another possibility allowing for a better saturation of the SA, which consists in placing the absorber close to antireflection-coated laser facet. This design is termed anticolliding mode-locking (ACML) and is inverse to that used for solid-state lasers where the absorber is usually placed close to a high reflection mirror, leading to self-colliding pulse mode-locking (SCML). In this latter situation, the pulse is allowed to interfere constructively with its own reflection onto the mirror leading to an improved absorber bleaching. For semiconductor laser diodes, however, such interferometric effects are rather weak due to the large value of the carrier diffusion coefficient which, in turn, reduces the coherent population grating created by the pulse self-interference as illustrated by Equation 32.47. The analysis of the ACML (Javaloyes and Balle, 2011) showed that this design leads to a substantial increase in output power, a reduction in amplitude and timing jitter, and an enlargement of the range of currents where stable PML can be obtained. The reason is a consequence of the strong increase of the laser field along the cavity axis and as it propagates toward the anti-reflection facet, yielding a more intense saturation of the absorber, and comparatively a weaker saturation of the gain. These regimes where the field is widely nonuniform along the cavity is one of the cases where a proper analysis demands using a TWM approach.

TWM was also used for analysis and optimization of the geometry of a number of other ML laser designs. Simulations of long-cavity ($F_{\text{rep}} \sim 10$ GHz) actively, passively, and hybridly mode-locked DBR lasers predicted that the optimal hybrid and PML performance could be achieved by placing the SA near the facet rather than at the DBR (Hasler et al., 2005), probably largely due to the more efficient absorber saturation as discussed above (Javaloyes and Balle, 2011). TWM studies were also used to predict, in agreement with experiment, that extended cavity laser designs containing a passive section were preferable to all-active ones in terms of pulse duration and chirp (Camacho et al., 1997).

32.5 Microscopic and Semimicroscopic Approaches in Mode-Locked Laser Modeling

32.5.1 The Basics of Microscopic Input in Mode-Locked Laser Simulations

As discussed above, phenomenological models of ML, either of delay-differential or traveling wave types, can be successfully used to predict a number of qualitative tendencies and, to a degree, qualitative parameters of the ML process. However, for detailed comparison with experiments, some microscopic input regarding the gain and SA section material is highly desirable. Some attempts at adding elements of microscopic analysis to models of the type described in the previous section has been undertaken for some time, see, e.g., Bischoff et al. (1997) for a relatively early example. The main challenge, and the main need for microscopic input, arguably, is the accurate modeling of interplay of the *spectral and temporal properties* of the material and laser dynamics. Most importantly, in most QW and, in some cases, QD lasers, the operating wavelength is at the sharp edge of the SA absorption spectrum. As the absorber is saturated, the absorption edge shifts, which lead to a change in the operating point, variation in the absorber saturation parameters affects self-phase modulation, etc. Likewise, due to the absorber selectivity, the operating point can be detuned from the gain peak in the amplifier section. In particular, this makes it difficult to construct a simulated map of regimes in the plane of "gain section bias current/absorber bias voltage," since the voltage affects simultaneously the absorber recovery time (relatively easy to model semimicroscopically as discussed below) and, more importantly, the spectral position of the absorption edge hence the unsaturated absorption and the cross section. This complex interplay of effects is difficult to capture with Equation 32.42 or any of its time-domain implementations, for the SA in particular, and to a degree for the gain section too. A more accurate (and thus ideally microscopically informed) model of gain and, particularly, saturable absorption spectrum as function of population (carrier density), and ideally temperature, is thus required.

The difficulty with this problem is that a realistic microscopic evaluation of gain/absorption in semiconductors is a formidable problem. To do so rigorously requires a many-body quantum mechanical approach (Chow and Koch, 1999), which is a task so complex from the point of view of the various physical effects at work, and so computationally resource-intensive that, at the time of writing, only a handful of research teams worldwide have both the expertise and the computer capacity to attempt it, and to our knowledge only one (the partnership between the Philipps-Universität Marburg and the University of Arizona) is performing this task routinely. The approach has been applied to some mode-locked laser problems, notably single-pass pulse modifications (Bottge et al., 2014) and steady-state ML pulse formation (Kilen et al., 2016) in ultrashort-pulse mode-locked VECSELs, and has informed some work on edge-emitting ML lasers, but to the best of our knowledge not yet been integrated into full analysis of high-bit-rate ML dynamics.

Even in the simpler single-particle approximation ignoring many-body effects, microscopic evaluation of the gain/absorption spectrum in a semiconductor material is, in general, rather nontrivial. For the case of a QW, or, with some reservations, QD active medium, the gain in the single-particle approach is calculated as

$$g(\hbar\omega) = A \sum_{\mu,\nu} \int \left| M^2_{\mu\nu} \right| (E_{\mathrm{e-h}}) \rho_{\mu\nu}(E_{\mathrm{e-h}}) \left(f_{\mathrm{e}}(E_{\mathrm{e}}^{(\mu)}) + f_h(E_{\mathrm{h}}^{(\nu)}) - 1 \right) L' \left(\hbar\omega - E_{\mathrm{e-h}} \right) \mathrm{d}E_{\mathrm{e-h}}. \tag{32.49}$$

Here, A is a proportionality coefficient, $E_{\mathrm{e-h}}$ is the electron–hole energy separation, $E_{\mathrm{e}}^{(\mu)}$ and $E_{\mathrm{h}}^{(\nu)}$ are the energies of the electron in the conduction subband (in the case of a QW material) or level (in the case of QD material) μ and the hole in the valence subband/level ν, which are fully defined by the subband/level numbers and energy separation. Furthermore, $M^2_{\mu\nu}$ and $\rho_{\mu\nu}$ are the squared matrix element of the transition and the reduced density of states, respectively (each of which, in general, varies with energy and is, in general, calculated from a transcendental equation in the case of nonparabolic QW subbands), f_{e} and f_{h} are the electron and hole distribution functions (which in the case of QW materials are very close to Fermi distributions given by the carrier density N and temperature T, and for QDs, in general, need to be calculated from separate kinetic equations), and L'' is the bell-shaped linewidth broadening function.

For a bulk semiconductor material (rarely used in mode-locked diode lasers), the summation over μ and ν is not necessary as there are no multiple subbands in the valence and conduction bands, and the reduced density of states and the matrix element are simple analytical functions of energy.

Still, the expression for the gain/absorption spectrum, in general, can only be evaluated numerically in all cases. The refractive index modulation, needed to consider self-phase modulation in mode-locked lasers, is then reconstructed using Kramers–Kronig relations.

The results are rather difficult to parametrize and to implement in a time-domain model. Simple linear or logarithmic approximations of the microscopic results, often used in the phenomenological models described in the previous section, capture the carrier density dependence of the gain peak and the saturated absorption, but not the spectrum.

The first step toward simplifying the problem is to approximate the linewidth broadening function L'' (which in the general case of a non-Markoffian phase relaxation characteristic of a highly populated semiconductor has quite a complex shape) by a simple Lorentzian form, which follows from a density matrix analysis with a Markoffian phase relaxation with a constant decay rate γ_T and is thus essentially identical to the function describing homogeneous broadening in a two-level system (Equation 32.42):

$$L'(\Delta\omega) = \mathrm{Re}L(\Delta\omega) = \frac{1}{\pi} \frac{\gamma_T}{\gamma_T^2 + \Delta\omega^2}; L(\Delta\omega) = \frac{1}{\pi} \frac{1}{\gamma_T - i\Delta\omega}. \tag{32.50}$$

This form of L' is advantageous because the dynamic correction to the refractive index or the real part of the dielectric permittivity (“imaginary gain”) is then calculated using the same equation as Equation 32.49 but using the complementary linewidth broadening function $L''(\Delta\omega) = ImL(\Delta\omega) = \frac{1}{\pi} \frac{\Delta\omega}{\gamma_T^2 + \Delta\omega^2}$, eliminating the need for Kramers–Kronig transformation and allowing both gain and refractive index variation to be implemented in time domain as discussed below.

Two main routes can then be successfully followed for using the results of Equation 32.49 in time-domain analysis required for ML simulations.

The first one consists in using a number of approximations in order to simplify the expression (Equation 32.49) (and its imaginary counterpart for the refractive index variation) to a level that permits the integration in energy to be performed analytically, and thus can be realistically parametrized as a (analytical) function of carrier density N and photon energy $\hbar\omega$. This has been achieved (Balle, 1998; Javaloyes and Balle, 2010a) by considering a QW material with a single, parabolic subband for both electrons and holes (hence constant reduced density of states), ignoring the energy dependence of M^2 (and γ_T) and considering (at least in the original version of the model) a very low (mathematically, zero) temperature. In the *frequency* domain (Balle, 1998), this results in a complex dielectric permittivity correction (with the refractive index correction as the real part and the gain as the imaginary part) in the form

$$\chi(\omega, N) = -\chi_0 \left[2\ln\left(1 - \frac{D}{u-i}\right) - \ln\left(1 - \frac{b}{u-i}\right)\right]. \tag{32.51}$$

Here, χ_0 is a constant proportional to the matrix element M^2 (Balle, 1998) (which in practice can be used as a fitting parameter in the model), $D = \frac{\pi d_{\mathrm{QW}}\hbar}{m\gamma_T}N$ is the reduced carrier density, d_{QW} and $m = \left(m_{\mathrm{e}}^{-1} + m_{\mathrm{h}}^{-1}\right)^{-1} \approx m_{\mathrm{e}}$ being the QW thickness and the reduced mass of the electron–hole pair; $b = \frac{\hbar k_m^2}{2m\gamma_T}$, k_m being the maximum wave vector in the first Brillouin zone of the semiconductor crystal; and the frequency of the optical transition is parameterized as $u = \frac{\hbar\omega - E_{\mathrm{g}}}{\hbar\gamma_T}$, where the bandgap E_{g} includes N- and T-dependent renormalization.

This result was used, among other applications, in the work in DDE modeling of mode-locked VECSELs (Mulet and Balle, 2005) discussed in Section 32.3.3. In such lasers, the operating frequency and the spectral selectivity of the laser cavity are mainly determined by the resonator properties of the gain chip. A full time-domain implementation of the gain spectrum was thus not required; instead, the complex dielectric permittivity was linearized around the spectral point of operation which, as mentioned in Section 32.3.3, was sufficient to achieve very good qualitative, and good quantitative, agreement with previously reported experiments.

In time domain, the general form of polarization has the same form as Equation 32.44

$$P_{\pm} = \int_0^{\infty} \chi(\tau)Y_{\pm}(t-\tau)\mathrm{d}\tau. \tag{32.52}$$

The time-domain version of Equation 32.51 results (Javaloyes and Balle, 2010a) in an integration kernel more complex than the simple exponential of Equation 32.44, specifically of the form

$$\chi_{\mathrm{g}}(\tau, N) = \chi_0 e^{-\left(\gamma_T + i\left(\omega_{\mathrm{g}} - \omega_{ref}\right)\right)\tau}\frac{2e^{-i\gamma_T D\tau} - 1 - e^{-i\gamma_T b\tau}}{\tau}. \tag{32.53}$$

Here, $\omega_{\mathrm{g}} = E_{\mathrm{g}}/\hbar$; the rest of parameters are the same as in Equation 32.51 and the normalized carrier density D is, strictly speaking, evaluated at the time $t - \tau$; however, since the carrier density even in the faster recovering SA, let alone the gain section, does not change noticeably on the timescale of $1/\gamma_{\mathrm{T}}$, using the value at time t is usually accurate enough (Javaloyes and Balle, 2010a). Despite being, strictly speaking, only applicable for very low temperatures such that $\kappa_{\mathrm{B}}T \ll \hbar\gamma_T$, the kernel (Equation 32.53) is a reasonable approximation for the highly degenerate gain medium. Its accuracy can be further improved to include the finite temperature, though this results in a more complex expression including special functions.

For the case of SAs, an alternative kernel expression was derived (Stolarz et al., 2011) with a finite temperature taken into account from the start, but with the Fermi distribution functions substituted by Boltzmann exponentials, which is usually a safe approximation for the weakly degenerate SA material; in this case,

$$\chi_{\alpha}(\tau, N) = \chi_0 e^{-\left(\gamma_T + i\left(\omega_{\mathrm{g}} - \omega_{\mathrm{ref}}\right)\right)\tau}\left\{D\left[\frac{a_c}{a_c + i\gamma_T\tau} + \frac{a_v}{a_v + i\gamma_T\tau}\right] - \frac{1 - e^{-i\gamma_T b\tau}}{i\gamma_T\tau}\right\}, \tag{32.54}$$

where $a_{c,v} = \frac{m}{m_{c,v}}\frac{\hbar\gamma_T}{\kappa_B T}$ and the rest of the parameters are as in Equation 32.53. Unlike the simple purely Lorentzian kernel of Equation 32.44, neither of the functions (Equations 32.53 and 32.54) allow for a simple iterative solution using only the previous point in time; instead, the integral must in practice go to an integration limit in τ of about $\tau_{\text{lim}} \sim (3\text{–}6)/\gamma_{\text{T}}$, making the number of previous points necessary for the integration $\tau_{\text{lim}}/\Delta t$.

Despite being managed/mitigated in part by the efficient numerical implementation (Javaloyes and Balle, 2010a), this approach is still quite taxing on the computer resources if implemented straightforwardly in a TWM, since the need for numerical integration in the dispersion operator adds to the general limitations of the TWM approach (the stiffness of the problem requiring a small time step, and computational timescaling as $1/\Delta t^2$ due to the $\Delta z = v_{\text{g}}\Delta t$ condition). Its efficiency can, however, be improved drastically with the *decimation* technique discussed in Section 32.5.2.

The semimicroscopical approach in this form was successfully used to model a number of mode-locked laser designs, including a "straightforward" two-section laser with an SA at a facet (Javaloyes and Balle, 2010b), a colliding-pulse mode-locked laser (Tandoi et al., 2013), and a specialist ring laser with an intracavity filter for spectral flattening and pulse duration reduction (Moskalenko et al., 2013); we shall consider the results of Tandoi et al. (2013) here in somewhat more detail as an illustration of the possibilities and certain limitations of the model.

While the quantitative agreement with experimental results is still not entirely accurate, microscopic models of this type have allowed a number of new possibilities compared to purely phenomenological approaches. The first of these is a much more meaningful, and directly comparable with experiments, prediction of the map of regimes in the "gain section current/absorber bias" plane. Indeed, in QW materials, the voltage applied to the SA has a double effect. First, it decreases the absorber recombination time. To a good accuracy, this is described with an exponential dependence, e.g., in Tandoi et al. (2013), an expression

$$\tau_{\text{SA}} \approx \tau_0 \exp[-F/(f F_0)] \tag{32.55}$$

was used, with $F_0 = (\kappa_{\text{B}} T)/(ed_{\text{QW}})$ being the activation field predicted by the simple thermionic excitation theory (Cavailles et al., 1992), and a correction factor f is introduced heuristically to take into account the interplay between tunneling and thermionic processes (and possibly any well profile distortion). The values $f = 0.5$ and $\tau_0 = 50$ ps were given by fitting the measured $\tau(V_0)$ but agreed well with estimates from more sophisticated theory (Nikolaev and Avrutin, 2003). Second, in materials where the quantum-confined Stark effect (QCSE, absorption shift with voltage) is significant, a voltage variation changes the bandgap, thus affecting the operating point of the laser (unsaturated absorption and saturation cross section), which the microscopic model successfully captures. This tends to be more pronounced in InGaAsP than in aluminum-containing quaternaries. Strictly speaking, QCSE affects also the shape of the absorption spectrum, but this is as yet to be included in the model, since to the best of our knowledge there is as yet no accurate calculation of absorption in a QW with both carriers and electric field present.

Finally, analysis of CW ML behavior requires, in addition to models of absorption spectrum and the recovery time, also a model of the current heating in the absorber. Indeed, the material bandgap that features in the kernels (Equations 32.53 and 32.54) is the renormalized temperature- (and carrier density-) dependent one, and the absorber can get significantly (up to about 10°) heated by the photocurrent flowing through it. This can be modeled (Tandoi et al., 2013) using a separate equation for the absorber bandgap:

$$\frac{\mathrm{d}E_{\text{g}}^{(\text{SA})}}{\mathrm{d}t} = -\gamma_{\text{therm}}\left(E_{\text{g}}^{(\text{SA})} - E_{\text{g0}}^{(\text{SA})} + 2\pi R_{\text{th}} j_{\text{SA}}\left(\left|V_{\text{SA}}\right| + V_{\text{bi}}\right)\right). \tag{32.56}$$

Here, R_{th} is the thermal resistance of the SA, the photocurrent density can be estimated as $j_{\text{SA}} = eN_{\text{SA}}d_{\text{QW}}/\tau_{\text{SA}}$, and the voltage seen by the active region in the SA is a sum of the built-in potential drop $V_{\text{bi}} = E_{\text{g0}}^{(\text{SA})}/e$ and the applied reverse bias $\left|V_{\text{SA}}\right|$. The parameters of the model were fitted in Tandoi et al. (2013) with the experimental measurements, and then the thermal relaxation rate γ_{therm}, which in reality is of the order of inverse microseconds, sped up to 10^8 s^{-1}, still slower than all characteristic times of laser dynamics but within the possibilities of the model.

Several features observed from the experimental results were well reproduced by the model, e.g., the optimal pulses around 600 fs long were found in the case of a SA occupying 4% of the cavity; some deviation from the best experimental results (430 fs) can be attributed to the approximate nature of the fitting of the gain curve with the two-band model (Equation 32.53).

With the microscopically informed gain and SA models as well as the voltage dependence on the SA recovery time, we constructed in Tandoi et al. (2013) maps of major characteristics of laser behavior in the "amplifier section current—absorber voltage" plane (Figure 32.13). The evolution of the ML quality as a function of the current followed the predictions of the DDE and TWM theories, with a minimal bias current to obtain the ML. The inspection of the time traces indicates that the degradation at high currents is due to the leading-edge instability and to the competition with other harmonics ML solutions, where three or four pulses propagate within the cavity. The evolution of the dynamical regimes as a function of the reverse bias was more subtle. In the AlGaAs materials studied in Tandoi et al. (2013), QCSE was weak and only the SA recovery time varied. This was, however, sufficient to capture the essential physical effects at work and to reproduce the output power decrease and the photocurrent density increase with V_{SA}. A faster SA has enough time to recover its full absorption between pulses and, therefore, presents the full amount of its unsaturated absorption to the pulse, which allows increased losses and photocurrent. Additionally, the photocurrent generated in the SA induces a detuning of the SA bandgap, which eventually hinders ML and decreases the output power. This is due to the fact that the SA can efficiently be modulated only within a limited spectral region around the bandgap and that the absorption increases on the blue side of the spectrum.

The alternative approach to including microscopically informed gain and absorption spectra into a mode-locked laser model has been so far implemented mainly in the context of QD lasers. In QD media, carriers are localized in individual dots and so the global carrier–carrier collisions that establish a common quasi Fermi level in a bulk or QW material are not present. QD lasers still show inhomogeneous

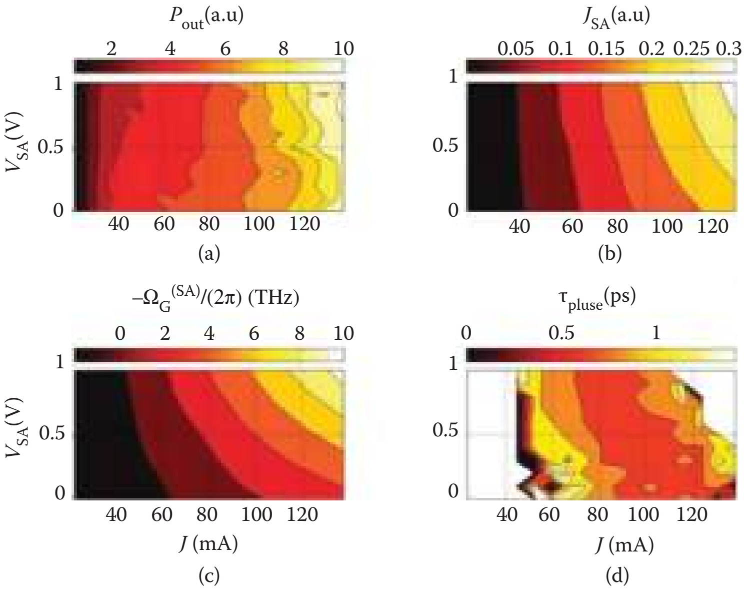


FIGURE 32.13 Parameter maps as a function of the SA reverse bias and of the gain section currently calculated using a TMW with decimation and analytical gain dispersion spectrum, for (a) output power P_{out}, (b) photocurrent density j_{SA} flowing out of the SA section, (c) bandgap of the SA section $-E_g^{(SA)}/h$, and (d) pulsewidth τ_{pulse}. The SA was 4% of the cavity length (optimal case, in agreement with measurements). (From Tandoi G. et al., *IEEE J Sel Top Quantum Electron*, 19, 1100608, 2013. Reproduced with permission).

broadening described mathematically by an equivalent of the reduced density of states in Equation 32.49, but the nature of this broadening is different from the case of bulk or QW materials and stems, not from a continuous energy spectrum of electrons and holes at each point in the material, but from the presence of a large number of dots, with fluctuating dot size and composition, and thus energy levels, in any microscopically large material sample, as already mentioned in Section 32.3.3. The Fermi energy distribution among the carriers occupying the multiple energy levels is established, not by the fast carrier–carrier collisions as in bulk and QW, but by slower processes of capture and escape and is thus only approximate at room temperatures and nonexistent at cryogenic temperatures. The gain thus cannot always be parameterized as a function of the total carrier density N. Instead, the populations of multiple individual levels need to be determined dynamically by solving rate equations including the kinetic processes of capture, interlevel relaxation, and escape, as well as spontaneous and stimulated recombination (see Rafailov et al., 2011 and the corresponding chapter in this handbook for a detailed description). In this case, trying to obtain an integral in Equation 32.49 analytically is problematic (some approximate methods have been discussed in Rafailov et al. (2011) for the room temperature case, but their accuracy has not been tested). Instead, a number of dynamic variables representing population of levels, or slices of the energy spectrum of the carriers in the material (different values of $E_{\text{e-h}}$ in Equation 32.49), at each point in time and space are introduced; the problem thus becomes not two, but (at least) three-dimensional: time, space, and transition energy. To reproduce a spectrum faithfully, the width of the slice should be smaller than, or at least comparable to, the homogeneous broadening γ_T. This approach to QD ML laser modeling is much more complex than the simpler one ignoring the inhomogeneous broadening as presented in Viktorov et al. (2006), but is also much more accurate. Its computational efficiency is comparable to that of using the analytical dielectric permittivity (Equations 32.52 through 32.54). Indeed, on the one hand, the number of variables characterizing carrier polarization (gain and refractive index modulation) at each point is not just the carrier density but the populations of the individual levels, but on the other hand, polarization relaxation of each individual level is Lorentzian (Equation 32.50), so an easy and efficient integration kernel of the type of Equation 32.44 can be implemented.

The resulting model has been used for mapping stability limits in a mode-locked laser and analysis of pulse shape and chirp (Rossetti et al., 2011a,b), as well as the analysis of the effects of laser cavity geometry (absorber location, cavity reflectivities) (Simos et al., 2013; Xu and Montrosset, 2013), in particular finding the optimum SA length (Xu and Montrosset, 2013).

The microscopic approach of this type can be implemented both in the traveling-wave formalism as in the papers above (particularly with decimation to improve the otherwise problematic numerical efficiency) and in a DDE model, particularly the multisection version (MS-DDE) as discussed in Section 32.3.3 (Rossetti et al., 2011b). With inhomogeneous broadening included, an MS-DDE model was used for the analysis of problems including the peculiarities of ML operation involving both ground- and excited-level transitions (Xu et al., 2012), as well as for modeling-tapered laser structures (Nikitichev et al., 2012), the latter in agreement with both a TWM simulation by the same authors and with experiments.

In Pimenov and Vladimirov (2014), a multilevel approach was presented for a TWM of a "generic" inhomogeneously broadened medium, not necessarily QD. It can be foreseen that potentially, it can be used for lasers with QW and bulk active layers, for example for a short-cavity QW laser where transitions from both first and second electron levels (disregarded in Equations 32.52 through 32.54) can become important. In principle, it could allow an arbitrary set of carrier density-dependent spectra, possibly calculated using a many-body approach and/or using realistic QW parameters such as multiple-level transitions, strain, etc., by necessity omitted in Equations 32.52 through 32.54, to be approximated and reproduced in the time-domain model. To a degree, the analytical spectrum approach (Equations 32.52 through 32.54) offers such an approximation possibility as well by treating the effective masses of carriers and matrix elements as fitting parameters, which has the advantages of involving much fewer parameters, but may be somewhat less flexible.

32.5.2 Improving the Numerical Efficiency of TWMs: Decimation/Space–Time Folding

So far, we have considered, on the one hand, the DDE models (Section 32.3.3) which are simple, efficient, instructive, and allow automated bifurcation analysis by numerical continuation and in the simplest cases also analytical insight, but are not always accurate in reproducing the behavior or realistic edge-emitting laser constructions, and, on the other hand, full TWM models (Section 32.3.4) which are potentially very accurate in representing a realistic laser, but rather computationally intensive and incompatible with automated bifurcation analysis.

The numerical technique of decimation presented below, which builds on an earlier model for laser designs where gain and SA dispersion are negligible compared to the effect of a lumped dispersive element (reflector) such as a DBR (Vladimirov et al., 2009) to apply to the general case of significant gain and SA dispersion (Javaloyes and Balle, 2012b), allows the advantages of both models to be combined to a large degree.

The starting point for the technique is a TWM in the general form of Equation 32.40, which can be compactly rewritten (keeping only the deterministic part so far) as

$$\pm\frac{\partial Y_{\pm}}{\partial z}+\frac{1}{v_{\mathrm{g}}}\frac{\partial Y_{\pm}}{\partial t}=Z_{\pm}=P_{\pm}^{t}-\alpha_{\mathrm{f}}Y_{\pm}, \tag{32.57}$$

where $\alpha_{\mathrm{f}}=\alpha_{int}/2$ is the "field" internal loss, and the "total" polarization $P_{\pm}^{t}$ includes the dynamic coupling terms:

$$P_{\pm}^{t}=\left(\frac{1}{2}\hat{g}_{\mathrm{mod}}+ik_{\mathrm{ref}}\widehat{\Delta\eta}_{\mathrm{mod}}\right)Y_{\pm}+i\hat{K}_{\pm,\mp}Y_{\mp}. \tag{32.58}$$

(Note that the dynamic coupling terms can include dispersion in the same way as the "copropagating" polarization terms, as is indeed the case in the semimicroscopic model used in Javaloyes and Balle (2010a,b), though this is not likely to have a major effect as the dynamic coupling is usually weak anyway.) The formal solution of Equation 32.51 is obtained in the form

$$Y_{\pm}(z,t)=Y_{\pm}(z\mp v_{\mathrm{g}}\tau,t-\tau)+\Xi_{\pm}(z,t);\quad \Xi_{\pm}(z,t)=\int_{0}^{\tau}Z_{\pm}(z\mp v_{\mathrm{g}}(\tau-\vartheta),t-\tau+\vartheta)\mathrm{d}\vartheta. \tag{32.59}$$

Then, assuming that the time interval is sufficiently short for the variation of the field along the traveling coordinate $\xi_{\mp}=z\mp v_{\mathrm{g}}t$ to be modest (crucially, no such requirement is made to the variations in z and t *separately*—in short-pulse ML, these can be quite strong due to the steep pulse fronts, but the evolution in the pulse shape as it propagates is relatively gentle meaning that variation along $\xi_{\mp}$ remains gentle), it is possible to approximate $\Xi_{\pm}(z,t)=\frac{v_{\mathrm{g}}\tau}{2}(Z_{\pm}(z\mp v_{\mathrm{g}}\tau,t-\tau)+Z_{\pm}(z,t))$. Then, using a (1,1) Padé approximation $\exp(x)\approx(1+x/2)/(1-x/2)$, one obtains a final formula relating the fields and polarizations at the previous and current points:

$$\left(1+\frac{\alpha_{\mathrm{f}}\tau}{2}\right)Y_{\pm}(z,t)=\left(1-\frac{\alpha_{\mathrm{f}}\tau}{2}\right)Y_{\pm}(z\mp v_{\mathrm{g}}\tau,t-\tau)+\frac{v_{\mathrm{g}}\tau}{2}(P_{\pm}^{t}(z\mp v_{\mathrm{g}}\tau,t-\tau)+P_{\pm}^{t}(z,t)). \tag{32.60}$$

This formula can be, and is, implemented directly in a numerical solver, in which a laser is separated into a number of segments n_{sect} with a length $v_{\mathrm{g}}\tau_{j}, j=1,\ldots,n_{\mathrm{sect}}$.

Accurate calculation of the polarization with the spectrum taken into account still requires a small *temporal* integration step Δt (commensurate with $1/\gamma_{T}$); however, crucially, when using Equation 32.59 as a basis of the numerical procedure, the *spatial* step $v_g\tau_j$ is limited only by the degree of uniformity with the field (roughly speaking, the requirement that $v_g\tau_j \left|g_{\text{mod}\,j}\right| << 1$, $g_{\text{mod}j}$ being the characteristic value of the modal gain (or absorption) in the section of the laser with the length $v_g\tau_j$). This is a much less stringent requirement, meaning that the space and time steps are essentially decoupled and the computational time therefore scales, not as $1/\Delta t^2$ as in the straightforward TWM, but only approximately as $1/\Delta t$.

From the numerical point of view, it is important to have an integer relation between the time and space steps:

$$\tau_j = n_{dj}\Delta t, \tag{32.61}$$

where the integer number n_{dj} is the *decimation factor* of the spatial section *j*. It is not necessary (nor advisable) for all the sections to have the same decimation factor; e.g., typically, in the SA part of a passively or hybridly mode-locked laser (where the field is less uniform), it is advisable to set the decimation factor smaller than in the more uniform gain sections. For numerical simplicity, identical decimation factors are typically used in all gain sections and in all SA sections, e.g., $n_{d(\text{gain})} = 71$ and $n_{d\text{SA}} = 8$ were used in Tandoi et al. (2013). This implies that the computation time required in a model with decimation is only a few percent of the time needed for the same calculation with a straightforward TWM model, which is indeed the case.

Mathematically speaking, a model with decimation is a set of n_{sect} delayed algebraic equations (DAEs) with the delay of $\tau_j = n_{dj}\Delta t$ in each section *j*. This section replaces having to deal with n_{dj} spatial steps which would be the case in a straightforward TWM analysis, which is why the alternative (original) term for the decimation technique is *folding space into time delay*. As the number of the DAEs in a typical simulation is not large ($<\sim 10$), the decimated model, unlike a straightforward TWM, is compatible with the numerical tools for bifurcation analysis.

The decimation/space-time folding technique of Tandoi et al. (2013) was not the first attempt at improving the efficiency of the traveling-wave solver. For example, this was attempted, with some degree of success, in the semianalytical TWM technique by Carroll et al. (Jones et al., 1995; Carroll et al., 1998), in which, however, the (homogeneous broadening type) digital filter was explicitly incorporated into the solver and the condition $\Delta z = v_g\Delta t$, eliminated in the decimated TWM, still stood. The decimation technique is thus both much more powerful and more versatile in that it can be used with *any arbitrary method* of calculating the polarization $P_\pm^t$, including either of the two microscopic approaches discussed in Section 32.5.1, or indeed a phenomenological two-level type dispersion representation of Equation 32.44.

However, the decimation technique is both particularly useful and particularly instructive with the analytical kernel technique of Equations 32.54 and 32.57, for which it was originally derived and used (which is the reason we have chosen to place the description of the model within the section on microscopic input). Figure 32.14 illustrates the "stencils" showing the time and space points required for propagating the fields in three major cases: (1) the TWM with a homogeneously broadened line (Figure 32.14a), (2) a straightforwardly implemented TWM with the analytical kernel (Figure 32.14b), and (3) the decimated model with the same analytical kernel and $n_{d(\text{gain})} = 8$, $n_{d\text{SA}} = 2$ (Figure 32.14c). The sparsity of the stencil clearly shows how the decimation allows the sophistication of the microscopic model to be retained while drastically improving the efficiency of the calculation.

We note, following Javaloyes and Balle (2012b), that the decimated TWM contains as particular cases the two other major model types discussed in previous sections. Indeed, with $n_{d(\text{gain})} = n_{d\text{SA}} = 1$, it reduces to a straightforward TWM of Section 32.3.4. For a unidirectional ring cavity (and particularly for a homogeneously broadened line), it reproduces the DDE approach of Section 32.3.3 (Javaloyes and Balle (2012b) noted also that for a single-frequency laser, their model can, in principle, reduce to a generalized complex rate equations formalism, but that is by definition not relevant for ML). A decimated TWM with an analytical kernel technique for gain/refractive index dispersion has been implemented to include a number of laser geometries including but not restricted to mode-locked lasers, and is available online as free software (MATLAB® toolkit) under GPL license (Javaloyes and Balle, 2012a).

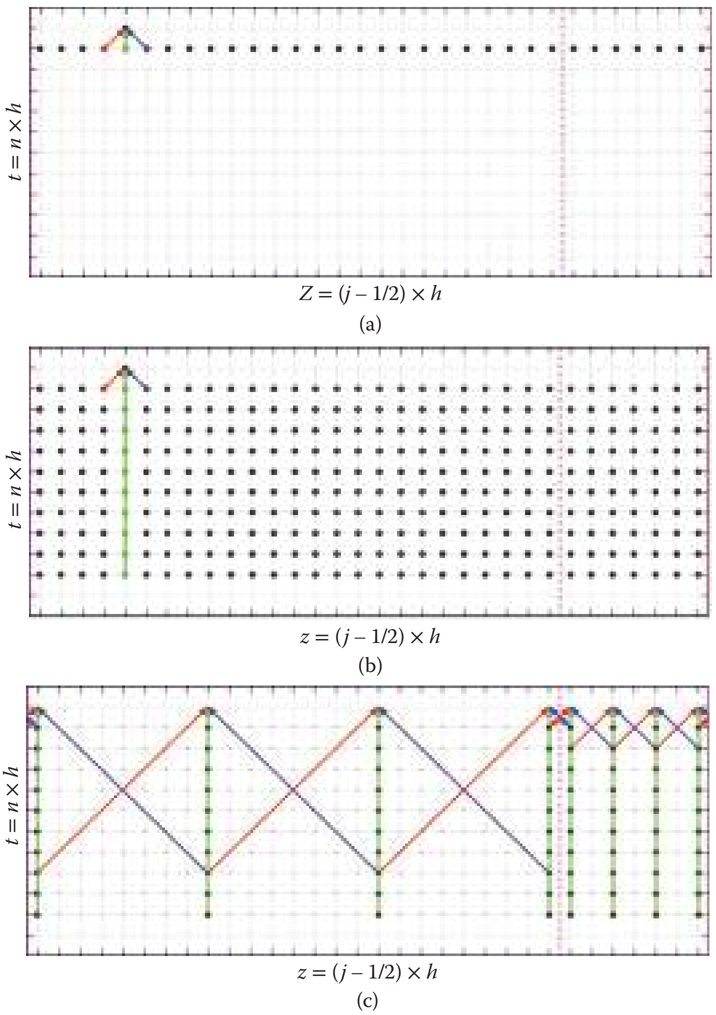


FIGURE 32.14 Calculation "stencils" illustrating field propagation in traveling wave modeling. (a) TWM with homogeneous broadening (Lorentzian gain spectrum), (b) analytical gain spectrum in a straightforward TWM, and (c) analytical gain spectrum in a TWM with decimation.

32.6 Some Novel Problems and Challenges in Mode-Locked Laser Modeling

32.6.1 Coherent Population Effects as a Possible Saturable Absorption Mechanism

One of the most important developments in mode-locked semiconductor laser technology in the recent years has been the direct generation of ultrashort (tens to a couple of hundreds of femtoseconds) pulses that until recently have been the domain of solid-state lasers only. Only one type of semiconductor lasers has been reported to produce such pulses so far, namely an optically pumped VECSEL, with the gain and SESAM chips in a folded (V-shaped) external cavity arrangement (Klopp et al., 2009, 2011; Quarterman et al., 2009; Wilcox et al., 2008). One of the teams that produced the femtosecond ML pulses attributed their results to coherent effects in the SESAM elements, namely the optical (ac) Stark effect. This effect, related to self-induced transparency, has the relaxation rate equal to the inverse of the dephasing (coherence decay) time. In semiconductors, this time is determined by carrier–carrier (and carrier–phonon) collisions and is of the order of 50–100 fs. This belongs to the *fast*, rather than slow, absorber regime, when the SA recovery time is shorter than the pulse duration. Indeed, theoretical analysis (Mihoubi et al., 2008; Wilcox et al., 2008), based on iterative small-signal time domain approach not dissimilar to that covered in Section 32.3.1 for a standard SA, showed that the pulse duration possible with this mechanism is about twice the dephasing time, which agrees well with the experiments. Further developments of this work are ongoing and are likely to rely not only on the experimental progress but also achieving better understanding of the limits and the requirements on the laser model through improved modeling.

There has also been work recently on the theoretical analysis of (so far, hypothetical) coherent ML through self-induced transparency effects in an *edge-emitting* class B (possibly semiconductor) laser geometry (see, e.g., Arkhipov et al., 2015a, 2016a,b), which promises femtosecond pulses but relies on *extending* the phase relaxation time, possibly by working under cryogenic conditions.

32.6.2 Spontaneous ML in Single-Section Lasers

Most of the discussions above concerned passive or hybrid ML constructions including an SA. However, in recent years, a very interesting development in ML laser technology occurred, when several teams have observed—and utilized—ML in single-section lasers without SA sections and without any external modulation either. The ML in this case was not of a pure AM ML type and the laser output only acquired short-pulse characteristics after external chirp compensation.

This effect has been observed in *quantum dash* (see, e.g., Duan et al., 2009; Gosset et al., 2006; Rosales et al., 2011), *QD* (Lu et al., 2011; Renaudier et al., 2005), QW (Sato, 2003; Yang, 2011), and even bulk materials (Yang, 2011), as well as in intersubband, far-infrared quantum cascade lasers (Faist et al., 2016 and references therein) showing that the effect is fairly generic. Further optimization of these lasers may partly depend on the establishment of full understanding of their behavior, yet at the time of writing, a full, universally agreed, theoretical explanation for ML in single-section lasers is still pending, and it is not impossible that different effects can play the main part in different constructions. In the past, some authors (Shore and Yee, 1991; Yee and Shore, 1993) used frequency-domain models with postulated, and differing, self- and cross- nonlinearity coefficients to show that nonlinearities in single-section semiconductor lasers could lead to a steady-state regime with fixed phases; this was predicted to produce, not the AM ML that corresponds to short-pulse emission, but the so-called frequency modulation ML, in which the phases of adjacent modes differ approximately by π, and the outcome is a CW-like regime with periodic carrier frequency oscillation. More recent works (Nomura et al., 2002; Renaudier et al., 2007), also using frequency-domain analysis with microscopically calculated (Nomura et al., 2002) or phenomenological (Renaudier et al., 2007) description of linear and nonlinear gain, predicted a possibility of ML-type signal generation, including AM ML for certain cavity lengths and active layer parameters, in a single-section laser with three modes involved in lasing, due to mode coupling by population pulsations/four-wave mixing effects. This appears to agree with traveling-wave modeling (Section 32.3.4) analysis for the case of a DBR laser without an SA (Bardella and Montrosset, 2005); Bardella and Montrosset (2005) also identified the role of four-wave mixing in their construction. Recently, a microscopic TWM analysis of a *QD* single-section laser has been presented (Gioannini et al., 2015), relying on both the nonequilibrium occupation of multiple levels due to inhomogeneous broadening and an additional phenomenological gain compression coefficient. The authors concluded that the suppressed carrier diffusion and the fast ground state (GS) gain recovery, typical of quantum dashes and QDs, were the mechanisms behind the phase-locking among the laser modes. Time-domain modeling on single-section ML (comb generation) in quantum cascade lasers has also been reported (Tzenov et al., 2016), though it has to be noted that the short upper state lifetime in those lasers makes their dynamics very different to that of other semiconductor lasers.

Further work extending the understanding achieved this far to other active media is desirable and hopefully forthcoming.

32.7 Concluding Remarks

We have attempted to review the most important developments in semiconductor mode-locked laser modeling. The topic is very dynamic and fast developing, so the choice of emphasis in this review was by necessity subjective; we apologize to those authors whose work may not have been given due prominence.

Acknowledgment

J.J. acknowledges financial support project COMBINA (TEC2015-65212-C3-3-P MINECO/FEDER UE)

References

Arkhipov R, Pimenov A, Radziunas M, Rachinskii D, Vladimirov AG, Arsenijevic D, Schmeckebier H, Bimberg D (2013) Hybrid mode locking in semiconductor lasers: Simulations, analysis, and experiments. *IEEE J Sel Top Quantum Electron* 19:1100208.

Arkhipov RM, Amann A, Vladimirov AG (2015b) Pulse repetition-frequency multiplication in a coupled cavity passively mode-locked semiconductor lasers. *Appl Phys B-Lasers Opt* 118:539–548.

Arkhipov RM, Arkhipov MV, Babushkin IV (2015a) On coherent mode-locking in a two-section laser. *Jetp Lett* 101:149–153.

Arkhipov RM, Arkhipov MV, Babushkin I (2016a) Self-starting stable coherent mode-locking in a two-section laser. *Opt Commun* 361:73–78.

Arkhipov RM, Arkhipov MV, Babushkin I, Rosanov NN (2016b) Self-induced transparency mode locking, and area theorem. *Opt Lett* 41:737–740.

Arkhipov RM, Habruseva T, Pimenov A, Radziunas M, Hegarty SP, Huyet G, Vladimirov AG (2016c) Semiconductor mode-locked lasers with coherent dual-mode optical injection: Simulations, analysis, and experiment. *J Opt Soc Am B-Opt Phys* 33:351–359.

Avrutin EA, Arnold JM, Marsh JH (1996) Analysis of dynamics of monolithic passively mode-locked laser diodes under external periodic excitation. *IEE Proc-Optoelectron* 143:81–88.

Avrutin EA, Arnold JM, Marsh JH (2003) Dynamic modal analysis of monolithic mode-locked semiconductor lasers. *IEEE J Sel Top Quantum Electron* 9:844–856.

Avrutin EA, Marsh JH, Portnoi EL (2000) Monolithic and multi-gigahertz mode-locked semiconductor lasers: Constructions, experiments, models and applications. *IEE Proc-Optoelectron* 147:251–278.

Avrutin EA, Nikolaev VV, Gallagher D (2005) Monolithic mode-locked semiconductor lasers. In: *Optoelectronic Devices—Advanced Simulation and Analysis* (Piprek J, ed.), pp. 185–215. New York, NY: Springer.

Avrutin and Panajotov, in preparation.

Avrutin EA, Portnoi EL, Chelnokov AV (1991) Effect of nonlinear amplification on characteristics of quality modulation regime in semiconducting laser with fast saturated absorbers. *Pisma V Zhurnal Tekhnicheskoi Fiziki* 17:49–54.

Avrutin EA, Rafailov EU (2012) Advances in mode-locked semiconductor lasers. In: *Advances in Semiconductor Lasers* (Coleman JJ, Bryce AC, Jagadish C, eds.), San Diego: Academic Press, pp. 93–147.

Avrutin EA, Russell BM (2009) Dynamics and spectra of monolithic mode-locked laser diodes under external optical feedback. *IEEE J Quantum Electron* 45:1456–1464.

Balle S (1998) Simple analytical approximations for the gain and refractive index spectra in quantum-well lasers. *Phys Rev A* 57:1304–1312.

Bandelow U, Radziunas M, Sieber J, Wolfrum M (2001) Impact of gain dispersion on the spatio-temporal dynamics of multisection lasers. *IEEE J Quantum Electron* 37:183–188.

Bandelow U, Radziunas M, Vladimirov A, Huttl B, Kaiser R (2006) 40 GHz mode-locked semiconductor lasers: Theory, simulations and experiment. *Opt Quantum Electron* 38:495–512.

Bardella P, Montrosset I (2005) Analysis of self-pulsating three-section DBR lasers. *IEEE J Sel Top Quantum Electron* 11:361–366.

Bischoff S, Mork J, Franck T, Brorson SD, Hofmann M, Frojdh K, Prip L, Sorensen MP (1997) Monolithic colliding pulse mode-locked semiconductor lasers. *Quantum Semiclass Opt* 9:655–674.

Bottge CN, Hader J, Kilen I, Moloney JV, Koch SW (2014) Ultrafast pulse amplification in mode-locked vertical external-cavity surface-emitting lasers. *Appl Phys Lett* 105:261105.

Brown CTA, Cataluna MA, Lagatsky AA, Rafailov EU, Agate MB, Leburn CG, Sibbett W (2004) Compact laser-diode-based femtosecond sources. *New J Phys* 6:175.

Camacho F, Avrutin EA, Cusumano P, Helmy AS, Bryce AC, Marsh JH (1997) Improvements in mode-locked semiconductor diode lasers using monolithically integrated passive waveguides made by quantum-well intermixing. *IEEE Photonics Technol Lett* 9:1208–1210.

Carroll JE, Whiteaway J, Plumb D (1998) *Distributed Feedback Semiconductor Lasers*, 2nd Edition. Stevenage: IET.

Cataluna MA, Nikitichev DI, Mikroulis S, Simos H, Simos C, Mesaritakis C, Syvridis D, Krestnikov I, Livshits D, Rafailov EU (2010) Dual-wavelength mode-locked quantum-dot laser, via ground and excited state transitions: Experimental and theoretical investigation. *Opt Express* 18:12832–12838.

Cataluna MA, Viktorov EA, Mandel P, Sibbett W, Livshits DA, Kovsh AR, Rafailov EU (2006) Temperature dependence of pulse duration in a mode-locked quantum-dot laser: Experiment and theory. In: *The 19th Annual Meeting of the IEEE Lasers and Electro-Optics Society* (IEEE, ed), pp. 798–799. Montreal, Canada.

Cataluna MA, Viktorov EA, Mandel P, Sibbett W, Livshits DA, Weimert J, Kovsh AR, Rafailov EU (2007) Temperature dependence of pulse duration in a mode-locked quantum-dot laser. *Appl Phys Lett* 90:101102–101103.

Cavailles JA, Miller DAB, Cunningham JE, Wa PLK, Miller A (1992) Simultaneous measurements of electron and hole sweep-out from quantum-wells and modeling of photoinduced field screening dynamics. *IEEE J Quantum Electron* 28:2486–2497.

Chow WW, Koch SW (1999) *Semiconductor Laser Fundamentals. Physics of the Gain Materials*. Berlin, Heidelberg: Springer.

Coldren LA, Corzine SW, Mashanovitch ML (2012) *Diode Lasers and Photonic Integrated Circuits*, 2nd Edition. New York, NY: Wiley.

Delpon EL, Oudar JL, Bouche N, Raj R, Shen A, Stelmakh N, Lourtioz JM (1998) Ultrafast excitonic saturable absorption in ion-implanted InGaAs/InAlAs multiple quantum wells. *Appl Phys Lett* 72:759–761.

Deryagin AG, Kuksenkov DV, Kuchinskii VI, Portnoi EL, Khrushchev IY, Frahm J (1994) Generation of high repetition frequency subpicosecond pulses at 1.535 mu-m by passive mode-locking of InGaAsP/InP laser diode with saturable absorber regions created by ion implantation. In: *14th IEEE International Semiconductor Laser Conference*, pp 107–108. New York, NY: IEEE.

Duan GH, Shen A, Akrout A, Van Dijk F, Lelarge F, Pommereau F, LeGouezigou O, Provost JG, Gariah H, Blache F, Mallecot F, Merghem K, Martinez A, Ramdane A (2009) High performance InP-based quantum dash semiconductor mode-locked lasers for optical communications. *Bell Labs Tech J* 14:63–84.

Dubbeldam JLA, Leegwater JA, Lenstra D (1997) Theory of mode-locked semiconductor lasers with finite absorber relaxation times. *Appl Phys Lett* 70:1938–1940.

Ell R, Morgner U, Kartner FX, Fujimoto JG, Ippen EP, Scheuer V, Angelow G, Tschudi T, Lederer MJ, Boiko A, Luther-Davies B (2001) Generation of 5-fs pulses and octave-spanning spectra directly from a Ti: Sapphire laser. *Opt Lett* 26:373–375.

Erneux T, Viktorov EA, Mandel P, Piwonski T, Huyet G, Houlihan J (2009) The fast recovery dynamics of a quantum dot semiconductor optical amplifier. *Appl Phys Lett* 94:113501.

Faist J, Villares G, Scalari G, Rosch M, Bonzon C, Hugi A, Beck M (2016) Quantum cascade laser frequency combs. *Nanophotonics* 5:272–291.

Fleck JA (1968) Emission of pulse trains by Q-switched lasers. *Phys Rev Lett* 21:131–133.

Gioannini M, Bardella P, Montrosset I (2015) Time-domain traveling-wave analysis of the multimode dynamics of quantum dot Fabry-Perot lasers. *IEEE J Sel Top Quantum Electron* 21:1–11.

Gosset C, Merghem K, Martinez A, Moreau G, Patriarche G, Aubin G, Landreau J, Lelarge E, Ramdane A (2006) Subpicosecond pulse generation at 134 GHz and low radiofrequency spectral linewidth in quantum dash-based Fabry-Perot lasers emitting at 1.5 μm. *Electron Lett* 42:91–92.

Green RP, Haji M, Hou LP, Mezosi G, Dylewicz R, Kelly AE (2011) Fast saturable absorption and 10 GHz wavelength conversion in Al-quaternary multiple quantum wells. *Opt Express* 19:9737–9743.

Hasler KH, Klehr A, Wenzel H, Erbert G (2005) Simulation of high-power pulse generation due to modelocking in long multisection lasers. *IEE Proc-Optoelectron* 152:77–85.

Haus HA (1975) Theory of mode-locking with a slow saturable absorber. *IEEE J Quantum Electron* 11:736–746.

Haus HA (2000) Mode-locking of lasers. *IEEE J Sel Top Quantum Electron* 6:1173–1185.

Haus HA, Silberberg Y (1985) Theory of mode-locking of a laser diode with a multiple-quantum-well structure. *J Opt Soc Am B-Opt Phys* 2:1237–1243.

Heck MJR, Bente E, Barbarin Y, Lenstra D, Smit MK (2006) Simulation and design of integrated femtosecond passively mode-locked semiconductor ring lasers including integrated passive pulse shaping components. *IEEE J Sel Top Quantum Electron* 12:265–276.

Hou LP, Avrutin EA, Haji M, Dylewicz R, Bryce AC, Marsh JH (2013) 160 GHz passively mode-locked AlGaInAs 1.55 mu m strained quantum-well lasers with deeply etched intracavity mirrors. *IEEE J Sel Top Quantum Electron* 19:1100409.

Hou LP, Haji M, Dylewicz R, Stolarz P, Qiu BC, Avrutin EA, Bryce AC (2010b) 160 GHz harmonic mode-locked AlGaInAs 1.55 mu m strained quantum-well compound-cavity laser. *Opt Lett* 35:3991–3993.

Hou LP, Haji M, Marsh JH (2014) Mode-locking and frequency mixing at THz pulse repetition rates in a sampled-grating DBR mode-locked laser. *Opt Express* 22:21690–21700.

Hou LP, Stolarz P, Dylewicz R, Haji M, Javaloyes J, Qiu BC, Bryce C (2010a) 160-GHz passively mode-locked AlGaInAs 1.55-mu m strained quantum-well compound cavity laser. *IEEE Photonics Technol Lett* 22:727–729.

Innerhofer E, Sudmeyer T, Brunner F, Haring R, Aschwanden A, Paschotta R, Honninger C, Kumkar M, Keller U (2003) 60-W average power in 810-fs pulses from a thin-disk Yb: YAG laser. *Opt Lett* 28:367–369.

Jaurigue L, Nikiforov O, Scholl E, Breuer S, Ludge K (2016) Dynamics of a passively mode-locked semiconductor laser subject to dual-cavity optical feedback. *Phys Rev E* 93:022205.

Jaurigue L, Pimenov A, Rachinskii D, Scholl E, Ludge K, Vladimirov AG (2015) Timing jitter of passively-mode-locked semiconductor lasers subject to optical feedback: A semi-analytic approach. *Phys Rev A* 92:053807.

Javaloyes J, Balle S (2010a) Quasiequilibrium time-domain susceptibility of semiconductor quantum wells. *Phys Rev A* 81:062505.

Javaloyes J, Balle S (2010b) Mode-locking in semiconductor Fabry-Perot lasers. *IEEE J Quantum Electron* 46:1023–1030.

Javaloyes J, Balle S (2011) Anticolliding design for monolithic passively mode-locked semiconductor lasers. *Opt Lett* 36:4407–4409.

Javaloyes J, Balle S (2012a) Freetwm: A simulation tool for semiconductor lasers. In: Freetwm: A simulation tool for semiconductor lasers.

Javaloyes J, Balle S (2012b) Multimode dynamics in bidirectional laser cavities by folding space into time delay. *Opt Express* 20:8496–8502.

Jones DJ, Zhang LM, Carroll JE, Marcenac DD (1995) Dynamics of monolithic passively mode-locked semiconductor lasers. *IEEE J Quantum Electron* 31:1051–1058.

Khalfin VB, Arnold JM, Marsh JH (1995) A theoretical-model of synchronization of a mode-locked semiconductor-laser with an external pulse stream. *IEEE J Sel Top Quantum Electron* 1:523–527.

Kilen I, Koch SW, Hader J, Moloney JV (2016) Fully microscopic modeling of mode locking in microcavity lasers. *J Opt Soc Am B-Opt Phys* 33:75–80.

Kim BS, Chung Y, Kim SH (1999) Dynamic analysis of mode-locked sampled-grating distributed Bragg reflector laser diodes. *IEEE J Quantum Electron* 35:1623–1629.

Kim I, Lau KY (1993) Frequency and timing stability of mode-locked semiconductor lasers-passive and active mode locking up to millimeter wave frequencies. *IEEE J Quantum Electron* 29:1081–1090.

Kim K, Lee S, Delfyett PJ (2005) 1.4kW high peak power generation from an all semiconductor mode-locked master oscillator power amplifier system based on eXtreme Chirped Pulse Amplification (X-CPA). *Opt Express* 13:4600–4606.

Klopp P, Griebner U, Zorn M, Klehr A, Liero A, Weyers M, Erbert G (2009) Mode-locked InGaAs-AlGaAs disk laser generating sub-200-fs pulses, pulse picking and amplification by a tapered diode amplifier. *Opt Express* 17:10820–10834.

Klopp P, Griebner U, Zorn M, Weyers M (2011) Pulse repetition rate up to 92 GHz or pulse duration shorter than 110 fs from a mode-locked semiconductor disk laser. *Appl Phys Lett* 98:071103.

Koumans R, vanRoijen R (1996) Theory for passive mode-locking in semiconductor laser structures including the effects of self-phase modulation, dispersion, and pulse collisions. *IEEE J Quantum Electron* 32:478–492.

Kuznetsov M (1985) Pulsations of semiconductor lasers with a proton bombarded segment: Well-developed pulsations. *IEEE J Quantum Electron* 21:587–592.

Lau KY, Paslaski J (1991) Condition for short pulse generation in ultrahigh frequency mode-locking of semiconductor lasers. *IEEE Photonics Technol Lett* 3:974–976.

Leegwater JA (1996) Theory of mode-locked semiconductor lasers. *IEEE J Quantum Electron* 32:1782–1790.

Lu ZG, Liu JR, Poole PJ, Jiao ZJ, Barrios PJ, Poitras D, Caballero J, Zhang XP (2011) Ultra-high repetition rate InAs/InP quantum dot mode-locked lasers. *Opt Commun* 284:2323–2326.

Malins DB, Gomez-Iglesias A, White SJ, Sibbett W, Miller A, Rafailov EU (2006) Ultrafast electroabsorption dynamics in an InAs quantum dot saturable absorber at 1.3 μm. *Appl Phys Lett* 89:171111–171113.

Marconi M, Javaloyes J, Balle S, Giudici M (2014) How lasing localized structures evolve out of passive mode locking. *Phys Rev Lett* 112:223901.

Marconi M, Javaloyes J, Camelin P, Gonzalez DC, Balle S, Giudici M (2015) Control and generation of localized pulses in passively mode-locked semiconductor lasers. *IEEE J Sel Top Quantum Electron* 21:1101210.

Martins JF, Avrutin EA, Ironside CN, Roberts JS (1995) Monolithic multiple colliding pulse mode-locked quantum-well lasers: Experiment and theory. *IEEE J Sel Top Quantum Electron* 1:539–551.

McDougall SD, Vogele B, Stanley CR, Ironside CN (1997) The crucial role of doping for high repetition rate monolithic mode locking of multiple quantum well GaAs/AlGaAs lasers. *Appl Phys Lett* 71:2910–2912.

Mecozzi A, Mork J (1997) Saturation effects in nondegenerate four-wave mixing between short optical pulses in semiconductor laser amplifiers. *IEEE J Sel Top Quantum Electron* 3:1190–1207.

Mihoubi Z, Daniell GJ, Wilcox KG, Tropper AC (2008) *Numerical Model of a Vertical-External-Cavity Surface-Emitting Semiconductor Lasers Mode-Locked by the Optical Stark Effect*. New York, NY: IEEE.

Mork J, Mecozzi A (1997) Theory of nondegenerate four-wave mixing between pulses in a semiconductor waveguide. *IEEE J Quantum Electron* 33:545–555.

Moskalenko V, Javaloyes J, Balle S, Smit M, Bente E (2013) Dynamics of colliding pulse passively semiconductor mode-locked ring lasers with an intra-cavity Mach-Zehnder modulator. In: *2013 Conference on and International Quantum Electronics Conference Lasers and Electro-Optics Europe*. Cleo Europe/Iqec.

Mulet J, Balle S (2005) Mode-locking dynamics in electrically driven vertical-external-cavity surface-emitting lasers. *IEEE J Quantum Electron* 41:1148–1156.

New GHC (1974) Pulse evolution in mode-locked quasicontinuous lasers. *IEEE J Quantum Electron* 10:115–124.

Nikitichev DI, Ding Y, Cataluna MA, Rafailov EU, Drzewietzki L, Breuer S, Elsaesser W, Rossetti M, Bardella P, Xu T, Montrosset I, Krestnikov I, Livshits D, Ruiz M, Tran M, Robert Y, Krakowski M (2012) High peak power and sub-picosecond Fourier-limited pulse generation from passively mode-locked monolithic two-section gain-guided tapered InGaAs quantum-dot lasers. *Laser Phys* 22:715–724.

Nikitichev DI, Ding Y, Ruiz M, Calligaro M, Michel N, Krakowski M, Krestnikov I, Livshits D, Cataluna MA, Rafailov EU (2011a) High-power passively mode-locked tapered InAs/GaAs quantum-dot lasers. *Appl Phys B-Lasers Opt* 103:609–613.

Nikitichev DI, Ruiz M, Ding Y, Tran M, Robert Y, Krakowski M, Rossetti M, Bardella P, Montrosset I, Krestnikov I, Lifshits D, Kataluna MA, Rafailov EU (2011b) Passively mode-locked monolithic two-section gain-guided tapered quantum-dot lasers: II. Record 15 Watt peak power generation. In: *CLEO Europe*, p CB3.4. Munich.

Nikolaev VV, Avrutin EA (2003) Photocarrier escape time in quantum-well light-absorbing devices: Effects of electric field and well parameters. *IEEE J Quantum Electron* 39:1653–1660.

Nikolaev VV, Avrutin EA (2004) Quantum-well design for monolithic optical devices with gain and saturable absorber sections. *IEEE Photonics Technol Lett* 16:24–26.

Nomura Y, Ochi S, Tomita N, Akiyama K, Isu T, Takiguchi T, Higuchi H (2002) Mode locking in Fabry-Perot semiconductor lasers. *Phys Rev A* 65:043807.

Otto C, Ludge K, Vladimirov AG, Wolfrum M, Scholl E (2012) Delay-induced dynamics and jitter reduction of passively mode-locked semiconductor lasers subject to optical feedback. *New J Phys* 14:113033.

Pimenov A, Vladimirov AG (2014) Theoretical analysis of passively mode-locked inhomogeneously broadened lasers. In: *14th International Conference on Numerical Simulation of Optoelectronic Devices* (Piprek J, Javaloyes J, eds.), pp. 151–152. 1–4 September 2014, Palma de Mallorca, Spain.

Quarterman AH, Wilcox KG, Apostolopoulos V, Mihoubi Z, Elsmere SP, Farrer I, Ritchie DA, Tropper A (2009) A passively mode-locked external-cavity semiconductor laser emitting 60-fs pulses. *Nat Photonics* 3:729–731.

Rafailov EU, Cataluna MA, Avrutin EA (2011) *Ultrafast Lasers Based on Quantum Dot Structures*. New York, NY/Berlin: Wiley.

Rebrova N, Huyet G, Rachinskii D, Vladimirov AG (2011) Optically injected mode-locked laser. *Phys Rev E* 83:8.

Renaudier J, Brenot R, Dagens B, Lelarge F, Rousseau B, Poingt F, Legouezigou O, Pommereau F, Accard A, Gallion P, Duan GH (2005) 45 GHz self-pulsation with narrow linewidth in quantum dot Fabry-Perot semiconductor lasers at 1.5 μm. *Electron Lett* 41:1007–1008.

Renaudier J, Duan GH, Landais P, Gallion P (2007) Phase correlation and linewidth reduction of 40 GHz self-pulsation in distributed Bragg reflector semiconductor lasers. *IEEE J Quantum Electron* 43:147–156.

Revin DG, Hemingway M, Wang Y, Cockburn JW, Belyanin A (2016) Active mode locking of quantum cascade lasers in an external ring cavity. *Nat Commun* 7:11440.

Rosales R, Merghem K, Martinez A, Akrout A, Tourrenc JP, Accard A, Lelarge F, Ramdane A (2011) InAs/InP quantum-dot passively mode-locked lasers for 1.55-mu m applications. *IEEE J Sel Top Quantum Electron* 17:1292–1301.

Rossetti M, Bardella P, Montrosset I (2011a) Time-domain travelling-wave model for quantum dot passively mode-locked lasers. *IEEE J Quantum Electron* 47:139–150.

Rossetti M, Bardella P, Montrosset I (2011b) Modeling passive mode-locking in quantum dot lasers: A comparison between a finite-difference traveling-wave model and a delayed differential equation approach. *IEEE J Quantum Electron* 47:569–576.

Salvatore RA, Sanders S, Schrans T, Yariv A (1996) Supermodes of high-repetition-rate passively mode-locked semiconductor lasers. *IEEE J Quantum Electron* 32:941–952.

Sato K (2003) Optical pulse generation using Fabry-Perot lasers under continuous-wave operation. *IEEE J Sel Top Quantum Electron* 9:1288–1293.

Schell M, Weber AG, Scholl E, Bimberg D (1991) Fundamental limits of sub-ps pulse generation by active mode-locking of semiconductor lasers: The spectral gain width and the facet reflectivities. *IEEE J Quantum Electron* 27:1661–1668.

Scollo R, Lobe HJ, Holzman JE, Robin E, Jackel H, Erni D, Vogt W, Gini E (2005) Mode-locked laser diode with an ultrafast integrated uni-traveling carrier saturable absorber. *Opt Lett* 30:2808–2810.

Scollo R, Lohe HJ, Robin F, Erni D, Gini E, Jackel H (2009) Mode-locked InP-based laser diode with a monolithic integrated UTC absorber for subpicosecond pulse generation. *IEEE J Quantum Electron* 45:322–335.

Shimizu T, Ogura I, Yokoyama H (1997) 860 GHz rate asymmetric colliding pulse modelocked diode lasers. *Electron Lett* 33:1868–1869.

Shimizu T, Wang XL, Yokoyama H (1995) Asymmetric colliding-pulse mode-locking in InGaAsP semiconductor lasers. *Opt Rev* 2:401–403.

Shore KA, Yee WM (1991) Theory of self-locking FM operation in semiconductor lasers. *IEE Proc J Optoelectron* 138:91–96.

Simos H, Rossetti M, Simos C, Mesaritakis C, Xu TH, Bardella P, Montrosset I, Syvridis D (2013) Numerical analysis of passively mode-locked quantum-dot lasers with absorber section at the low-reflectivity output facet. *IEEE J Quantum Electron* 49:3–10.

Stolarz PM, Javaloyes J, Mezosi G, Hou L, Ironside CN, Sorel M, Bryce AC, Balle S (2011) Spectral dynamical behavior in passively mode-locked semiconductor lasers. *IEEE Photonics J* 3:1067–1082.

Tandoi G, Javaloyes J, Avrutin E, Ironside CN, Marsh JH (2013) Subpicosecond colliding pulse mode locking at 126 GHz in monolithic GaAs/AlGaAs quantum well lasers: Experiments and theory. *IEEE J Sel Top Quantum Electron* 19:1100608.

Thompson MG, Rae A, Sellin RL, Marinelli C, Penty RV, White IH, Kovsh AR, Mikhrin SS, Livshits DA, Krestnikov IL (2006) Subpicosecond high-power mode locking using flared waveguide monolithic quantum-dot lasers. *Appl Phys Lett* 88:133119–133113.

Thompson MG, Rae AR, Xia M, Penty RV, White IH (2009) InGaAs quantum-dot mode-locked laser diodes. *IEEE Journal Sel Top Quantum Electron* 15:661–672.

Tzenov P, Burghoff D, Hu Q, Jirauschek C (2016) Time domain modeling of terahertz quantum cascade lasers for frequency comb generation. *Opt Express* 24:23232–23247.

Vasil'ev P (1995) *Ultrafast Diode Lasers: Fundamentals and Applications*. Boston: Artech House.

Viktorov EA, Erneux T, Mandel P, Piwonski T, Madden G, Pulka J, Huyet G, Houlihan J (2009) Recovery time scales in a reversed-biased quantum dot absorber. *Appl Phys Lett* 94:263502.

Viktorov EA, Mandel P, Vladimirov AG, Bandelow U (2006) Model for mode locking in quantum dot lasers. *Appl Phys Lett* 88:201102–201103.

Vladimirov AG, Pimenov AS, Rachinskii D (2009) Numerical study of dynamical regimes in a monolithic passively mode-locked semiconductor laser. *IEEE J Quantum Electron* 45:462–468.

Vladimirov AG, Turaev D (2005) Model for passive mode locking in semiconductor lasers. *Phys Rev A* 72:033808.

Vladimirov AG, Turaev D, Kozyreff G (2004) Delay differential equations for mode-locked semiconductor lasers. *Opt Lett* 29:1221–1223.

Wilcox KG, Mihoubi Z, Daniell GJ, Elsmere S, Quarterman A, Farrer I, Ritchie DA, Tropper A (2008) Ultrafast optical Stark mode-locked semiconductor laser. *Opt Lett* 33:2797–2799.

Williams KA, Sarma J, White IH, Penty RV, Middlemast I, Ryan T, Laughton FR, Roberts JS (1994) Q-switched bow-tie lasers for high-energy picosecond pulse generation. *Electron Lett* 30:320–321.

Xu TH, Montrosset I (2013) Quantum dot passively mode-locked lasers: Relation between intracavity pulse evolution and mode locking performances. *IEEE J Quantum Electron* 49:65–71.

Xu TH, Rossetti M, Bardella P, Montrosset I (2012) Simulation and analysis of dynamic regimes involving ground and excited state transitions in quantum dot passively mode-locked lasers. *IEEE J Quantum Electron* 48:1193–1202.

Yang WG (2011) Single-section Fabry-Perot mode-locked semiconductor lasers. *Adv Optoelectron* 2011:780373.

Yanson DA, Street MW, McDougall SD, Thayne IG, Marsh JH, Avrutin EA (2002) Ultrafast harmonic mode-locking of monolithic compound-cavity laser diodes incorporating photonic-bandgap reflectors. *IEEE J Quantum Electron* 38:1–11.

Yee WM, Shore KA (1993) Multimode analysis of self-locked FM operation in laser diodes. *IEE Proc J Optoelectron* 140:21–25.

Zarrabi JH, Portnoi EL, Chelnokov AV (1991) Passive-mode locking of a multistripe single quantum-well GaAs-laser diode with an intracavity saturable absorber. *Appl Phys Lett* 59:1526–1528.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

33

Quantum Cascade Lasers: Electrothermal Simulation

Song Mei

Yanbing Shi

Olafur Jonasson

and

Irena Knezevic

33.1 Introduction.. 235
Lasing in QCLs • Recent Developments in High-Power QCLs • QCL Modeling: An Overview
33.2 Electronic Transport.. 238
Semiclassical Techniques • Quantum Techniques • EMC with Nonequilibrium Phonons
33.3 Thermal Transport.. 243
Thermal Conductivity in a QCL Device • Device-Level Electrothermal Simulation
33.4 Device-Level Electrothermal Simulation: An Example........... 249
Electronic Simulation Results • Representative Electrothermal Simulation Results
33.5 Conclusion... 255

33.1 Introduction

Quantum cascade lasers (QCLs) are high-power, coherent light sources emitting in the mid-infrared (mid-IR) and terahertz (THz) frequency ranges [1]. QCLs are electronically driven, unipolar devices whose active core consists of tens to hundreds of repetitions of a carefully designed stage. The QCL active core can be considered a superlattice (SL) in which each stage is a multiple quantum well (MQW) heterostructure where confined electronic states with specific energy levels are formed because of quantum confinement. The concept of achieving lasing in semiconductor SLs was first introduced by Kazarinov and Suris [2] in 1971. The first working QCL was demonstrated by Faist et al. [1] two decades later.

QCLs are typically III-V material systems grown on GaAs or InP substrates. Molecular beam epitaxy (MBE) [3] and metal-organic chemical vapor deposition (MOCVD) [4] are the techniques that enable precise growth of thin layers of various III-V alloys. It is also possible to incorporate strain into the structure as long as the total strain in a stage is balanced. Both the precision and the possibility of introducing strain bring great flexibility to the design of the QCL active core, so lasing over a wide range of wavelengths (from 3 to 190 μm) has been achieved. The growth techniques produce high-quality interfaces with atomic-level roughness.

Mid-IR QCLs (wavelength range 3–12 μm) have widespread military and commercial applications. A practical portable detector requires mid-IR QCLs to operate at room temperature (RT) in continuous wave (CW) mode and with high (watt-level) output power. Furthermore, these QCLs must also have high wall-plug efficiency (WPE) (the ratio of emitted optical power to the electrical power pumped in) and long-term reliability under these high-stress operating conditions. As the stress likely stems from excessive nonuniform heating while lasing [5,6], improving device reliability and lifetime goes hand in hand with improving the WPE.

33.1.1 Lasing in QCLs

In QCLs, multiple conduction subbands are formed in the active core by means of quantum confinement. QCLs are unipolar devices, meaning that lasing is achieved through radiative intersubband transitions (transitions between two conduction subbands) instead of radiative interband transitions (transitions between the conduction and valence bands) in traditional quantum well (QW) semiconductor lasers. As a result, electrons do not combine with holes after the radiative transitions and can be used to emit another photon. In order to reuse electrons, the same MQW heterostructure is repeated many times (25–70) in the QCL active core (the so-called cascading scheme).

Figure 33.1 depicts a typical conduction band diagram of two adjacent stages in a QCL under an electric field. Each stage consists of an injector region and an active region. The injector region has several thin wells separated by thin barriers (10–30 Å), so a miniband is formed with multiple subbands that are close in energy and whose associated wavefunctions have high spatial overlap. Typically, the lowest few energy levels in the miniband are referred to as the injector levels. The injector levels collect the electrons that come from the previous stage and inject them into the active region. The active region usually consists of 2–3 wider wells (40–50 Å) separated by thin barriers. Consequently, a minigap forms in the active region between the upper lasing level (3) and the lower lasing level (2). Another important energy level in the active region is the ground state (1). There is a thin barrier (usually the thinnest among all layers) between the injecting region and the active region, called the injection barrier.

By design, the injector levels are close in energy and strongly coupled to the upper lasing level because of the thin injection barrier. The upper and lower lasing levels have large spatial overlap, which allows a radiative transition between the two levels; the wavelength of the emitted light is determined by the energy spacing between these two levels. The lower lasing level overlaps with the ground state for efficient electron extraction. Electron emission of longitudinal optical (LO) phonons is the dominant mechanism for electron extraction, so the energy spacing between the lower lasing level and the ground state is designed to be close to the LO phonon energy to facilitate extraction. With careful design, the electron lifetime in the upper lasing level is longer than in the lower lasing level, so population inversion can be

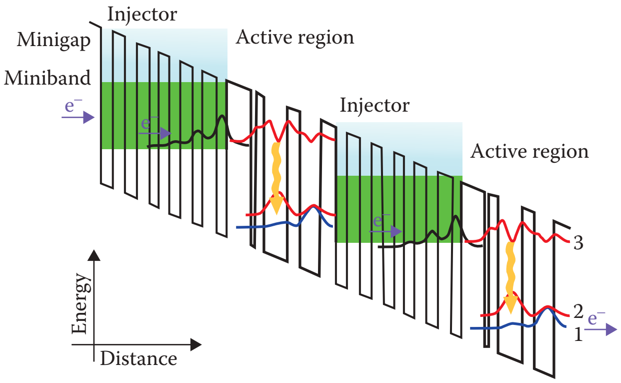


FIGURE 33.1 A typical conduction band diagram of two adjacent QCL stages under an applied electric field. Each stage consists of an injector region and an active region. A miniband is formed in the injector region while a minigap is formed in the active region (between the upper and lower lasing levels). Lasing is associated with a radiative transition from the upper (3) to the lower (2) lasing level. Electrons in the lower lasing level depopulate quickly to the ground level (1) by emission of longitudinal optical phonons.

achieved. After reaching the ground state, electrons tunnel through the injector into the upper lasing level of the next stage, and the process is repeated. Of course, the lasing mechanism description above is idealized. In reality, the efficiency of the radiative transition between the upper and lower lasing levels is very low [7,8].

33.1.2 Recent Developments in High-Power QCLs

In recent years, considerable focus has been placed on improving the WPE and output power of QCLs for RT CW operation. Bai et al. [9] showed 8.4% WPE and 1.3 W output power around 4.6 μm in 2008. Shortly thereafter, Lyakh et al. [10] reported 12.7% WPE and 3 W power at 4.6 μm. Watt-level power with 6% WPE at 3.76 μm and then lower power at 3.39 and 3.56 μm are reported by Bandyopadhyay et al. [11,12]. Bai et al. [13] demonstrated 21% WPE and 5.1 W output power around 4.9 μm in 2011. Much higher WPE and/or output power has been achieved at lower temperatures or at pulsed mode [7,14] near 4.8 μm. A summary of recent developments can be found in review papers [8,15].

While good output powers and WPEs have been achieved, long-term reliability of these devices under RT CW operation remains a critical problem [5,6]. These devices are prone to catastrophic breakdown owing to reasons that are not entirely understood, but are likely related to thermal stress that stems from prolonged high-power operation [5]. This kind of thermal stress is worst in short-wavelength devices that have high strain and high thermal-impedance mismatch between layers [11,12,16,17].

In addition to improved device lifetime, we seek better CW temperature performance (higher characteristic temperatures T_0 and T_1, defined next) [6]. The first aspect is a weaker temperature dependence of the threshold current density. Empirically, the threshold current density (the current density at which the device starts lasing) has an exponential dependence on the operating temperature T: $J_{\mathrm{th}} \propto \exp\left(\frac{T}{T_0}\right)$. Higher characteristic temperature T_0 is preferred in QCL design, as it means less variation in J_{th} as the temperature changes.

Another key temperature-dependent parameter is the differential quantum efficiency (also called the slope efficiency or external quantum efficiency), defined as the amount of output optical power dP per unit increase in the pumping current dI: $\eta_{\mathrm{d}} = \frac{\mathrm{d}P}{\mathrm{d}I} \propto \exp\left(-\frac{T}{T_1}\right)$. The differential quantum efficiency is directly proportional to the WPE ($\mathrm{WPE} = \eta_{\mathrm{d}}\eta_{\mathrm{f}}$, where η_{f} is the feeding efficiency). Therefore, the higher the T_1, the closer the η_{d} is to unity, and the higher the WPE. Recently, deep-well structures with tapered active regions have demonstrated significant improvements in T_0 and T_1 with respect to the conventional 4.6-μ*m* device [9], underscoring that the suppression of leakage plays a key role in temperature performance [6,18,19]. Still, the microscopic mechanisms and leakage pathways that contribute to these empirical performance parameters remain unclear.

33.1.3 QCL Modeling: An Overview

Under high-power RT CW operation, both electron and phonon systems in QCLs are far away from equilibrium. In such nonequilibrium conditions, both electronic and thermal transport modeling are important for understanding and improving QCL performance.

Electron transport in both mid-IR and THz QCLs has been successfully simulated via semiclassical (rate equations [20–22] and Monte Carlo [23–26]) and quantum techniques (density matrix [27–33], nonequilibrium Green's functions (NEGF) [34–36], and lately Wigner functions [37]). InP-based mid-IR QCLs have been addressed via semiclassical [38] and quantum transport approaches (8.5-μm [39] and 4.6-μm [35,36] devices). There has been a debate whether electron transport in QCLs can be described using semiclassical models, in other words, how much of the current in QCLs is coherent. Theoretical work by Iotti and Rossi [23,40] shows that the steady-state transport in mid-IR QCLs is largely incoherent. Monte Carlo simulation [41] has also been used to correctly predict transport near threshold. However, short-wavelength structures [9] have pronounced coherent features, which cannot be addressed semiclassically

[33]. NEGF simulations accurately and comprehensively capture quantum transport in these devices but are computationally demanding. Density matrix approaches have considerably lower computational overhead than NEGF but are still capable of capturing coherent transport features. A comprehensive review of electron transport modeling was recently written by Jirauschek and Kubis [42].

Electronic simulations that ignore radiative transitions are applicable for modeling QCLs below or near threshold where the interaction between electrons and the laser electromagnetic field can be ignored. Such simulations are useful for predicting quantities such as threshold current density and T_0. However, in order to accurately model QCLs under lasing operations, the effect of the laser field on electronic transport would have to be included. In some cases, the effects of the laser field can be very strong [39], especially for high-WPE devices, where the field-induced current can be dominant [38]. When included in simulations, the laser field is typically modeled either as an additional scattering mechanism [38,43] or as a time-dependent sinusoidal electric field [39,44]. In this work, we ignore the effect of the laser field on electron dynamics.

Thermal transport in QCLs is often described through the heat diffusion equation, which requires accurate thermal conductivity in each region, a challenging task for the active core that contains many interfaces [17,45–48]. It is also very important to include nonequilibrium effects, such as the nonuniform heat generation rate stemming from the nonuniform temperature distribution [47] and the feedback that the nonequilibrium phonon population has on electron transport [26].

In this chapter, we present a multiphysics (coupled electronic and thermal transport) and multiscale (bridging between a single stage and device level) simulation framework that enables the description of QCL performance under far-from-equilibrium conditions [49]. We present the electronic (Section 33.2) and thermal (Section 33.3) transport models, and then bring them together for electrothermal simulation of a real device structure (Section 33.4). We strive to cover the basic ideas while pointing readers to the relevant references for derivation and implementation details.

33.2 Electronic Transport

Depending on the desired accuracy and computational burden, one can model electronic transport in QCLs with varying degrees of complexity. The goal is to determine the modal gain (proportional to the population inversion between the upper and lower lasing levels) under various pumping conditions (current or voltage) and lasing conditions (pulsed wave or CW). A typical electron transport simulator relies on accurately calculated quasibound electronic states and associated energies in the direction of confinement. Electronic wavefunctions and energies are determined by solving the Schrödinger equation or the Schrödinger equation combined with the Poisson's equation in highly doped systems. Section 33.2.3.1 introduces a $\mathbf{k} \cdot \mathbf{p}$ Schrödinger solver coupled with a Poisson solver. More information about other solvers for electronic states can be found in the review paper [42] and references therein.

The simulations of electronic transport fall into two camps depending on how the electron single-particle density matrix is treated. The diagonal elements of the density matrix represent the occupation of the corresponding levels and off-diagonal elements represent the "coherence" between two levels. Transport is semiclassical or incoherent when the off-diagonal coherences are much smaller than the diagonal terms and can be approximated as proportional to the diagonal terms times the transition rates between states [50]. In that case, the explicit calculation of the off-diagonal terms is avoided and only the diagonal elements are tracked, which simplifies the simulation considerably. However, when the off-diagonal terms are appreciable, transport is partially coherent and has to be addressed using quantum transport techniques.

33.2.1 Semiclassical Techniques

Semiclassical approaches assume that electronic transport between stages is largely incoherent "hopping" transport. The key quantities are populations of electronic states that are confined in the QCL growth direction, and electrons transfer between them due to scattering events. The scattering rates can be obtained empirically or more rigorously via Fermi's golden rule. Common semiclassical approaches are the rate equations and ensemble Monte Carlo (EMC), the latter solving a Boltzmann-like transport equation stochastically.

33.2.1.1 Rate Equations

In the rate equation approach [20–22], scattering between relevant states, i.e., the injector level, the upper and lower lasing levels, and the ground state, is captured through transition rates. The rates include all relevant (radiative and nonradiative) scattering mechanisms and can be either empirical parameters or calculated [51,52]. The computational requirements of rate equation models are low, so they are suitable for fast numerical design and optimization of different structures [22].

33.2.1.2 Ensemble Monte Carlo

The heterostructure in the QCL active core is a quasi-two-dimensional (quasi-2D) system, where electrons are free to move in the x–y plane, while confined cross-plane in the z direction; the confinement results in the formation of quasibound states and discrete energy levels corresponding to the bottoms of 2D energy subbands. The electron wavefunctions in 3D are plane waves in the x–y plane and confined wavefunctions in z direction. Electronic transport is captured by a Boltzmann-like semiclassical transport equation [23], which can be solved via the stochastic EMC technique assuming instantaneous hops between states in 3D due to scattering [53]. The simulation explicitly tracks the energy level and in-plane momentum of each particle in the simulation ensemble (typically $\sim 10^5$ particles). Tracking in-plane dynamics makes it more detailed than the rate equation model. The transition rates are generally computed directly from the appropriate interaction Hamiltonians and therefore depend on the energy levels as well as the wavefunction overlaps between different electronic states [42,53,54]. EMC allows us to include nonequilibrium effect into transport, which is covered in more detail in Section 33.2.3.

33.2.2 Quantum Techniques

Density matrix and NEGF are the two most widely used techniques to describe quantum transport in QCLs. Recently, a Wigner function approach was also successfully used to model a SL [37].

33.2.2.1 Density Matrix Approaches

In semiclassical approaches, the central quantity of interest is the distribution function $f_n^{E_\mathrm{k}}(t)$, the probability of an electron occupying an eigenstate n and having an in-plane kinetic energy E_k. The quantum-mechanical analogue is the single-electron density matrix, $\rho_{nm}^{E_\mathrm{k}}(t)$, where the diagonal elements $\rho_{nn}^{E_\mathrm{k}}(t) = f_n^{E_\mathrm{k}}(t)$ are occupations and the off-diagonal elements $\rho_{nm}^{E_\mathrm{k}}(t)$ are the spatial coherences between states n and m at the in-plane energy E_k. When employing semiclassical methods, off-diagonal matrix elements are assumed to be much smaller than diagonal elements. This approximation may fail in some cases, e.g., when two eigenstates with a large spatial overlap have similar energies. This scenario often arises when modeling THz QCLs [28,32,55], but can also come up in mid-IR QCLs [33]. In these cases, semiclassical models fail.

The density matrix models that have been employed for QCL modeling can be categorized into two groups. The first includes hybrid methods, where transport is treated semiclassically within a region of the device (typically a single stage) while the effects of tunneling between different regions, separated by barriers, are treated quantum mechanically using a density matrix formalism with phenomenological dephasing times [28,31,55]. The second group involves completely quantum-mechanical methods that rely on microscopically derived Markovian master equations that guarantee positivity of the density matrix [32,33]. Both methods are more computationally expensive than their semiclassical counterparts, because the density matrix contains many more elements than its diagonal semiclassical analogue.

33.2.2.2 Nonequilibrium Green's Functions

The NEGF technique (see a good overview in Reference [42]) relies on the relationships between single-particle time-ordered Green's functions and correlation functions [34–36]. The correlation function $G^{<}_{\alpha,\beta}(k;t_1,t_2)$, often referred to as the lesser Green's function [44,56], is one of the central quantities and can be understood as a two-time generalization of the density matrix, where k refers to the magnitude of in-plane wave vector. The correlation function contains both spatial correlations (terms with $\alpha \neq \beta$) as well as temporal correlations between times t_1 and t_2 (not included in semiclassical or density matrix models). Typically, the potential profile is assumed to be time independent, in which case the correlation function only depends on the time difference $G^{<}_{\alpha,\beta}(k;t_1,t_2) = G^{<}_{\alpha,\beta}(k;t_1 - t_2)$. Fourier transform over the time difference into the energy domain gives the energy-resolved correlation function $G^{<}_{\alpha,\beta}(k,E)$, which is the quantity that is usually solved for numerically [35,44,56]. The main advantages of the NEGF formalism are that it provides spectral (energy-resolved) information and it includes the effects of collisional broadening (the broadening of energy levels due to scattering), which is particularly important when the states are close in energy. These advantages carry a considerable computational cost, so NEGF calculations are much more time-consuming than density matrix approaches [39].

33.2.3 EMC with Nonequilibrium Phonons

Here, we focus on presenting semiclassical modeling of electron transport in QCL structures via EMC [25,26,57]. The solver consists of two parts: a coupled Schrödinger–Poisson solver and a transport kernel. We solve for the electronic states using the coupled Schrödinger–Poisson solver and feed the energy levels and the wavefunctions of the relevant electronic states to the transport kernel. The transport kernel keeps track of the electron momentum, energy, and distribution among subbands. If the electron density inside the device is high, transport kernel will periodically feed the electron distribution back to the Schrödinger–Poisson solver and update the electronic states. This loop is repeated until the electron distribution converges. By doing so, we solve for both the electron transport and the electronic band structure self-consistently.

Since the active QCL core consists of repeated stages, the wavefunctions in any stage can be obtained from the wavefunctions in any other stage by translation in space and energy. This translational symmetry makes it possible to simulate electron transport in only one generic central stage instead of in the whole QCL core [53]. Typically, electronic states in nonadjacent stages have negligible overlap, which also means that the transition rates between them are negligible. As a result, it is sufficient to limit interstage scattering events to only those between adjacent stages.

Figure 33.2 shows a schematic of three adjacent stages under an applied field. We simulate electron transport in the central stage λ, while nearest-neighbor interstage ($\lambda \rightleftarrows \lambda \pm 1$) and intrastage ($\lambda \to \lambda$) scattering is allowed. Periodic boundary conditions (PBCs) are applied in the simulation, i.e., whenever one electron scatters from the central stage into the next stage (process ②), an electron scatters from the previous stage into the central stage (process ①) and vice versa (process ③ and process ④). PBCs are justified by the cascading scheme.

33.2.3.1 Electronic Bandstructure Solver

We employ the **k·p** method to solve the Schrödinger equation and couple it to a Poisson solver [25,53,57]. The **k·p** method is an efficient way to solve for the electronic band structure near the band edges, where the transport happens in QCLs. The **k·p** method considers the contribution from the conduction band (C), light-hole band (LH), and the spin-orbit split-off band (SO) (the heavy-hole band (HH) decouples from the other three at the band edge) [58]. The contributions from the LH and SO are especially important for narrow-gap materials, such as InP. Moreover, in modern QCLs, strain-balanced structures have been employed to obtain enhanced performance. In these structures, alternate layers are compressively or tensilely strained while the whole structure is strain free, with carefully designed thickness of each

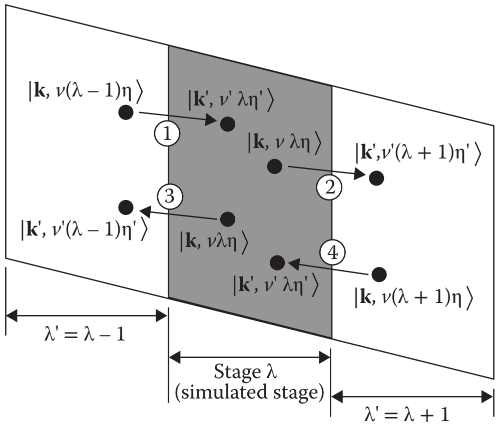

$|\mathbf{k}, \nu(\lambda - 1)\eta\rangle$
$|\mathbf{k}', \nu' \lambda\eta'\rangle$
①
$|\mathbf{k}, \nu \lambda\eta\rangle$
$|\mathbf{k}', \nu'(\lambda + 1)\eta'\rangle$
②
③
$|\mathbf{k}', \nu'(\lambda - 1)\eta'\rangle$
$|\mathbf{k}, \nu\lambda\eta\rangle$
④
$|\mathbf{k}', \nu' \lambda\eta'\rangle$
$|\mathbf{k}, \nu(\lambda + 1)\eta\rangle$
$\lambda' = \lambda - 1$
Stage λ
(simulated stage)
$\lambda' = \lambda + 1$

FIGURE 33.2 Schematic of the three simulated stages in a QCL active core under an applied field. Scattering is limited to nearest-neighbor stages and periodic boundary conditions are justified by the cascading scheme; therefore, only three stages are needed in the EMC transport kernel.

layer. The **k·p** method allows for convenient inclusion of the effects of strain on the band structure. The implementation details of the **k·p** solver can be found in Reference [53].

The **k·p** solver can only solve for a finite structure rather than an infinite periodic one. As a result, we need to simulate a finite number of stages and add artificially high barriers to the two ends to confine all the states. If a stage is far enough from the boundaries, the calculated band structure in it should be the same as if we were to solve for the whole periodic structure. Tests have confirmed that three stages, which we also used in EMC, are enough when solving for the electronic states to ensure that the central-stage states are unaffected by the simulation-domain potential boundaries. The states from the central stage are then translated in energy and position to the neighboring stages according to the stage length and the applied electric field.

When we need to solve for electron transport and electronic states self-consistently, it is necessary for the solver to be able to automatically pick out the electronic states belonging to the central stage. One intuitive criterion is to calculate the "center of mass" for each state (the expectation value of the cross-plane coordinate, $\langle z \rangle$) and assign those falling in the central stage to that stage. However, in our three-stage scheme, this method may pick up the states that are too close to the boundary. One can either extend the number of stages in the **k·p** solver to five, so the three stages in the middle are all far from the boundary, or use additional criteria such as that there be more than 50% possibility of finding an electron in the central stage, based on the probability density distribution, or requiring that the location of the probability density peak be in the central stage. Additional criteria requiring strong confinement of states have been explored in [32].

33.2.3.2 Transport Kernel with Nonequilibrium Phonons

The EMC kernel tracks the hopping transitions of electrons between subbands and stages until convergence and outputs the transport information for us to calculate the experimentally relevant quantities such as current and modal gain [53]. In the transport kernel, both electron–electron interactions and electron–LO-phonon interactions are considered. Other scattering processes, such as intervalley scattering, impurity scattering, and interface roughness scattering can be considered under different circumstances [42]. Photon emission is not considered, either. Because EMC tracks individual particles, nonequilibrium electron transport can be automatically captured. (EMC tracks individual simulation particles, each of which might represent thousands of real electrons.)

The most important scattering mechanism in QCLs is electron–LO-phonon scattering, which facilitates the depopulation of the lower lasing level. As shown in Reference [59], phonon confinement has little effect on the electronic transport; therefore, for simplicity, LO phonons are treated as bulklike dispersionless phonons with energy $\hbar\omega_0$. The transition rate between an initial state $\phi_i(z)$ with energy E_i and a final state $\phi_f(z)$ withenergy E_f can be derived from the Fermi's golden rule as follows:

$$\Gamma_{a(-),e(+)} = \frac{e^2\hbar\omega_0 m_f^*}{8\pi^2\hbar^3}\left(\frac{1}{\epsilon_\infty}-\frac{1}{\epsilon_0}\right)\int_0^{2\pi} d\theta \int_{-\infty}^{\infty} dq_z \int_0^{\infty} dE_{kf} N_{\mathbf{q}} \frac{|\mathcal{I}_{\mathrm{if}}(q_z)|^2}{\mathbf{q}_{\|}^2 + q_z^2}\delta(E_f - E_i \mp \hbar\omega_0), \qquad (33.1)$$

where e is the electronic charge while ϵ_0 and ϵ_∞ are static and high-frequency electronic permittivities of the material, respectively. The integrals are over the in-plane kinetic energy E_k of the final state and the cross-plane momentum transfer q_z. $\mathbf{q}_{\|} = \mathbf{k}'_{\|} - \mathbf{k}_{\|}$ is the in-plane momentum transfer.

$$|\mathcal{I}_{\mathrm{if}}(q_z)|^2 = \left| \int_0^d dz \phi_f^*(z)\phi_i(z)e^{-izq_z} \right|^2 \qquad (33.2)$$

is defined as the overlap integral (OI) between the initial and final states, where q_z is the cross-plane momentum transfer. The integration is over the angle between initial and final in-plane momenta $\mathbf{k}_{\|}$ and $\mathbf{k}'_{\|}$ (θ), cross-plane momentum component of the final state (k_{fz}), and the kinetic energy of the final state (E_{kf}). $N_{\mathbf{q}}$ represents the number of LO phonons with momentum $\mathbf{q} = (\mathbf{q}_{\|}, q_z)$. The expression can be further simplified in the equilibrium case, where $N_{\mathbf{q}}$ follows the Bose–Einstein distribution [53]. In order to model nonequilibrium phonon effects, we numerically integrated the expression using a phonon number histogram according to both $q_{\|}$ and q_z [54].

According to the uncertainty principle, position and momentum both cannot be determined simultaneously. Since our electrons are all confined in the central stage ($\Delta z'$ is finite), the cross-plane momentum is not exactly conserved during the scattering process ($q_z \neq k'_z - k_z$) [60]. This analysis does not affect the momentum conservation in the x–y plane, because we assume infinite uncertainty in position there. Previously, the cross-plane momentum conservation has been considered through the momentum conservation approximation (MCA) [61,62] and a broadening of q_z according to the well width [60]. The MCA forbids a phonon emitted between subbands i and f to be reabsorbed by another transition between i' and f' if $i \neq i'$ or $f \neq f'$, and thus might underestimate the electron–LO interaction strength [54]. The concept of well width is hard to apply in an MQW structure such as the QCL active core [54]. We observe that the probability of a phonon with cross-plane momentum q_z being involved in an interaction is proportional to the OI in Equation 33.2. Figure 33.3 depicts the typical OIs for both intersubband ($i_1 \rightarrow 3$ and $2 \rightarrow 1$) and intrasubband ($3 \rightarrow 3$) transitions. As a result, in each electron–LO-phonon scattering event, we randomly select a q_z following the distribution from the OI (Figure 33.3). Depending on the mechanism (absorption or emission), a phonon with $(\mathbf{q}_{\|}, q_z)$ is removed or added to the histogram according to the 2D density of states (DOS) and the effective simulation area [54]. Once the phonons with a certain momentum are depleted, transitions involving such phonons become forbidden.

In order to couple the EMC solver to the thermal transport solver, we need to keep a detailed log of heat generation during electron transport. In all the relevant scattering events, electron–LO-phonon scattering is the only inelastic mechanism and therefore is the only mechanism that contributes to heat generation. As a result, the total energy emitted and absorbed in the form of LO phonons is recorded during each step of the EMC simulation. The nonequilibrium phonons decay into acoustic longitudinal acoustic (LA) phonons via a three-phonon anharmonic decay process. The formulation and the parameters here follow [63]. The simulation results of EMC including nonequilibrium phonons are shown in Section 33.4.

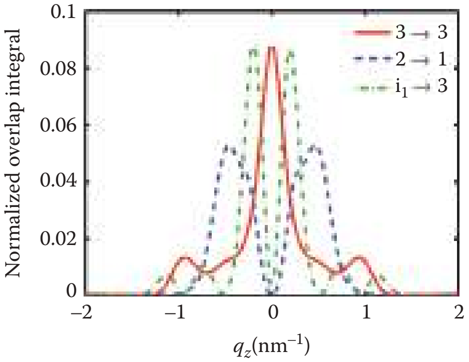


FIGURE 33.3 Normalized overlap integral $|\mathcal{I}_{\text{if}}|^2$ from Equation 33.2 versus cross-plane phonon wave vector q_z for several transitions (intersubband $i_1 \rightarrow 3$ and $2 \rightarrow 1$; intrasubband $3 \rightarrow 3$). (Reprinted with permission from Y. B. Shi and I. Knezevic. Nonequilibrium phonon effects in midinfrared quantum cascade lasers. *Journal of Applied Physics,* 116(12):123105, 2014. Copyright 2014, American Institute of Physics.)

33.3 Thermal Transport

The dominant path of heat transfer in a QCL structure is depicted in Figure 33.4. The operating electric field of a typical QCL is high, which means that considerable energy is pumped into the electronic system. These energetic, "hot" electrons relax their energy largely by emitting LO phonons. LO phonons have high energies but flat dispersions, so their group velocities are low and they are poor carriers of heat. An LO phonon decays into two LA phonons via a three-phonon process referred to as anharmonic decay. LA phonons have low energy but high group velocity and are the main carriers of heat in semiconductors [47,63]. If we neglect the diffusion of optical phonons, the flow of energy in a QCL can be described by the following equations:

$$\frac{\partial W_{\text{A}}}{\partial t} = \nabla \cdot (\kappa_{\text{A}} \nabla T_{\text{A}}) + \left.\frac{\partial W_{\text{LO}}}{\partial t}\right|_{\text{coll}}; \quad \frac{\partial W_{LO}}{\partial t} = \nabla \cdot (\kappa_{\text{A}} \nabla T_{\text{A}}) + \left.\frac{\partial W_{\text{e}}}{\partial t}\right|_{\text{coll}} - \left.\frac{\partial W_{\text{LO}}}{\partial t}\right|_{\text{coll}}, \tag{33.3}$$

where W_{LO}, W_{A}, and W_{e} are the LO phonon, acoustic phonon, and electron energy densities, respectively. κ_{A} is the thermal conductivity in the system and T_{A} is the acoustic phonon (lattice) temperature. The term $\nabla \cdot (\kappa_{\text{A}} \nabla T_{\text{A}})$ describes heat diffusion, governed by acoustic phonons. We have also used the fact that the rate of increase in the LO-phonon energy density equals the difference between the rate of its generation by electron–LO-phonon scattering and the rate of anharmonic decay into LA phonons.

In a nonequilibrium steady state, both the LO and LA energy densities are constant, so

$$-\nabla \cdot (\kappa_{\text{A}} \nabla T_{\text{A}}) = \left.\frac{\partial W_{\text{e}}}{\partial t}\right|_{\text{coll}}. \tag{33.4}$$

As described in the previous section, the right-hand side of Equation 33.4 is the heat generation rate Q and can be obtained by recording electron–LO-phonon scattering events in electronic EMC [47,64]:

$$Q = \left.\frac{\partial W_{\text{e}}}{\partial t}\right|_{\text{coll}} = \frac{N_{\text{3D}}}{N_{\text{sim}} t_{\text{sim}}} \sum (\hbar\omega_{\text{ems}} - \hbar\omega_{\text{abs}}), \tag{33.5}$$

where $N_{\text{3D}} = N_s/D_{\text{stage}}$ is the electron density (N_{s} is the sheet density and D_{stage} is the length of a single stage) while N_{sim} and t_{sim} are the number of simulation particles and the simulation time, respectively,

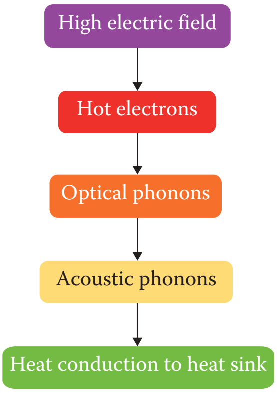


FIGURE 33.4 Flow of energy in a QCL.

and $\hbar\omega_{\text{ems}}$ and $\hbar\omega_{\text{abs}}$ are the energies of the emitted and absorbed LO phonons, respectively. To solve Equation 33.4, we need information on both the thermal conductivity κ_A and the heat generation rate Q; they are discussed in Subsections 33.3.1 and 33.3.2, respectively.

33.3.1 Thermal Conductivity in a QCL Device

33.3.1.1 Active Core: A III-V Superlattice

The QCL active core is an SL: It contains many identical stages, each with several thin layers made from different materials and separated by heterointerfaces. The thermal conductivity tensor of an SL system reduces to two values: the in-plane thermal conductivity $\kappa_{\|}$ (in-plane heat flow is assumed isotropic) and the cross-plane thermal conductivity $\kappa_{\perp}$. Experimental results have shown that, in SLs, the thermal conductivity is very anisotropic [65] ($\kappa_{\|} \gg \kappa_{\perp}$) while both $\kappa_{\|}$ and $\kappa_{\perp}$ are smaller than the weighted average of the constituent bulk materials [66–70]. Both effects can be attributed to the interfaces between adjacent layers [71,72].

Here, we discuss a semiclassical model for describing the thermal conductivity tensor of III-V SL structures. Note that the model described here is in principle applicable to SLs in other material systems as long as they have high-quality interface and thermal transport is mostly incoherent [48,73–75]. In particular, we focus on thermal transport in III-arsenide-based SLs, as they are most commonly used in mid-IR QCL active cores [48].

Under QCL operation conditions of interest (>77 K, and typically near RT), thermal transport is dominated by acoustic phonons and is governed by the Boltzmann transport equation (BTE). To obtain the thermal conductivity, we solve the phonon BTE with full phonon dispersion in the relaxation-time approximation [48].

33.3.1.2 Twofold Influence of Effective Interface Roughness

To capture both the anisotropic thermal transport and the reduced thermal conductivity in SL systems, we need to observe the twofold influence of the interface. First, it reduces $\kappa_{\|}$ by affecting the acoustic phonon population close to the interfaces [76]. Second, it introduces an interface thermal boundary resistance (ITBR), which is still very difficult to model [65,77]. Common models are the acoustic mismatch model (AMM) and the diffuse mismatch model (DMM) [65,76]; the former assumes a perfectly smooth interface and only considers the acoustic mismatch between the two materials, while the latter assumes complete randomization of momentum after phonons hit the interface. As most III-V-based QCLs are

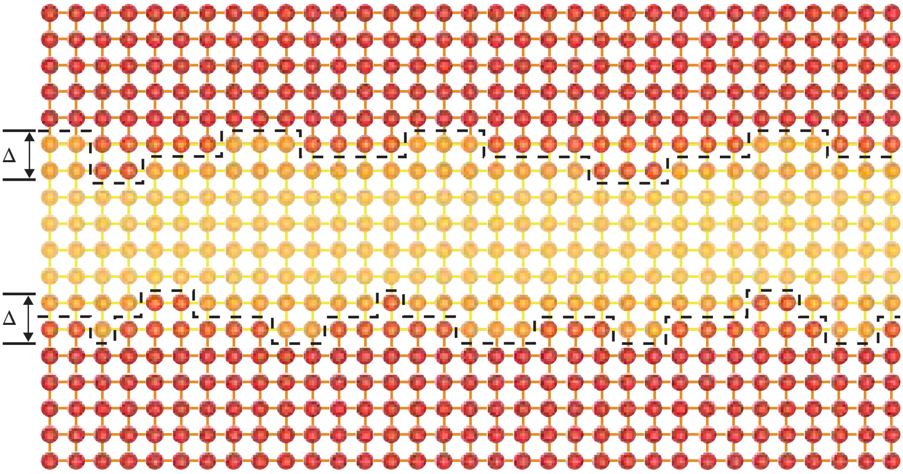


FIGURE 33.5 Even between lattice-matched crystalline materials, there exist nonuniform transition layers that behave as an effective atomic-scale interface roughness with some rms roughness Δ. This effective interface roughness leads to phonon momentum randomization and to interface resistance in cross-plane transport. (From S. Mei and I. Knezevic, *Journal of Applied Physics*, 118, 175101, 2015. With the permission of AIP Publishing.)

grown by MBE or MOCVD, both well-controlled techniques allowing consistent atomic-level precision, neither AMM nor DMM captures the essence of a III-V SL interface. Figure 33.5 shows a schematic of interface roughness in a lattice-matched SL. The jagged dashed boundaries depict transition layers of characteristic thickness Δ between the two materials.

We introduce a simple model that calculates a more realistic ITBR (a key part in calculating $\kappa_\perp$) by interpolating between the AMM and DMM transmission rates using a specularity parameter p_{spec}. The model has a single fitting parameter: the effective interface rms roughness Δ. Since the growth environment is well controlled, using one Δ to describe all the interfaces is justified. We use Δ to calculate a momentum-dependent specularity parameter:

$$p_{\text{spec}}(\vec{q}) = \exp(-4\Delta^2|\vec{q}|^2\cos^2\theta), \tag{33.6}$$

where $|\vec{q}|$ is the magnitude of the phonon wave vector and θ is the angle between $\vec{q}$ and the normal direction to the interface. Consistent with the twofold impact of interface roughness, Δ affects the thermal conductivity through two channels. Apart from calculating the ITBR, an effective interface scattering rate $\tau^{-1}_{\text{interface}}(\vec{q})$ dependent on the same specularity parameter $p_{\text{spec}}(\vec{q})$ is added to the internal scattering rate to calculate modified $\kappa_{\|}$ (see detailed derivations in [48]). By adjusting only Δ, typically between 1 and 2 Å, the calculated thermal conductivity using this model fits a number of different experiments [66,68,69].

33.3.1.3 $\kappa_{\|}$ and $\kappa_\perp$ of a QCL Active Core

Thermal transport inside the active core of a QCL is usually treated phenomenologically: $\kappa_{\|}$ is typically assumed to be 75% of the weighted average of the bulk thermal conductivities of the constituent materials, while $\kappa_\perp$ is treated as a fitting parameter (constant for all temperatures) to best fit the experimentally measured temperature profile [17,78]. We calculated the thermal conductivity tensor of a QCL active core [78] and showed that the typical assumption is not accurate and that the degree of anisotropy is temperature dependent (Figure 33.6).

The ratio between $\kappa_{\|}$ and the averaged bulk value (inset to Figure 33.6) varies between 45% and 70% over the temperature range of interest. $\kappa_\perp$ has a weak dependence on temperature in keeping with the

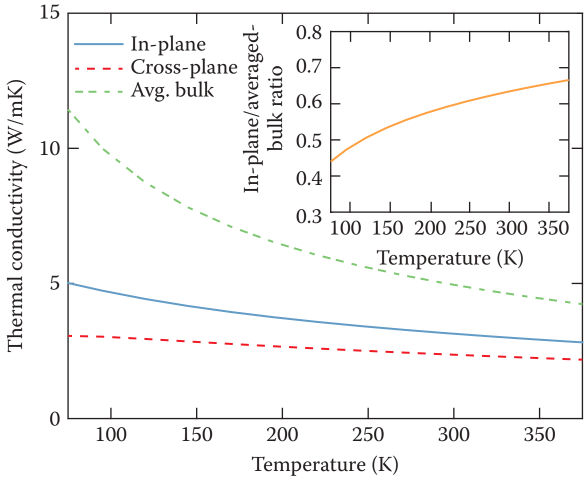


FIGURE 33.6 Thermal conductivity of a typical QCL active region [78] as a function of temperature. A single stage consists of 16 alternating layers of $In_{0.53}Ga_{0.47}As$ and $In_{0.52}Al_{0.48}As$. The solid curve, dashed curve, and dashed-dotted curve show the calculated in-plane, cross-plane, and averaged bulk thermal conductivity, respectively. $\Delta = 1$ Å in the calculations. The inset shows the ratio between the calculated in-plane and the averaged bulk thermal conductivities. (From S. Mei and I. Knezevic, *Journal of Applied Physics*, 118, 175101, 2015. With the permission of AIP Publishing.)

TABLE 33.1 Thermal Conductivity as a Function of Temperature for Materials in a QCL Structure

Materials	Thermal Conductivity (W/mK)
Au	$337–600 \times 10^{-4}T$
Si_3N_4	$30–1.4 \times 10^{-2}T$
In solder	$93.9–6.96 \times 10^{-2}T + 9.86 \times 10^{-5}T^2$

common assumption in simplified models; the weak temperature sensitivity means that ITBR dominates cross-plane thermal transport. These results show that it is important to carefully calculate the thermal conductivity tensor in QCL thermal simulation, and we use this thermal conductivity model in the device-level simulation.

33.3.1.4 Other Materials

The active core is not the only region we need to model in a device-level thermal simulation. Figure 33.7 shows a typical schematic (not to scale) of a QCL device in thermal simulation with a substrate-side mounting configuration [79]. The active core (in this case, consisting of 36 stages and 1.6-µm thick) with width W_{act} is embedded between two cladding layers (4.5-µm-thick GaAs). The waveguide is supported by a substrate (GaAs) with thickness D_{sub}. An insulation layer (Si_3N_4) with thickness D_{ins} is deposited around the waveguide and then etched away from the top to make the contact. Finally, a contact layer (Au) with thickness D_{cont} and a thin layer of solder (D_{sold}) are deposited on top. There is no heat generation in the regions other than the active core. Further, these layers are typically thick enough to be treated as bulk materials. Bulk substrate (GaAs or InP) thermal conductivities are readily obtained for III-V materials from experiment, as well as from relatively simple theoretical models [17,48,78,80] (Table 33.1).

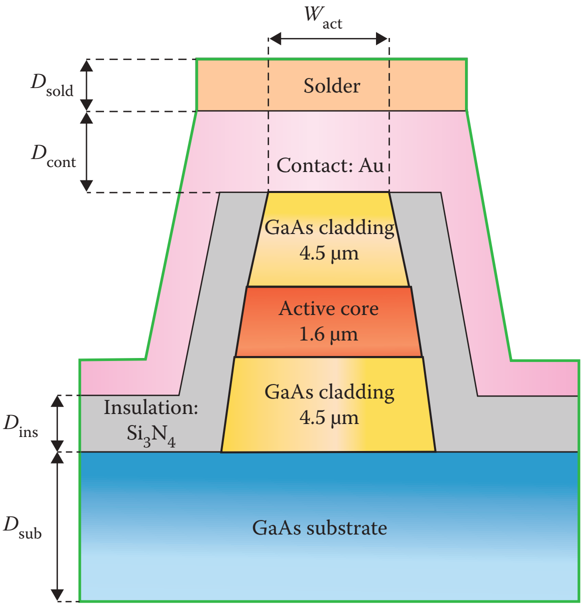


FIGURE 33.7 Schematic of a typical GaAs-based mid-IR QCL structure with a substrate (not to scale).

33.3.2 Device-Level Electrothermal Simulation

33.3.2.1 Device Schematic

The length of a QCL device is much greater than its width, therefore we can assume the length is infinite and carry out a 2D thermal simulation. The schematic of the simulation domain (not to scale) is shown in Figure 33.7. The boundary of the simulation region is delineated, and certain boundary conditions (heat sink at fixed temperature, convective boundary condition, or adiabatic boundary condition) can be applied (independently) to each boundary. Typically, the bottom boundary of the device is connected to a heat sink while other boundaries have the convective boundary condition at the environment temperature (single-device case) or the adiabatic boundary condition (QCL array case). Typical values for the layers thickness are $W_{act} = 15$ μm, $D_{sub} = 50$ μm, $D_{ins} = 0.3$ μm, $D_{cont} = 3$ μm, and $D_{sold} = 1.5$ μm.

We use the finite element method to solve for the temperature distribution. The whole device is divided into different regions according to their materials properties. Each stage of the active region is treated as a single unit with the heat generation rate tabulated in the device table in order to capture the nonuniform behavior among stages. The active core is very small but is also the only region with heat generation, small thermal conductivity, and spatial nonuniformity. To capture the behavior of the active region while saving computational time, we use a nonuniform mesh in the finite element solver to emphasize the active core region. Figure 33.8 shows a mesh generated in the simulation.

33.3.2.2 Simulation Algorithm

It is known that among all the stages in the active core, the temperature T_i and the electric field F_i (i represents the stage index) are not constant [56,78], but we have no *a priori* knowledge of how they depend on the stage index. However, we know that the charge–current continuity equation must hold, and in the steady state $\nabla \cdot \mathbf{J} = 0$; this implies that the current density J must be uniform, as the current flow is essentially in one dimension, along z. This insight is key to bridging the single-stage and device-level simulations.

From Section 33.2.3, we can obtain the heat generation rate Q inside the active core by running the single-stage EMC simulation. Each single-stage EMC is carried out at a specific electric field F and

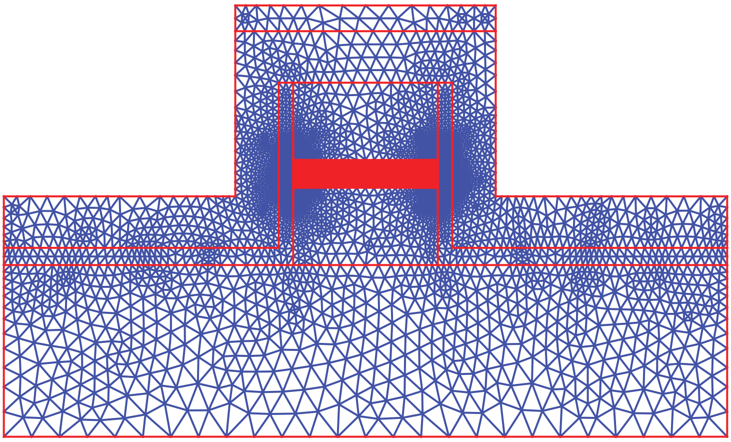


FIGURE 33.8 A typical nonuniform finite-element mesh of the simulated GaAs-based mid-IR QCL structure.

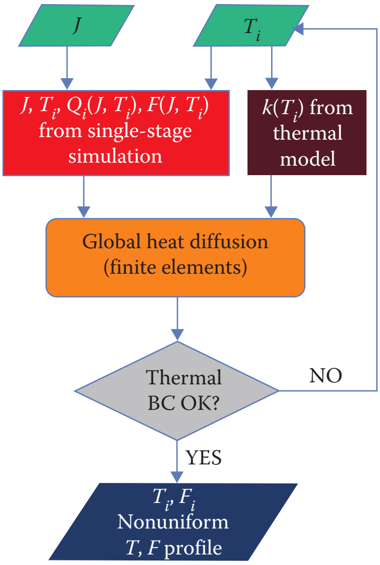

J
T_i
$J, T_i, Q_i(J, T_i), F(J, T_i)$ from single-stage simulation
$k(T_i)$ from thermal model
Global heat diffusion (finite elements)
Thermal BC OK?
NO
YES
T_i, F_i Nonuniform T, F profile

FIGURE 33.9 Flowchart of the device-level thermal simulation. We start by assuming a certain current density J and temperature profile T_i across the whole device. Based on the tabulated information from the single-stage simulation and assumed (J, T_i), we get stage-by-stage profiles for the electric field F_i and the heat generation rate Q_i. An accurate temperature-dependent thermal conductivity model, which includes the boundary resistances of layers, and the temperature profile guess are used as input to the heat diffusion equation, which is then iteratively solved (with updated temperature profile in each step) until the thermal boundary conditions are satisfied.

temperature T and outputs both the current density $J(F, T)$ and the heat generation rate $Q(F, T)$. By sweeping F and T in range of interest, we obtain a table connecting different field and temperature (F, T) to appropriate current density and heat generation rate (J, Q) $[(F, T) \rightarrow (J, Q)]$. However, based on the discussion above, the input in the thermal simulation needs to be the constant parameter J. Therefore, we "flip" the recorded $(F, T) \rightarrow (J, Q)$ table to a so-called device table $(J, T) \rightarrow (Q, F)$, suitable for coupled simulation [49].

Figure 33.9 depicts the flowchart of the device-level electrothermal simulation [49]. Before the simulation, we obtain the device table $[(J, T) \rightarrow (Q, F)]$, as discussed above. We also have to calculate the thermal conductivities ($\kappa_{\|}$ and $\kappa_{\perp}$) of the active region as a function of temperature and tabulate them, based on the model described in Section 33.3.1. We also need the bulk thermal conductivity of other materials in the device (cladding layer, substrate, insulation, contact, and solder) as a function of temperature. These material properties are standard and already well characterized.

Each device-level thermal simulation is carried out in a certain environment (i.e., for a given set of boundary conditions) and with a certain current density J. At the beginning of the simulation, an initial temperature profile is assigned. With the input from the device table and the thermal conductivity data in each region, we use a finite element method to iteratively solve the heat diffusion equation until convergence. At the end of the simulation, we obtain a thermal map of the whole device. Further, from the temperature T_i in each stage and the injected current density J, we obtain the nonuniform electric field distribution F_i. With the electric field in each stage and given the stage thickness, we can accurately calculate the voltage drop across the device and obtain the current–voltage characteristic. By changing the mounting configuration (W_{act}, D_{sub}, D_{ins}, D_{cont}, D_{sold}) or the boundary conditions, the temperature profile can be changed.

33.4 Device-Level Electrothermal Simulation: An Example

In this section, we present detailed simulation results of a 9-μm GaAs/$\text{Al}_{0.45}\text{Ga}_{0.55}\text{As}$ mid-IR QCL [79] based on a conventional three-well active region design. The chosen structure has 36 repetitions of the single stage; each stage has 16 layers. Starting from the injection barrier, the layer thicknesses in one stage (in Å) are $\mathbf{46}/19/\mathbf{11}/54/\mathbf{11}/48/\mathbf{28}/34/\mathbf{17}/30/\underline{\mathbf{18}}/\underline{28}/\underline{\mathbf{20}}/\underline{30}/\mathbf{26}/30$. Here, the barriers ($\text{Al}_{0.45}\text{Ga}_{0.55}\text{As}$) are in bold while the wells (GaAs) are in normal font; the underlined layers are doped to a sheet density of $n_{\text{Si}} = 3.8 \times 10^{11}\ \text{cm}^{-2}$. The results at 77 K are shown here.

33.4.1 Electronic Simulation Results

33.4.1.1 Band Structure

Figure 33.10 shows the electronic states of the chosen structure under the designed operating field of 48 kV/cm, calculated from the coupled **k**·**p**–Poisson solver (see Section 33.2.3.1). The active region states of the central stage are represented by dark solid curves 1, 2, and 3 are the ground state and the lower

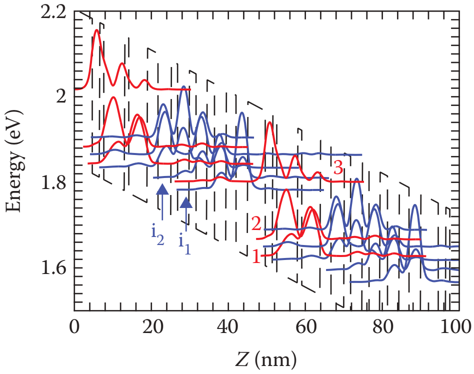


FIGURE 33.10 Energy levels and wavefunction moduli squared of Γ-valley subbands in two adjacent stages of the simulated GaAs/AlGaAs-based structure. The dark solid curves denote the active region states (1, 2, and 3 represent the ground state and the lower and upper lasing levels, respectively). The lighter solid curves represent injector states, with i_1 and i_2 denoting the lowest two. (From Y. B. Shi and I. Knezevic, *Journal of Applied Physics*, 116, 123105, 2014. With the permission of AIP Publishing.)

and upper lasing levels, respectively. Injector states are labeled i_1 and i_2. Other lightly colored solid curves denote the states that form the miniband. (When the electron density in the QCL is high, the electronic bands have to be calculated self-consistently with EMC.)

33.4.1.2 *J*-*F* Curve

The current density J versus field F curve, one of the key QCL characteristics at a given temperature, is intuitive to obtain in EMC. After calculating the electronic band structure at a certain field F, the wavefunctions, energy levels, and effective masses of each subband and each stage are fed into the EMC solver. In the EMC simulation, we include all the scattering mechanisms described in Section 33.2.3. Since we employ PBCs, the current density J can be extracted from how many electrons cross the stage boundaries in a certain amount of time in the steady state. The net flow n_{net} of electrons is calculated by subtracting the flow between the central stage and the previous stage (n_{backward}) from the flow between the central stage and the next stage (n_{forward}) in each time step. The current density is then calculated as follows:

$$J = \frac{e n_{\text{net}}}{A_{\text{eff}} \delta t} = \frac{e(n_{\text{forward}} - n_{\text{backward}})}{A_{\text{eff}} \delta t}, \tag{33.7}$$

where δt is the time interval during which the flow is recorded. A_{eff} is the effective in-plane area of the simulated device. Since doping is the main source of electrons, the area is calculated as follows:

$$A_{\text{eff}} = \frac{N_{\text{ele}}}{N_{\text{s}}}, \tag{33.8}$$

where N_{ele} is the number of simulated electrons and N_{s} is the sheet doping density (in cm^{-2}) in the fabricated device. In the current simulation, $N_{\text{ele}} = 50,000$ and $N_{\text{s}} = 3.8 \times 10^{11}\ \text{cm}^{-2}$.

Because of the stochastic nature of EMC, we need to average the current density over multiple time steps. In practice, one can record the net cumulative number of electrons per unit area that leave a stage over time and obtain a linear fit to this quantity in the steady state; the slope yields the steady-state current density.

From each individual simulation, we extract the current density at a given electric field and temperature. To obtain the *J*-*F* curve at that temperature, we sweep the electric field. To demonstrate the importance of including nonequilibrium phonons effects, we carry out the simulation with thermal phonons alone and with both thermal and excess nonequilibrium phonons. Figure 33.11 is the *J*-*F* curve for the simulated structure with (filled squares) and without (empty squares) nonequilibrium phonons at 77 K. It can be seen that the current density at a given field considerably increases when nonequilibrium phonons are included and the trend holds up to 60 kV/cm. This difference is prominent at low temperatures (<200 K) and goes away at RT [26].

33.4.1.3 Modal Gain (G_{m}) and Threshold

We calculate the modal gain as follows [22]:

$$G_{\text{m}} = \frac{4\pi e^2 \langle z_{32} \rangle^2 \Gamma_{\text{w}} \Delta n}{2\varepsilon_0 \underline{n} \gamma_{32} L_p \lambda}, \tag{33.9}$$

where ε_0 is the permittivity of free space. Some constants are obtained from experiment: waveguide confinement factor $\Gamma_{\text{w}} = 0.31$, stage length $L_p = 45$ nm, optical-mode refractive index $\underline{n} = 3.21$, and full width at half maximum $\gamma_{32}(T_{\text{L}}) \approx 8.68\ \text{meV} + 0.045\ \text{meV/K} \times T_{\text{L}}$ [26,79]. The dipole matrix element between the upper and lower lasing levels ($\langle z_{32} \rangle = 1.7$ nm) and the emission wavelength ($\lambda = 9\ \mu\text{m}$) are

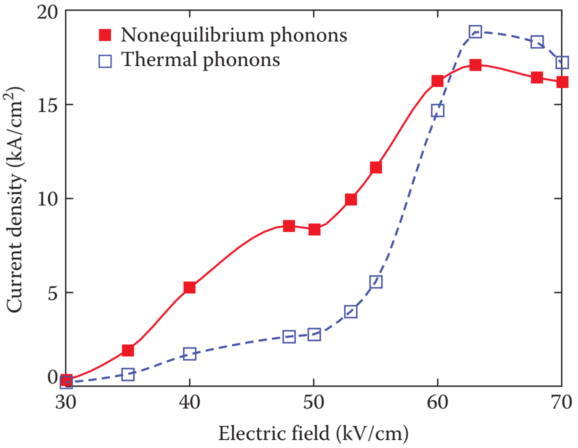


FIGURE 33.11 The current density versus electric field ($J - F$) curve of the simulated device with (filled squares) and without (empty squares) the nonequilibrium phonon effect at 77 K. The inclusion of nonequilibrium phonons considerably increases the current density at a given field up to 60 kV/cm.

also estimated in experiment [79], but we calculate these two terms directly. The dipole matrix element is calculated as follows:

$$\langle z_{32} \rangle = \int_0^d z\varphi_3^*(z)\varphi_2(z)dz\,. \tag{33.10}$$

The value is slightly different at different fields, as the band structure changes. At 48 kV/cm, the calculated matrix element is $\langle z_{32} \rangle = 1.997$ nm. Similarly, the wavelength of emitted photon also changes at different fields. One can calculate the value from the energy difference between the upper and lower lasing levels. The calculated wavelength at 48 kV/cm is 8.964 μm. $\Delta n = n_{\text{upper}} - n_{\text{lower}}$ is the population inversion obtained from EMC. Again, due to the randomness of EMC, the population inversion needs to be averaged over a period after the steady state has been reached.

Figure 33.12 shows the modal gain of the device with nonequilibrium (filled squares) and thermal (empty squares) phonons as a function of (1) electric field and (2) current density at 77 K. Horizontal dotted line indicates the total estimated loss in the device, which is used to help find the threshold current density, J_{th}. Lasing threshold is achieved when the modal gain G_{m} equals the total loss α_{tot}. We consider two sources of loss, mirror (α_{m}) and waveguide (α_{w}), so the total loss is $\alpha_{\text{tot}} = \alpha_{\text{m}} + \alpha_{\text{w}}$. The intercepts between the total loss line and the G_{m} versus F (Figure 33.12a) and G_{m} versus J (Figure 33.12b) curves give the threshold field F_{th} and threshold current density J_{th}, respectively. Like the current density, the modal gain of the device is also considerably higher when nonequilibrium phonons are considered, which leads to a lower F_{th} and a lower J_{th}. The reason for the increased current density and modal gain with nonequilibrium phonons can be attributed to the enhanced injection selectivity and efficiency [26].

33.4.1.4 Heat Generation Rate

The way to obtain the heat generation rate Q is similar to how we get the current density J. We record the cumulative net energy emission as a function of time and fit a straight line to the region where the simulation has reached a steady state. The slope of the line is used in place of $\sum(\hbar\omega_{\text{ems}} - \hbar\omega_{\text{abs}})/t_{\text{sim}}$. Figure 33.13 shows the heat generation rate as a function of electric field at 77 K. The filled squares and the empty squares depict the situation with and without nonequilibrium phonons, respectively.

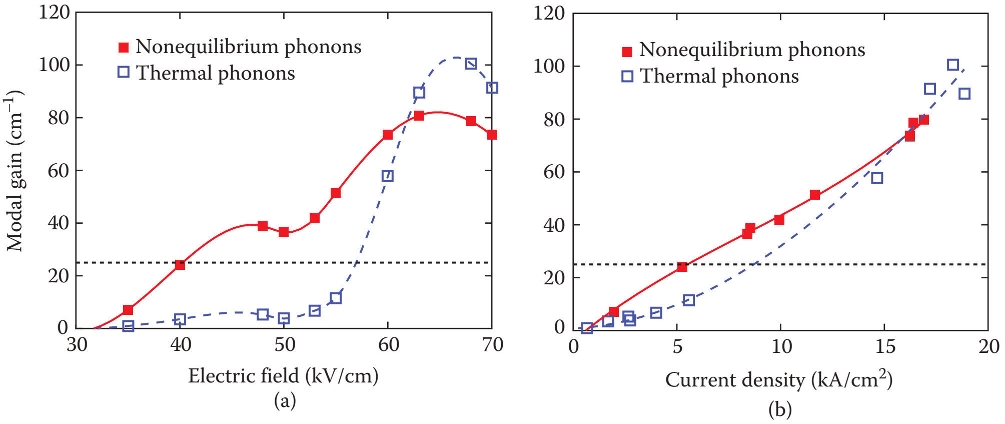


FIGURE 33.12 The modal gain (G_m) of the simulated device with nonequilibrium (filled squares) and thermal (empty squares) phonons as a function of (a) applied electric field and (b) current density *J* at 77 K. Horizontal dotted line shows the total estimated loss of the device.

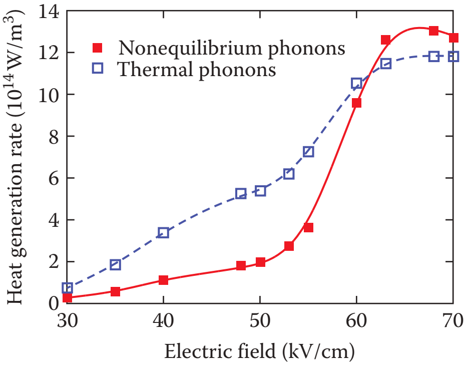


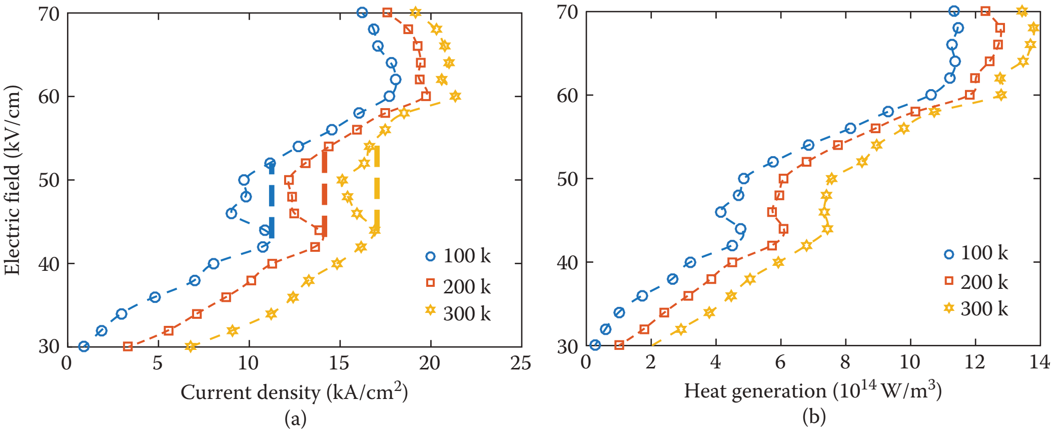
FIGURE 33.13 The heat generation rate of the simulated device as a function of electric field *F* at 77 K with (filled squares) and without (empty squares) nonequilibrium phonons.

33.4.2 Representative Electrothermal Simulation Results

This section serves to illustrate how the described simulation is implemented in practice and what type of information it provides at the single-stage and device levels.

First, the single-stage-coupled simulation has to be performed at different temperatures, as in Figure 33.14a. We noted the calculated *J*-*F* curves show a negative differential conductance region, which is typical for calculations, but generally not observed in experiment. Instead, a flat *J*-*F* dependence is typically recorded [30]. At every temperature and field, we also record the heat generation rate, as depicted in Figure 33.14b.

Second, the thermal model for the whole structure is developed. Considering that growth techniques improve over time, structures grown around the same time period should have similar properties. Since the device studied here was built in 2001 [79], we assume the active core should have similar effective rms roughness Δ to other lattice-matched GaAs/AlAs SLs built around the same time [69,70]. From our



(a)

(b)

FIGURE 33.14 The field versus current density (a) and heat generation rate versus current density (b) characteristics for the simulated device at 100, 200, and 300 K, as obtained from single-stage simulation with nonequilibrium phonons.

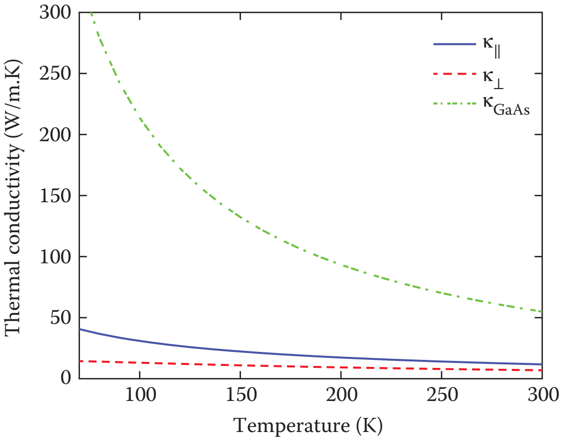


FIGURE 33.15 Calculated in-plane ($\kappa_{\|}$; solid line) and cross-plane ($\kappa_{\perp}$; dashed line) thermal conductivities of the active core along with the bulk thermal conductivity of the GaAs substrate (dash-dotted line). The effective rms roughness Δ is taken to be 5 Å.

previous simulation work on fitting the SL thermal conductivities [48], we choose an effective rms roughness $\Delta = 5$ Å in this calculation. Figure 33.15 shows the calculated thermal conductivities $\kappa_{\|}$ (solid line) and $\kappa_{\perp}$ (dashed line) along with the calculated bulk thermal conductivity (dash-dotted line) for the substrate GaAs.

The structure we considered operated in pulsed mode at 77 K. Depending on the duty cycle, the temperature distribution in the device can differ considerably. Figure 33.16 depicts a typical temperature profile across the device, while Figure 33.17 depicts the profile across the active core alone at duty cycles of 100% (essentially CW lasing, if the device achieved it) and 0.01% (as in experiment [79]). Clearly, CW operation would result in dramatic heating of the active region. Finally, Figure 33.18 shows the J–V curve of the entire simulated device at 77 K with duty cycles of 0.01%, 100%, and as observed in experiment [79].

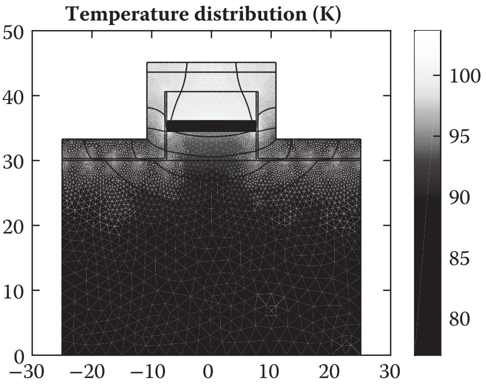


FIGURE 33.16 A typical temperature profile across the structure. At the bottom of the device is a heat sink held at 77 K, while adiabatic boundary conditions are applied elsewhere. The current density is 6 kA/cm^2 and the duty cycle is 100%.

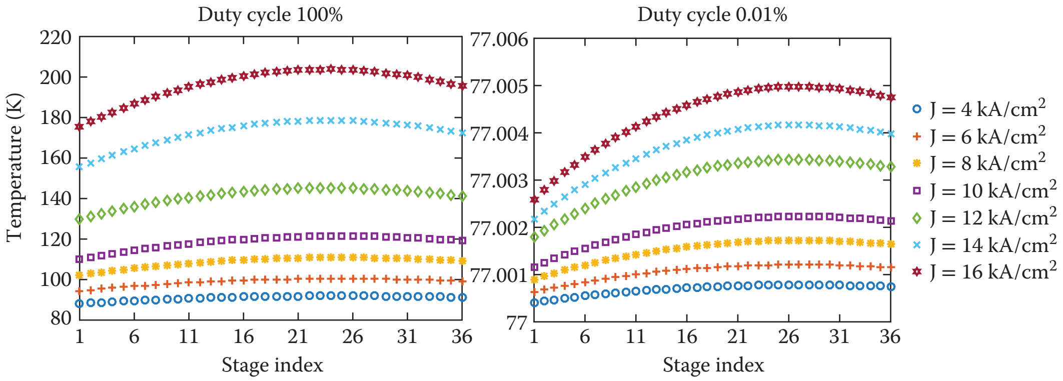


FIGURE 33.17 Temperature profile inside the active region at 100% duty cycle (left) and 0.01% duty cycle (right) for the QCL from [79].

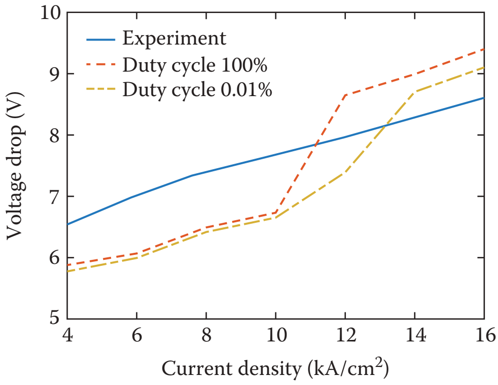


FIGURE 33.18 The current density versus voltage drop for the simulated device in experiment (solid curve) and as calculated at 100% (dashed curve) and 0.01% (dot-dashed curve) duty cycles. The bottom of the device is placed on a heat sink held at 77 K while adiabatic boundary conditions are assumed on the rest of the boundaries (see Section 33.3.2).

33.5 Conclusion

We overviewed electronic and thermal transport simulation of QCLs, as well as recent efforts in device-level electrothermal modeling of these structures, which is appropriate for transport below threshold, where the effects of the optical field are negligible. We specifically focused on mid-IR QCLs in which electronic transport is largely incoherent and can be captured by the EMC technique. The future of QCL modeling, especially for near-RT CW operation, will likely include improvements on several fronts: (1) further development of computationally efficient yet rigorous quantum transport techniques for electronic transport to fully account for coherent transport features that are important in short-wavelength mid-IR devices; (2) a better understanding and better numerical models for describing the role of electron–electron interaction, impurities, and interface roughness on device characteristics; and (3) holistic modeling approaches in which electrons, phonons, and photons are simultaneously and self-consistently captured within a single simulation. The goal of QCL simulation should be nothing less than excellent predictive value of device operation across a range of temperatures and biasing conditions, along with unprecedented insight into the fine details of exciting nonequilibrium physics that underscores the operation of these devices.

Acknowledgment

The authors gratefully acknowledge support by the U.S. Department of Energy, Basic Energy Sciences, Division of Materials Sciences and Engineering, Physical Behavior of Materials Program, Award No. DE-SC0008712. The work was performed using the resources of the UW-Madison Center for High Throughput Computing (CHTC).

References

1. J. Faist, F. Capasso, D. L. Sivco, C. Sirtori, A. L. Hutchinson, and A. Y. Cho. Quantum cascade laser. *Science*, 264(5158):553–556, 1994.
2. R. A. Suris and R. F. Kazarinov. Possibility of the amplification of electromagnetic waves in a semiconductor with a superlattice. *Soviet Physics: Semiconductors*, 5(4):707–709, 1971.
3. K. Y. Cheng. Molecular beam epitaxy technology of III-V compound semiconductors for optoelectronic applications. *Proceedings of the IEEE*, 85(11):1694–1714, 1997.
4. K. H. Goetz, D. Bimberg, H. Jürgensen, J. Selders, A. V. Solomonov, G. F. Glinskii, and M. Razeghi. Optical and crystallographic properties and impurity incorporation of $\mathrm{Ga}_x\mathrm{In}_{1-x}\mathrm{As}$ ($0.44 < x < 0.49$) grown by liquid phase epitaxy, vapor phase epitaxy, and metal organic chemical vapor deposition. *Journal of Applied Physics*, 54(8):4543–4552, 1983.
5. Q. Zhang, F. Q. Liu, W. Zhang, Q. Lu, L. Wang, L. Li, and Z. Wang. Thermal induced facet destructive feature of quantum cascade lasers. *Applied Physics Letters*, 96(14):141117, 2010.
6. D. Botez, J. C. Shin, J. D. Kirch, C. C. Chang, L. J. Mawst, and T. Earles. Multidimensional conduction-band engineering for maximizing the continuous-wave (CW) wallplug efficiencies of mid-infrared quantum cascade lasers. *IEEE Journal of Selected Topics in Quantum Electronics*, 19(4):1200312–1200312, 2013.
7. P. Q. Liu, A. J. Hoffman, M. D. Escarra, K. J. Franz, J. B. Khurgin, Y. Dikmelik, X. Wang, J.-Y. Fan, and C. F. Gmachl. Highly power-efficient quantum cascade lasers. *Nature Photonics*, 4:95–98, 2010.
8. A. J. Hoffman, Y. Yao, and C. F. Gmachl. Mid-infrared quantum cascade laser. *Nature Photonics*, 6(1749–4885):432–439, 2012.
9. Y. Bai, S. R. Darvish, S. Slivken, W. Zhang, A. Evans, J. Nguyen, and M. Razeghi. Room temperature continuous wave operation of quantum cascade lasers with watt-level optical power. *Applied Physics Letters*, 92(10):101105, 2008.

10. A. Lyakh, R. Maulini, A. Tsekoun, R. Go, C. Pflügl, L. Diehl, Q. J. Wang, F. Capasso, and C. Kumar N. Patel. 3 W continuous-wave room temperature single-facet emission from quantum cascade lasers based on nonresonant extraction design approach. *Applied Physics Letters*, 95(14):141113, 2009.
11. N. Bandyopadhyay, Y. Bai, B. Gokden, A. Myzaferi, S. Tsao, S. Slivken, and M. Razeghi. Watt level performance of quantum cascade lasers in room temperature continuous wave operation at $\lambda \sim 3.76$ μm. *Applied Physics Letters*, 97(13):131117, 2010.
12. N. Bandyopadhyay, S. Slivken, Y. Bai, and M. Razeghi. High power, continuous wave, room temperature operation of $\lambda \sim 3.4$ μm and $\lambda \sim 3.55$ μm InP-based quantum cascade lasers. *Applied Physics Letters*, 100(21):212104, 2012.
13. Y. Bai, N. Bandyopadhyay, S. Tsao, S. Slivken, and M. Razeghi. Room temperature quantum cascade lasers with 27% wall plug efficiency. *Applied Physics Letters*, 98(18):181102, 2011.
14. Y. Yu, W. Xiaojun, F. Jen-Yu, and C. F. Gmachl. High performance continuum-to-continuum quantum cascade lasers with a broad gain bandwidth of over 400 cm^{-1}. *Applied Physics Letters*, 97(8):081115, 2010.
15. M. Razeghi, Q. Y. Lu, N. Bandyopadhyay, W. Zhou, D. Heydari, Y. Bai, and S. Slivken. Quantum cascade lasers: From tool to product. *Optics Express*, 23(7):8462–8475, 2015.
16. M. Wienold, M. P. Semtsiv, I. Bayrakli, W. T. Masselink, M. Ziegler, K. Kennedy, and R. Hogg. Optical and thermal characteristics of narrow-ridge quantum-cascade lasers. *Journal of Applied Physics*, 103(8):083113, 2008.
17. H. K. Lee and J. S. Yu. Thermal analysis of short wavelength InGaAs/InAlAs quantum cascade lasers. *Solid-State Electronics*, 54(8):769–776, 2010.
18. J. D. Kirch, J. C. Shin, C. C. Chang, L. J. Mawst, D. Botez, and T. Earles. Tapered active-region quantum cascade lasers ($\lambda = 4.8$ μm) for virtual suppression of carrier-leakage currents. *Electronics Letters*, 48(4):234–235, 2012.
19. D. Botez, C. C. Chang, and L. J. Mawst. Temperature sensitivity of the electro-optical characteristics for mid-infrared ($\lambda = 3$–16 μm)-emitting quantum cascade lasers. *Journal of Physical D: Applied Physics*, 49:043001, 2016.
20. D. Indjin, P. Harrison, R. W. Kelsall, and Z. Ikonić. Self-consistent scattering theory of transport and output characteristics of quantum cascade lasers. *Journal of Applied Physics*, 91(11):9019–9026, 2002.
21. D. Indjin, P. Harrison, R. W. Kelsall, and Z. Ikonic. Influence of leakage current on temperature performance of GaAs/AlGaAs quantum cascade lasers. *Applied Physics Letters*, 81(3):400–402, 2002.
22. A. Mircetic, D. Indjin, Z. Ikonic, P. Harrison, V. Milanovic, and R. W. Kelsall. Towards automated design of quantum cascade lasers. *Journal of Applied Physics*, 97(8):084506, 2005.
23. R. Claudia Iotti and F. Rossi. Nature of charge transport in quantum-cascade lasers. *Physical Review Letters*, 87:146603, 2001.
24. H. Callebaut, S. Kumar, B. S. Williams, Q. Hu, and J. L. Reno. Importance of electron-impurity scattering for electron transport in terahertz quantum-cascade lasers. *Applied Physics Letters*, 84(5):645–647, 2004.
25. X. Gao, D. Botez, and I. Knezevic. X-valley leakage in GaAs-based midinfrared quantum cascade lasers: A Monte Carlo study. *Journal of Applied Physics*, 101(6):063101, 2007.
26. Y. B. Shi and I. Knezevic. Nonequilibrium phonon effects in midinfrared quantum cascade lasers. *Journal of Applied Physics*, 116(12):123105, 2014.
27. H. Willenberg, G. H. Döhler, and J. Faist. Intersubband gain in a Bloch oscillator and quantum cascade laser. *Physical Review B*, 67:085315, 2003.
28. S. Kumar and Q. Hu. Coherence of resonant-tunneling transport in terahertz quantum-cascade lasers. *Physical Review B*, 80:245316, 2009.
29. C. Weber, A. Wacker, and A. Knorr. Density-matrix theory of the optical dynamics and transport in quantum cascade structures: The role of coherence. *Physical Review B*, 79:165322, 2009.

30. E. Dupont, S. Fathololoumi, and H. C. Liu. Simplified density-matrix model applied to three-well terahertz quantum cascade lasers. *Physical Review B*, 81:205311, 2010.
31. R. Terazzi and J. Faist. A density matrix model of transport and radiation in quantum cascade lasers. *New Journal of Physics*, 12(3):033045, 2010.
32. O. Jonasson, F. Karimi, and I. Knezevic. Partially coherent electron transport in terahertz quantum cascade lasers based on a Markovian master equation for the density matrix. *Journal of Computational Electronics*, 15:1192, 2016.
33. O. Jonasson, S. Mei, F. Karimi, J. Kirch, D. Botez, L. Mawst, and I. Knezevic. Quantum transport simulation of high-power 4.6-μm quantum cascade lasers. *Photonics*, 3(2):38, 2016.
34. S.-C. Lee and A. Wacker. Nonequilibrium Green's function theory for transport and gain properties of quantum cascade structures. *Physical Review B*, 66:245314, 2002.
35. M. Bugajski, P. Gutowski, P. Karbownik, A. Kolek, G. Hałdaś, K. Pierściński, D. Pierścińska, J. Kubacka-Traczyk, I. Sankowska, A. Trajnerowicz, K. Kosiel, A. Szerling, J. Grzonka, K. Kurzydłowski, T. Slight, and W. Meredith. Mid-IR quantum cascade lasers: Device technology and non-equilibrium Green's function modeling of electro-optical characteristics. *Physica Status Solidi (b)*, 251(6):1144–1157, 2014.
36. A. Kolek, G. Haldas, M. Bugajski, K. Pierscinski, and P. Gutowski. Impact of injector doping on threshold current of mid-infrared quantum cascade laser–non-equilibrium greens function analysis. *IEEE Journal of Selected Topics in Quantum Electronics*, 21(1):124–133, 2015.
37. O. Jonasson and I. Knezevic. Dissipative transport in superlattices within the Wigner function formalism. *Journal of Computational Electronics*, 14:879–887, 2015.
38. A. Matyas, P. Lugli, and C. Jirauschek. Photon-induced carrier transport in high efficiency midinfrared quantum cascade lasers. *Journal of Applied Physics*, 110(1):013108, 2011.
39. M. Lindskog, J. M. Wolf, V. Trinite, V. Liverini, J. Faist, G. Maisons, M. Carras, R. Aidam, R. Ostendorf, and A. Wacker. Comparative analysis of quantum cascade laser modeling based on density matrices and non-equilibrium Green's functions. *Applied Physics Letters*, 105(10):103106, 2014.
40. R. C. Iotti and F. Rossi. Microscopic theory of semiconductor-based optoelectronic devices. *Reports on Progress in Physics*, 68(11):2533, 2005.
41. C. Jirauschek and P. Lugli. Monte-Carlo-based spectral gain analysis for terahertz quantum cascade lasers. *Journal of Applied Physics*, 105(12):123102, 2009.
42. C. Jirauschek and T. Kubis. Modeling techniques for quantum cascade lasers. *Applied Physics Reviews*, 1(1):011307, 2014.
43. C. Jirauschek. Monte Carlo study of carrier-light coupling in terahertz quantum cascade lasers. *Applied Physics Letters*, 96(1), 2010.
44. A. Wacker, M. Lindskog, and D. O. Winge. Nonequilibrium Green's function model for simulation of quantum cascade laser devices under operating conditions. *IEEE Journal of Selected Topics in Quantum Electronics*, 19(5):1–11, 2013.
45. C. A. Evans, D. Indjin, Z. Ikonic, P. Harrison, M. S. Vitiello, V. Spagnolo, and G. Scamarcio. Thermal modeling of terahertz quantum-cascade lasers: Comparison of optical waveguides. *IEEE Journal of Quantum Electronics*, 44(7):680–685, 2008.
46. H. K. Lee and J. S. Yu. Thermal effects in quantum cascade lasers at $\lambda \sim 4.6$ μm under pulsed and continuous-wave modes. *Applied Physics B*, 106(3):619–627, 2012.
47. Y. B. Shi, Z. Aksamija, and I. Knezevic. Self-consistent thermal simulation of GaAs/$\mathrm{Al_{0.45}Ga_{0.55}As}$ quantum cascade lasers. *Journal of Computational Electronics*, 11(1):144–151, 2012.
48. S. Mei and I. Knezevic. Thermal conductivity of III-V semiconductor superlattices. *Journal of Applied Physics*, 118(17):175101, 2015.
49. Y. B. Shi, S. Mei, O. Jonasson, and I. Knezevic. Modeling quantum cascade lasers: Coupled electron and phonon transport far from equilibrium and across disparate spatial scales. *Fortschritte der Physik – Progress of Physics*, 1600084, 2016. doi:10.1002/prop.201600084

50. S.-C. Lee, F. Banit, M. Woerner, and A. Wacker. Quantum mechanical wavepacket transport in quantum cascade laser structures. *Physical Review B*, 73:245320, 2006.
51. K. Faist, D. Hofstetter, M. Beck, T. Aellen, M. Rochat, and S. Blaser. Bound-to-continuum and two-phonon resonance, quantum-cascade lasers for high duty cycle, high-temperature operation. *IEEE Journal of Quantum Electronics*, 38(6):533–546, 2002.
52. M. Yamanishi, T. Edamura, K. Fujita, N. Akikusa, and H. Kan. Theory of the intrinsic linewidth of quantum-cascade lasers: Hidden reason for the narrow linewidth and line-broadening by thermal photons. *IEEE Journal of Quantum Electronics*, 44(1):12–29, 2008.
53. X. Gao. Monte Carlo Simulation of Electron Dynamics in Quantum Cascade Lasers. PhD thesis, University of Wisconsin–Madison, 2008.
54. Y. Shi. Electrothermal Simulation of Quantum Cascade Lasers. PhD thesis, University of Wisconsin–Madison, 2015.
55. H. Callebaut and Q. Hu. Importance of coherence for electron transport in terahertz quantum cascade lasers. *Journal of Applied Physics*, 98(10):104505, 2005.
56. A. Wacker. Semiconductor superlattices: A model system for nonlinear transport. *Physics Reports*, 357(1):1–111, 2002.
57. X. Gao, D. Botez, and I. Knezevic. X-valley leakage in GaAsAlGaAs quantum cascade lasers. *Applied Physics Letters*, 89(19):191119, 2006.
58. S. L. Chuang. *Physics of Optoelectronic Devices*. New York, NY: Wiley, 1995.
59. X. Gao, D. Botez, and I. Knezevic. Confined phonon scattering in multivalley Monte Carlo simulation of quantum cascade lasers. *Journal of Computational Electronics*, 7(3):209–212, 2008.
60. P. Lugli, P. Bordone, L. Reggiani, M. Rieger, P. Kocevar, and S. M. Goodnick. Monte Carlo studies of nonequilibrium phonon effects in polar semiconductors and quantum wells. I. Laser photoexcitation. *Physical Review B*, 39:7852–7865, 1989.
61. B. K. Ridley. The electron-phonon interaction in quasi-two-dimensional semiconductor quantum-well structures. *Journal of Physics C: Solid State Physics*, 15(28):5899, 1982.
62. J. T. Lü and J. C. Cao. Monte Carlo simulation of hot phonon effects in resonant-phonon-assisted terahertz quantum-cascade lasers. *Applied Physics Letters*, 88(6):061119, 2006.
63. S. Usher and G. P. Srivastava. Theoretical study of the anharmonic decay of nonequilibrium LO phonons in semiconductor structures. *Physical Review B*, 50:14179–14186, 1994.
64. E. Pop, S. Sinha, and K. E. Goodson. Heat generation and transport in nanometer-scale transistors. *Proceedings of the IEEE*, 94(8):1587–1601, 2006.
65. D. G. Cahill, P. V. Braun, G. Chen, D. R. Clarke, S. Fan, K. E. Goodson, P. Keblinski, W. P. King, G. D. Mahan, A. Majumdar, H. J. Maris, S. R. Phillpot, E. Pop, and L. Shi. Nanoscale thermal transport. II. 2003–2012. *Applied Physics Reviews*, 1(1):011305, 2014.
66. T. Yao. Thermal properties of AlAs/GaAs superlattices. *Applied Physics Letters*, 51(22):1798–1800, 1987.
67. G. Chen, C. L. Tien, X. Wu, and J. S. Smith. Thermal diffusivity measurement of GaAs/AlGaAs thin-film structures. *Journal of Heat Transfer*, 116:325–331, 1994.
68. X. Y. Yu, G. Chen, A. Verma, and J. S. Smith. Temperature dependence of thermophysical properties of GaAs/AlAs periodic structure. *Applied Physics Letters*, 67(24):3554–3556, 1995.
69. W. S. Capinski and H. J. Maris. Thermal conductivity of GaAs/AlAs superlattices. *Physica B*, 219–220:699–701, 1996.
70. W. S. Capinski, H. J. Maris, T. Ruf, M. Cardona, K. Ploog, and D. S. Katzer. Thermal conductivity measurements of GaAs/AlAs superlattices using a picosecond optical pump-and-probe technique. *Physical Review B*, 59:8105–8113, 1999.
71. G. Chen. Size and interface effects on thermal conductivity of superlattices and periodic thin-film structures. *Journal of Heat Transfer*, 119(2):220, 1997.
72. G. Chen. Thermal conductivity and ballistic-phonon transport in the cross-plane direction of superlattices. *Physical Review B*, 57:14958–14973, 1998.

73. S. T. Huxtable, A. R. Abramson, C. L. Tien, A. Majumdar, C. LaBounty, X. Fan, G. Zeng, J. E. Bowers, A. Shakouri, and E. T. Croke. Thermal conductivity of Si/SiGe and SiGe/SiGe superlattices. *Applied Physics Letters*, 80(10):1737–1739, 2002.
74. Y. Wang, H. Huang, and X. Ruan. Decomposition of coherent and incoherent phonon conduction in superlattices and random multilayers. *Physical Review B*, 90:165406, 2014.
75. Y. Wang, C. Gu, and X. Ruan. Optimization of the random multilayer structure to break the random-alloy limit of thermal conductivity. *Applied Physics Letter*, 106:073104, 2015.
76. Z. Aksamija and I. Knezevic. Thermal conductivity of Si1$-x$Gex/Si1$-y$Ge$y$superlattices: Competition between interfacial and internal scattering. *Physical Review B*, 88:155318, 2013.
77. E. T. Swartz and R. O. Pohl. Thermal boundary resistance. *Reviews of Modern Physics*, 61:605–668, 1989.
78. A. Lops, V. Spagnolo, and G. Scamarcio. Thermal modeling of GaInAs/AlInAs quantum cascade lasers. *Journal of Applied Physics*, 100(4):043109, 2006.
79. H. Page, C. Becker, A. Robertson, G. Glastre, V. Ortiz, and C. Sirtori. 300 k operation of a GaAs-based quantum-cascade laser at $\lambda \approx 9$ μm. *Applied Physics Letters*, 78(22):3529, 2001.
80. V. Spagnolo, A. Lops, G. Scamarcio, M. S. Vitiello, and C. Di Franco. Improved thermal management of mid-IR quantum cascade lasers. *Journal of Applied Physics*, 103(4):043103, 2008.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

34

Vertical-Cavity Surface-Emitting Lasers

34.1 Introduction 261
34.2 Physical Phenomena 262
34.3 Light Confinement in VCSELs 263
34.4 Fundamental Equations 264
34.5 VCSEL Structure 265
34.6 One-Dimensional Scalar Approach 266
34.7 Two-Dimensional Scalar Approach 270
34.8 Three-Dimensional Vectorial Approach 273
34.9 Impact of Oxide Aperture Diameter and Position 277
34.10 Conclusion 278

Tomasz Czyszanowski

Leszek Frasunkiewicz

and

Maciej Dems

34.1 Introduction

The history of vertical-cavity surface-emitting lasers (VCSELs) can be traced back almost as far as that of semiconductor lasers themselves. In 1965, Melngailis demonstrated coherent emission parallel to the direction of the current in an $n^{+}pp^{+}$ InSb diode laser.[1] The polished surface on the emission side of InSb and the gold current contact on its opposite side provided sufficient optical feedback to enhance laser emission. In this very early work, a design was proposed for the integration in an array that would provide coherent emission over a large area, with a small beam angle. Unfortunately, this idea was forgotten and not rediscovered until 1979 by Soda et al.,[2] who demonstrated a surface-emitting InP-based diode laser with two metallic mirrors. Further development was initiated by Ogura et al.,[3] who introduced highly reflective distributed Bragg reflectors (DBRs) in 1983.

Unlike typical edge-emitting lasers, in VCSELs the optical cavity is formed between the mirrors above and below the active region (Figure 34.1). The laser light resonates in the vertical direction. Inside the laser structure, the light passes the active region in the vertical direction—i.e., gain is provided over a short distance only and the amplification per photon round-trip is small. Therefore, the mirrors must be highly reflective (over 99% for the emitting mirror and almost 100% for the opposite one) so that the photons make many round-trips before they are emitted. High reflectivity is provided by DBRs composed of stacks of two alternating layers with high refractive index contrast. With quarter-wavelength layer thicknesses, the reflected waves from all DBR interfaces add up constructively, allowing for a total DBR reflectance of over 99%. Such layers are composed of 50 or more semiconductor layers. The current is injected vertically, similar to edge-emitting devices. However, in VCSELs, optical resonance occurs in the vertical direction.

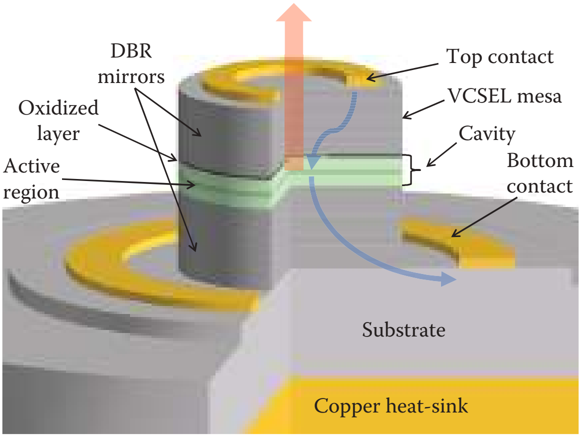


FIGURE 34.1 Intersection of a VCSEL structure showing key construction elements. The arrows inside VCSEL structure illustrate the path of the current flow and vertical arrow the emitted beam.

Optical resonance and current injection taking place in the same direction is the main weakness of these devices. Theoretically, preferential injection of the carriers to the active region may be achieved by placing two electrodes on opposite sides of the laser. However, in such configuration, the stimulated light resonates between the mirrors and no emission occurs, since the metal contacts are highly absorptive. To enable emission, a ring contact is typically used on the top DBR, through which the light is emitted. However, current flows through the shortest path, so maximal stimulated emission overlaps with the top contact, leading to high absorption in the metal and deterioration of the emission beam. To overcome this problem, the carriers must be forced to flow through the central part of the active region. This can be achieved either by selective proton implantation or by selective oxidation of the lateral regions in the VCSEL. Both of these approaches turn the semiconductor layers into electrical insulators. With selective processes, the central part of the VCSEL is conductive and radiative recombination takes place within the central part of the active region. Both methods also allow for cylindrical aperture formation, supporting cylindrical beam emission that cannot be achieved with typical edge-emitting lasers. Such symmetry of the aperture, together with the use of ring contacts, favors the choice of cylindrical geometry for numerical approaches, which are used to simulate the phenomena that take place in VCSELs during laser operation.

34.2 Physical Phenomena

VCSELs are very complex systems due to their multilayered structure (comprising in excess of a hundred layers). Often with nonplanar or buried-type architectures, they feature many heterojunctions, graded layers, quantum wells (QWs), quantum dots (QDs) or quantum wires, superlattices, oxide and oxidized layers, mesa structures, photonic microstructures, and so on. To analyze the operation of VCSELs precisely, therefore, one needs to consider four main classes of phenomena in the device: electrical, thermal, optical, and recombination, which together create a nonlinear network of mutual interrelations. Each class of phenomena is responsible for the following:

1. Optical: Cavity modes, their modal gain and loss; emission wavelengths and optical field distributions within the laser cavity; formation of the output beam
2. Electrical: Current spreading between the top and bottom contacts through the centrally located active region; injection of carriers of both polarities into the active region and their subsequent radiative or nonradiative recombination after radial out-diffusion in the active region; overbarrier carrier leakage

3. Thermal: Generation of heat flux (nonradiative recombination, reabsorption of spontaneous radiation, as well as volume and barrier Joule heating); heat spreading within heat sinks
4. Recombination: Material gain as a result of stimulated recombination (to determine the active region optical gain spectrum); nonradiative monomolecular and Auger recombinations

The main effects induced by nonlinear interaction between the models are thermal focusing and emitted power rollover. The first is related to the dependence of the refractive index on temperature. The heat induced predominantly in the central part of the active region dissipates in all directions, but mostly toward the bottom heat sink. This, in turn, induces different temperatures in the laser cavity. Increasing the temperature of semiconductor materials triggers an increase in their refractive index. The most significant increase of the refractive index is in the central part of the laser, which is heated most intensely. This causes the waveguiding effect, whereby the modes squeeze radially into the central part of the laser. On the one hand, stronger focusing of the modes in the active region increases their interaction with carriers that are injected into the central part of the active region, reducing the current necessary for laser operation (threshold current). On the other hand, if there are more optical modes confined and excited within the central part, this contributes to a broadening of the emission spectrum, since each mode relates to a slightly different emission wavelength. The quality of the output beam also deteriorates, since interference between different modes occurs at various emission angles.

The second effect—emission power rollover—limits the emitted power of the laser, mostly due to thermal effects. The active regions of VCSELs are designed in such a way that the wavelength corresponding to maximal material gain overlaps with the wavelength of the optical resonance induced by the VCSEL cavity. A change in the temperature of the device, due to heat dissipation or ambient temperature, may modify the bandgap and consequently change the wavelength corresponding to the maximum of the material gain. The optical modes do not follow the maximum of the material gain, as is the case in edge-emitting lasers, but follow the optical resonance induced by the cavity. The resonance wavelength also changes with temperature, since the refractive indices of the cavity change. The changes in maximal material gain and of cavity resonance, however, are of different proportions. Typically, the change in the gain spectrum is more sensitive to temperature changes than in the case of optical resonance. Hence, material gain can be insufficient at resonant wavelength to sustain laser operation for higher operational temperatures and the laser ceases to emit stimulated radiation. This effect is manifested as a reduction in emitted power as the injected current increases.

To simulate such complex behaviors, a comprehensive model including electrical, thermal, recombination, and optical submodels is needed. Three of these submodels can be solved in a very similar manner to the optoelectronic devices considered previously in this book. However, special attention must be given to the optical model, since electromagnetic field resonance takes place in the vertical direction, which is orthogonal to the resonance that occurs in typical edge-emitting lasers. In the remainder of this chapter, we have therefore focused on optical simulation.

34.3 Light Confinement in VCSELs

VCSELs inherently emit in single longitudinal mode due to their very short cavities. Controlling the modes in the transversal direction is more complicated. In typical VCSELs made from arsenide-based materials, well-established wet oxidation technology can be used,[4] which produces the waveguide effect through a selectively oxidized Al-rich layer that also plays the role of an electric insulator for improved current funneling to the active region. Another less material dependent technique is proton implantation,[5] which defines highly resistive regions. The waveguide effect is produced by thermal focusing induced by thermal increase of the refractive index in regions where nonradiative recombination and Joule heating occur. Other methods of transverse mode confinement require more sophisticated and expensive technologies. Possible solutions include micro-optical structures, such as surface relief,[6] antiresonant patterning,[7] tunnel junction patterning,[8] photonic crystals[9] and highcontrast gratings.[10] Each of these methods requires a separate analytical approach, since different phenomena are responsible for mode confinement: the waveguide effect, total interior reflections, Bragg reflections or Fano resonance. Since not all those phenomena can be observed in all VCSEL constructions with equal intensity, numerical models should be properly chosen to provide a balance between generalization and efficiency.

34.4 Fundamental Equations

Interaction between a VCSEL and the optical field is governed by Maxwell's equations. The boundary conditions correspond to an optical field that decays to 0 in infinity. These boundary conditions mean that Maxwell's equations turn into an eigenvalue problem. Its solutions are discrete eigenvalues and the corresponding eigenvectors. Eigenvalues of Maxwell's equations are vacuum wavevectors (k_0), corresponding angular frequencies (ω), or wavelengths in a vacuum (λ). Eigenvectors are distributions of electric (**E**) and magnetic (**H**) fields of electromagnetic waves corresponding to particular eigenvalues.

The three-dimensional electric **E** and magnetic **H** field vectors satisfy the following partial differential equations,[11] which correspond to Faraday's and Amper's laws, respectively:

$$\nabla \times \mathbf{E} = -\mu\mu_0 \frac{\partial \mathbf{H}}{\partial t}, \tag{34.1}$$

$$\nabla \times \mathbf{H} = \varepsilon\varepsilon_0 \frac{\partial \mathbf{E}}{\partial t}, \tag{34.2}$$

where μ and μ_0 are the magnetic permittivity of the material and the vacuum, respectively, and ε and ε_0 are analogous dielectric constants.

Let us consider a monochromatic wave only, of which the angular frequency is ω. The time dependency of **E** and **H** is given by the following relations:

$$\mathbf{E} = \mathbf{E} \exp(i\omega t), \tag{34.3}$$

$$\mathbf{H} = \mathbf{H} \exp(i\omega t). \tag{34.4}$$

Embedment of Equations 34.3 and 34.4 to Equations 34.1 and 34.2 results in a time-independent set of equations:

$$\nabla \times \mathbf{E} = -i\omega\mu\mu_0 \mathbf{H}, \tag{34.5}$$

$$\nabla \times \mathbf{H} = i\omega\varepsilon\varepsilon_0 \mathbf{E}. \tag{34.6}$$

We assume that μ is uniform and equal to 1, which is always true for the dielectrics and nonferromagnetic metals typically used in VCSELs. Combining Equations 34.5 and 34.6, we can reduce the set of six equations for all vector components to a set of three equations for the electric field only:

$$\nabla \times \nabla \times \mathbf{E} = k_0^2 n^2 \mathbf{E}, \tag{34.7}$$

where k_0 is the wavenumber of the mode in a vacuum and

$$\sqrt{\varepsilon} = n = n_{\mathrm{re}} + i n_{\mathrm{im}} \tag{34.8}$$

is the complex refractive index. Its imaginary part accounts for the loss and the gain of the medium in which the wave propagates. The relation between the imaginary part of n and losses α is given by

$$n_{\mathrm{im}} = -\frac{\alpha}{2\,k_0}. \tag{34.9}$$

Equation 34.9 can be rewritten in an equivalent form:

$$\nabla(\nabla\cdot\mathbf{E})-\nabla^2\mathbf{E}=\omega^2\mu\varepsilon\mathbf{E},$$
$$\nabla(\nabla\cdot\mathbf{E})-\nabla^2\mathbf{E}=k_0^2n^2\mathbf{E}. \tag{34.10}$$

VCSELs are typically processed to be circular in the epitaxial plane to support symmetric beam emission. This geometry allows the complexity of Equation 34.7 to be reduced by expressing it in a cylindrical coordinate system where the angular dependence of **E** can be given in simple analytical terms. However, in recent years, numerous confinement approaches based on photonic microstructures have been proposed. Many of these structures cannot be analyzed using the cylindrical coordinate system and therefore the three-dimensional Cartesian system must be used. We will therefore now show how Equations 34.5, 34.6, and 34.10 can be solved using three approaches of incremental complexity. The first approach is based on a one-dimensional analysis of the optical field along the optical axis. The second assumes that the VCSEL has cylindrical geometry and considers both the scalar optical field in two dimensions along the optical axis and the radius of the VCSEL. The third considers the vectorial optical field in the three-dimensional Cartesian coordinate system.

34.5 VCSEL Structure

Figure 34.2 shows the VCSEL structure with construction details listed in Table 34.1. This structure was used in References[12,13] as a benchmark for comparative analysis of optical numerical models, however it must be stressed here, the structure is significantly simplified in comparison to real-life VCSELs. The structure consists of one wavelength-long GaAs cavity with a 5-nm active GaAs layer at the antinode of the standing wave and a selectively oxidized AlAs layer ~15 nm in thickness. The cavity is sandwiched between 24 and 29.5 pairs of GaAs/AlAs DBRs. In comparison to a real-life VCSEL, there is no vertical carrier confinement that would capture and support optical recombination of the carriers. There is also no doping scheme that would allow efficient injection of the current to the active region.

In what follows, we will be using two coordinate systems: Cartesian and cylindrical (Figure 34.1). These share a z-axis along the optical axis of the device. The Cartesian plane $x-y$ and cylindrical plane $r-\varphi$ are parallel to the epitaxial layers of the VCSEL, and the x-axis relates to $\varphi=0$.

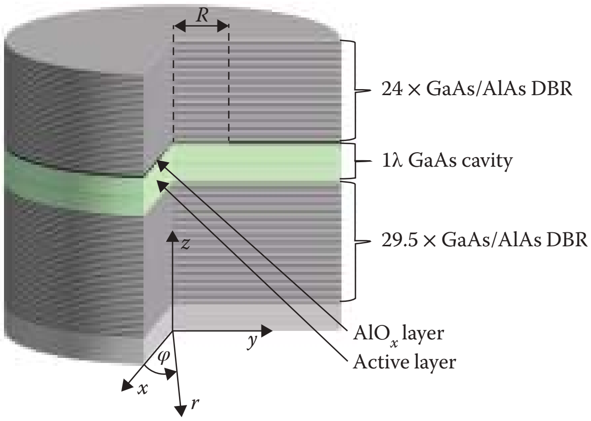


FIGURE 34.2 Intersection of the VCSEL structure with assigned radius (R) of the oxide aperture and definition of Cartesian (x, y, z) and cylindrical (r, φ, z) coordinate systems. DBR = distributed Bragg reflectors.

TABLE 34.1 Refractive Indices and Thicknesses of VCSEL Layers

Refractive Index	Thickness (nm)	Repetition
1.0	∞	–
3.53	69.49	24.5
3.08	79.63	
3.08	63.71	–
2.95 for $r < R$ 1.6 for $r > R$	15.93	–
3.53	136.49	–
3.53 (active)	5.00	–
3.53	136.49	–
3.08	79.63	29.5
3.53	69.49	
3.53	∞	–

Note: The two outer layers are assumed to be infinite.

Vectors **E** and **H** are given in the Cartesian system as

$$\mathbf{E} = \begin{bmatrix} E_x \\ E_y \\ E_z \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} H_x \\ H_y \\ H_z \end{bmatrix}, \tag{34.11}$$

and in the cylindrical system as

$$\mathbf{E} = \begin{bmatrix} E_r \\ E_\phi \\ E_z \end{bmatrix}, \quad \mathbf{H} = \begin{bmatrix} H_r \\ H_\varphi \\ H_z \end{bmatrix}. \tag{34.12}$$

34.6 One-Dimensional Scalar Approach

The simplest approach is to consider the distribution of light within the VCSEL along the z-axis only, assuming an electromagnetic field in the form of a plane wave within the device and that the layers are uniform in the lateral direction. This extremely simplified approach allows only the longitudinal modes to be determined. However, it is of tremendous importance in VCSEL wafers designing, as it enables one to determine the size of the cavity and the number of necessary DBRs, as well as the positions of active or highly doped layers.

The assumption of uniform layers corresponds to the Gauss's law relating to electrical fields:

$$\nabla \cdot \varepsilon \mathbf{E} = 0. \tag{34.13}$$

Since ε is constant within any particular layer of the VCSEL, Gauss's law can be reduced to the following form:

$$\nabla \cdot \mathbf{E} = 0, \tag{34.14}$$

which simplifies Equation 34.10:

$$\nabla^2 \mathbf{E} = k_0^2 n^2 \mathbf{E}. \tag{34.15}$$

The operator ∇^2 can be expressed in either coordinate system. However, for the sake of clarity in what follows, we shall express it in the cylindrical system:

$$\nabla^2 \equiv \begin{bmatrix} \partial_r^2 + \frac{1}{r}\partial_r + \frac{1}{r^2}\partial_\varphi^2 + \partial_z^2 & 0 & 0 \\ 0 & \partial_r^2 + \frac{1}{r}\partial_r + \frac{1}{r^2}\partial_\varphi^2 + \partial_z^2 & 0 \\ 0 & 0 & \partial_r^2 + \frac{1}{r}\partial_r + \frac{1}{r^2}\partial_\varphi^2 + \partial_z^2 \end{bmatrix}. \tag{34.16}$$

Substituting Equation 34.16 to Equation 34.15, one obtains three independent equations for each field component of **E**:

$$\left(\partial_r^2 + \frac{1}{r}\partial_r + \frac{1}{r^2}\partial_\varphi^2 + \partial_z^2\right) E_u = k_0^2 n^2 E_u, \tag{34.17}$$

where $u = r, \varphi, z$. An additional assumption with regard to the plane wave imposes $E_r = \text{const}$ and $E_\varphi = \text{const}$, which simplifies Equation 34.17 to a one-dimensional expression:

$$\partial_z^2 E_u = k_0^2 n^2 E_u. \tag{34.18}$$

Each of the equations can be solved separately; although by imposing the assumption of layer uniformity, we lose information on the relations between the components of the electric field. Hence, it is useful to assume that when linearly polarized light propagates along the z-axis then

$$E_z = 0$$

and

$E_\varphi = E_r \cot\varphi$ for x polarization or

$E_\varphi = E_r \tan\varphi$ for y polarization.

To solve Equation 34.18, we used the transfer matrix method (TMM),[14] which assumes that within the jth layer the solution of Equation 34.18 can be expressed by the function (Figure 34.3):

$$E_r^j = A_j \exp\left(ik_j z\right) + B_j \exp\left(-ik_j z\right), \tag{34.19}$$

where $k_j = k_0 n_j$ and E_r is continuous and smooth at the interfaces between the consecutive layers of the VCSEL (Figure 34.3). This leads to the following relations:

$$A_j \exp\left(ik_j d_j\right) + B_j \exp\left(-ik_j d_j\right) = A_{j+1} \exp\left(ik_{j+1} 0\right) + B_{j+1} \exp\left(-ik_{j+1} 0\right), \tag{34.20}$$

$$\frac{\mathrm{d}}{\mathrm{d}z}\left(A_j \exp\left(ik_j z\right) + B_j \exp\left(-ik_j z\right)\right)\bigg|_{z=d_j} = \frac{\mathrm{d}}{\mathrm{d}z}\left(A_{j+1} \exp\left(ik_{j+1} z\right) + B_{j+1} \exp\left(-ik_{j+1} z\right)\right)\bigg|_{z=0}, \tag{34.21}$$

at neighboring layers expressed in local coordinates within layers; k_0 is unknown and corresponds to the vacuum wavevector of the mode.

As a consequence, we can write the relation that transforms E_r between the layers as follows:

$$\begin{matrix} A_j \frac{k_{j+1}+k_j}{2k_{j+1}} \exp\left(ik_j d_j\right) + \frac{k_{j+1}-k_j}{2k_{j+1}} B_j \exp\left(-ik_j d_j\right) = A_{j+1} \\ A_j \frac{k_{j+1}-k_j}{2k_{j+1}} \exp\left(ik_j d_j\right) + \frac{k_{j+1}+k_j}{2k_{j+1}} B_j \exp\left(-ik_j d_j\right) = B_{j+1} \end{matrix}, \tag{34.22}$$

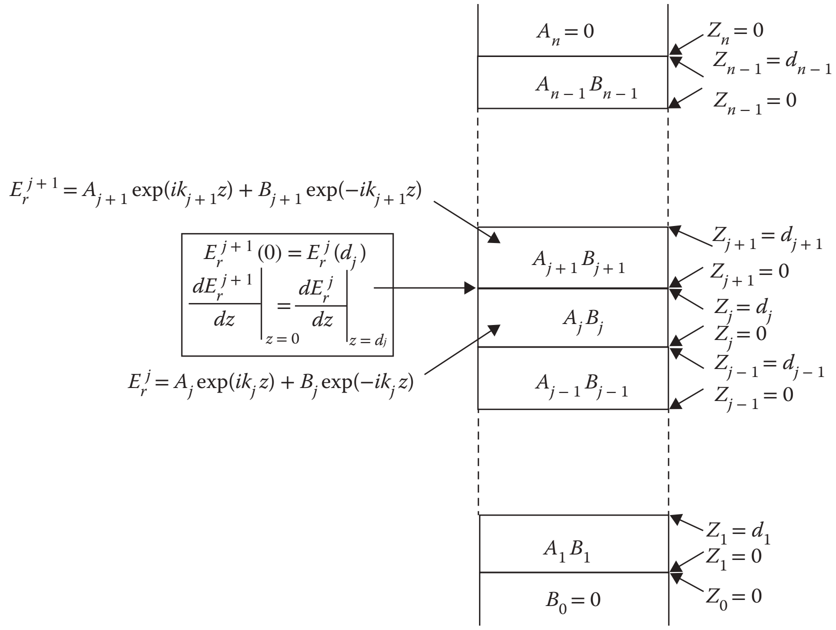

$A_n = 0$
$Z_n = 0$
$Z_{n-1} = d_{n-1}$
$A_{n-1}B_{n-1}$
$Z_{n-1} = 0$
$E_r^{j+1} = A_{j+1}\exp(ik_{j+1}z) + B_{j+1}\exp(-ik_{j+1}z)$
$E_r^{j+1}(0) = E_r^j(d_j)$
$\left.\frac{dE_r^{j+1}}{dz}\right|_{z=0} = \left.\frac{dE_r^j}{dz}\right|_{z=d_j}$
$Z_{j+1} = d_{j+1}$
$A_{j+1}B_{j+1}$
$Z_{j+1} = 0$
A_jB_j
$Z_j = d_j$
$Z_j = 0$
$Z_{j-1} = d_{j-1}$
$E_r^j = A_j\exp(ik_jz) + B_j\exp(-ik_jz)$
$A_{j-1}B_{j-1}$
$Z_{j-1} = 0$
$Z_1 = d_1$
A_1B_1
$Z_1 = 0$
$B_0 = 0$
$Z_0 = 0$

FIGURE 34.3 Schematic view of structure illustrating parameters and variables used in TMM algorithm.

and this can be given in a matrix form:

$$\begin{bmatrix} A_{j+1} \\ B_{j+1} \end{bmatrix} = \begin{bmatrix} \frac{k_{j+1}+k_j}{2k_{j+1}} \exp\left(ik_jd_j\right) & \frac{k_{j+1}-k_j}{2k_{j+1}} \exp\left(-ik_jd_j\right) \\ \frac{k_{j+1}-k_j}{2k_{j+1}} \exp\left(ik_jd_j\right) & \frac{k_{j+1}+k_j}{2k_{j+1}} \exp\left(-ik_jd_j\right) \end{bmatrix} \begin{bmatrix} A_j \\ B_j \end{bmatrix} = \mathbf{M_j} \begin{bmatrix} A_j \\ B_j \end{bmatrix}. \tag{34.23}$$

Matrix $\mathbf{M}_j$ connects the amplitudes of the electric field from layers j and $j + 1$. All layers of the VCSEL can be combined using the recursive relation:

$$\begin{bmatrix} A_n \\ B_n \end{bmatrix} = \prod_{j=0}^{n} \mathbf{M}_j \begin{bmatrix} A_0 \\ B_0 \end{bmatrix} = \mathbf{M} \begin{bmatrix} A_0 \\ B_0 \end{bmatrix}. \tag{34.24}$$

Boundary conditions impose that the electric field decays to infinity, so the exponential elements responsible for infinite values of E in $+\infty$ and $-\infty$ must be eliminated:

$$\begin{bmatrix} 0 \\ B_n \end{bmatrix} = \mathbf{M} \begin{bmatrix} A_0 \\ 0 \end{bmatrix} = \begin{bmatrix} m_{11} & m_{12} \\ m_{21} & m_{22} \end{bmatrix} \begin{bmatrix} A_0 \\ 0 \end{bmatrix}. \tag{34.25}$$

This relation implies that

$$m_{11}\left(k_0\right) = 0, \tag{34.26}$$

which is the condition in the numerical algorithm responsible for finding the longitudinal modes of the VCSEL. k_0, which fulfills the condition (Equation 34.26), corresponds to the longitudinal mode of the VCSEL and is interpreted as the wavenumber of the longitudinal mode in the vacuum. Simple relations give the emitted wavelength as well as the frequency of the mode. The imaginary part of k_0 corresponds to the total losses or gain of the mode. If $\mathrm{im}(k_0) \geq 0$, the mode reaches the threshold and can be observed in the emitted spectrum.

The vertical mode distribution for the structure detailed in Table 34.1, calculated using TMM, is given in Figure 34.4. Mode intensity is defined as follows:

$$I = E_u E_u^*, \tag{34.27}$$

where E^* is a complex conjugate of the electric field of an electromagnetic wave. Figure 34.4 shows typical oscillatory decay of optical field intensity within the top and bottom DBRs. In the simplified example considered here, in which the cavity is uniform, optical intensity is maximal and the amplitude is constant in the cavity. Typically, VCSELs are designed to ensure the coincidence of optical field intensity and the active region to maximize stimulated emission.

Using this simple approach, one can design the dominant wavelength of the VCSEL wafer by tuning the length of the cavity. Figure 34.5 illustrates the dependence of the resonant wavelength and modal gain as functions of detuning the cavity length from its optimal length. The cavity length and DBR stopband were originally designed for 980 nm. Detuning the cavity length and leaving the parameters of the DBRs unchanged shift the emitted wavelength. However, modal gain remains close to its maximal value in the limited range of detuning. Greater detuning induces significant shift of the resonant wavelength, which is more weakly reflected by the DBRs, and modal gain decreases.

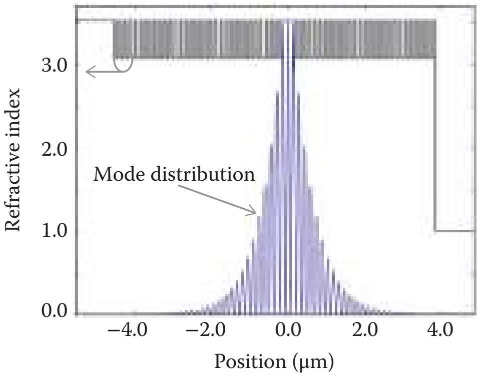


FIGURE 34.4 Distribution of the longitudinal mode and distribution of the refractive index along the optical axis of the VCSEL (Table 34.1).

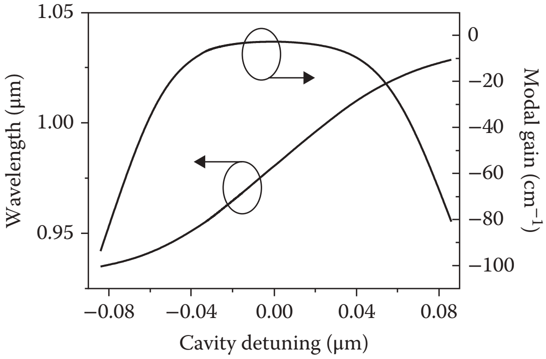


FIGURE 34.5 Modification of the resonant wavelength and modal gain induced by modification of the cavity length.

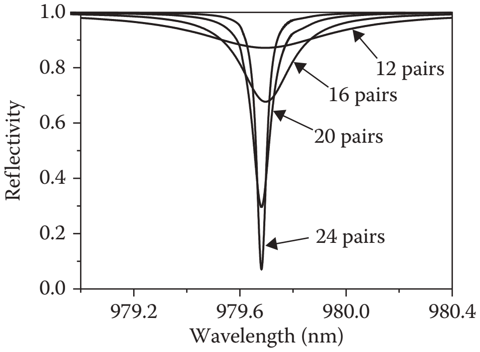


FIGURE 34.6 Power reflectance of the VCSEL wafer as a function of the incident wavelength for different numbers of DBR pairs.

Using the same method, but changing the boundary conditions to allow wave propagation toward the structure, one can calculate the reflectance or transmittance of the VCSEL wafer. In the case of a structure with an unpumped active region (in our calculations, absorption of 3000 cm^{-1} in the active region was assumed), there is a characteristic dip in the reflectance spectrum (Figure 34.6). If the dip is narrow and tends toward 0 reflectance, this indicates that the structure provides strong resonance for incident light that is absorbed by the active region. Such calculations can be compared directly with experimental measurements taken from the reflectance spectrum of the VCSEL wafer.

34.7 Two-Dimensional Scalar Approach

A more comprehensive approach, allowing lateral mode determination but still requiring moderate usage of computational memory and time, assumes that the electric field can be expressed independently using three coordinates: r, φ, and z:

$$E_u = E_u^{\text{rad}}(r)E_u^{\text{azim}}(\varphi)E_u^{ax}(z). \tag{34.28}$$

If the geometry of the device is cylindrical, one can assume the periodic dependence of

$$E_u^{\text{azim}}(\varphi) \sim \sin(m\varphi) \text{ or } E_u^{\text{azim}}(\varphi) \sim \cos(m\varphi), \tag{34.29}$$

which in practice reduces the problem to two dimensions (r, z).

Another assumption that makes use of the one-dimensional solution of (Equation 34.18) axial dependence can be expressed as

$$E_u^{ax} = A\exp\left(ik_{ax}z\right) + B\exp\left(-ik_{ax}z\right), \tag{34.30}$$

where k_{ax} is calculated for each cylinder (Figure 34.7) separately if the lateral distributions of the refractive indices in each cylinder are uniform in the radial direction. Equation 34.17 can then be transformed into

$$\left(\partial_r^2 + \frac{1}{r}\partial_r + k_{ax}^2 - k_0^2n^2 - \frac{m^2}{r^2}\right)E_u = 0, \tag{34.31}$$

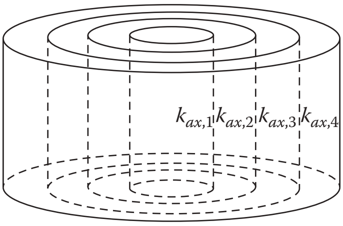


FIGURE 34.7 Schematic arrangement of cylinders. Wavenumbers $k_{ax,i}$ were calculated using one-dimensional TMM.

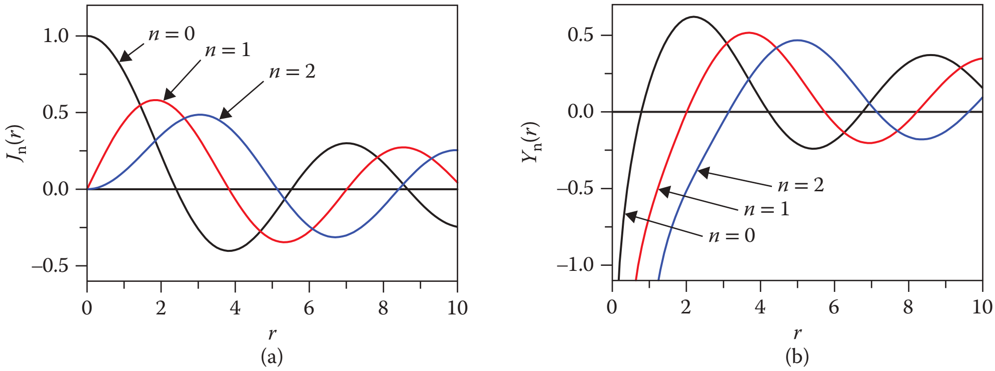


FIGURE 34.8 Bessel functions of the first (a) and second (b) kinds for $n = 0, 1, 2$.

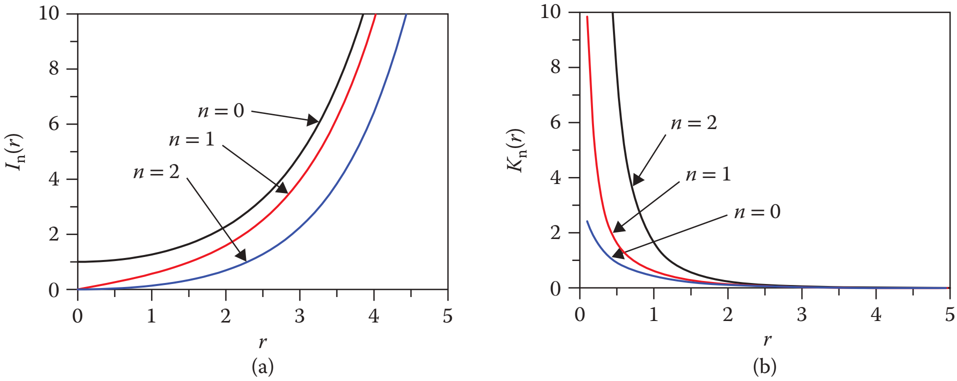


FIGURE 34.9 Modified Bessel function of the first (a) and second (b) kinds for $n = 0, 1, 2$.

which is solved separately for different m. The solutions of these Bessel differential equations are combinations of Bessel functions of the first $J_m(ar)$ and second kind $Y_m(ar)$ (Figure 34.8), where

$$a = k_{ax}^2 - k_0^2 n^2. \tag{34.32}$$

When $a < 0$, it is convenient to use modified Bessel functions of the first $I_m(-ar)$ and second kind $K_m(-ar)$ (Figure 34.9).

The following sums give the solution for an arbitrary layer:

$$E_{u,j}^{\mathrm{rad}} = A_j J_m\left(a_j r\right) + B_j Y_m\left(a_j r\right) \tag{34.33}$$

or

$$E_{u,j}^{\mathrm{rad}} = A_j I_m\left(-a_j r\right) + B_j K_m\left(-a_j r\right). \tag{34.34}$$

Implementing conditions for continuity of E_u^{rad} and continuity of their derivatives at the interfaces between cylinders, one can construct a transfer matrix between the electric fields in separate cylinders. In a similar manner as in Equation 34.24, one can construct a relation combining the fields from the first and last cylinders. The boundary conditions correspond to a finite value for the field at the center of the structure and to a decaying field for $r \to \infty$. The first condition eliminates $Y_m(ar)$ from the solution in the first cylinder and $I_m(-ar)$ in the last, which implies a condition similar to Equation 34.26. The set of k_0's that fulfill this condition correspond to lateral modes. Algorithms based on these approaches are very efficient and can be performed on personal computers.

The intensities of the lateral modes and the corresponding emission wavelengths of the VCSEL (Table 34.2), calculated using the algorithm detailed in this chapter, are presented in Figure 34.10. The set of modes illustrated were calculated assuming $E_u^{\mathrm{azim}}(\varphi) \sim \cos(m\varphi)$. A corresponding set can be calculated for $E_u^{\mathrm{azim}}(\varphi) \sim \sin(m\varphi)$, for which the emitted wavelengths will be exactly the same as for $\sim \cos(m\varphi)$. Superposition of these solutions and their complex conjugation leads to the elimination of angular dependence and only the oval shapes of the modes can be observed. However, in real-life structures in which perfect cylindrical symmetry can be broken by the geometry of the optical aperture or by the crystallographic structure of the layers, separate modes corresponding to both solutions can be observed.

TABLE 34.2 Resonant Wavelengths of Successive Lateral Modes Calculated Using Two-Dimensional Scalar Model

Mode Number	Wavelength (nm)
LP_{01}	979.3025
LP_{11}	978.8136
LP_{21}	978.175
LP_{02}	977.9555
LP_{31}	977.3976
LP_{12}	976.9519
LP_{41}	976.4901
LP_{22}	975.8181
LP_{03}	975.6216
LP_{51}	975.46
LP_{32}	974.5793
LP_{61}	974.3159
LP_{13}	974.2325
LP_{42}	973.3101
LP_{71}	973.0707
LP_{81}	971.7489

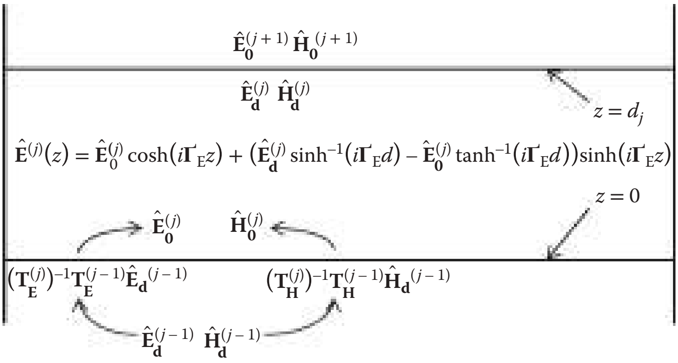

$\hat{\mathbf{E}}_{\mathbf{0}}^{(j+1)}\ \hat{\mathbf{H}}_{\mathbf{0}}^{(j+1)}$
$\hat{\mathbf{E}}_{\mathbf{d}}^{(j)}\ \hat{\mathbf{H}}_{\mathbf{d}}^{(j)}$
$z = d_j$
$\hat{\mathbf{E}}^{(j)}(z) = \hat{\mathbf{E}}_0^{(j)}\cosh(i\Gamma_E z) + (\hat{\mathbf{E}}_{\mathbf{d}}^{(j)}\sinh^{-1}(i\Gamma_E d) - \hat{\mathbf{E}}_{\mathbf{0}}^{(j)}\tanh^{-1}(i\Gamma_E d))\sinh(i\Gamma_E z)$
$z = 0$
$\hat{\mathbf{E}}_{\mathbf{0}}^{(j)}$
$\hat{\mathbf{H}}_{\mathbf{0}}^{(j)}$
$(\mathbf{T}_{\mathbf{E}}^{(j)})^{-1}\mathbf{T}_{\mathbf{E}}^{(j-1)}\hat{\mathbf{E}}_{\mathbf{d}}^{(j-1)}$
$(\mathbf{T}_{\mathbf{H}}^{(j)})^{-1}\mathbf{T}_{\mathbf{H}}^{(j-1)}\hat{\mathbf{H}}_{\mathbf{d}}^{(j-1)}$
$\hat{\mathbf{E}}_{\mathbf{d}}^{(j-1)}\ \hat{\mathbf{H}}_{\mathbf{d}}^{(j-1)}$

FIGURE 34.10 Definition of symbols for a single, uniform layer. $\hat{\mathbf{E}}_0$ and $\hat{\mathbf{H}}_0$ are virtual fields for the bottom interface of the layer, $\hat{\mathbf{E}}_d$ and $\hat{\mathbf{H}}_d$ are virtual fields for the upper interface, and d_j— is thickness of the layer.

34.8 Three-Dimensional Vectorial Approach

The three-dimensional vectorial approach relies on relations (Equations 34.5 and 34.6). To consider these in general geometries, we use the Cartesian coordinate system and assume that the medium in which field propagates is nonisotropic. The curl operator ∇× in the Cartesian coordinate system is defined as follows:

$$\nabla\times \equiv \begin{bmatrix} 0 & -\partial_z & \partial_y \\ \partial_z & 0 & -\partial_x \\ -\partial_y & \partial_x & 0 \end{bmatrix}. \tag{34.35}$$

We can eliminate E_z and H_z:

$$\begin{aligned} H_z &= \begin{bmatrix} -\frac{i}{\omega\mu_z\mu_0}\partial_y & \frac{i}{\omega\mu_z\mu_0}\partial_x \end{bmatrix}\begin{bmatrix} E_x \\ E_y \end{bmatrix} \\ E_z &= \begin{bmatrix} -\frac{i}{\omega\varepsilon_z\varepsilon_0}\partial_x & \frac{i}{\omega\varepsilon_z\varepsilon_0}\partial_y \end{bmatrix}\begin{bmatrix} H_y \\ H_x \end{bmatrix}, \end{aligned} \tag{34.36}$$

which leads to the relations:

$$\begin{aligned} \partial_z\bar{\mathbf{E}} &= \begin{bmatrix} -\partial_x\frac{i}{\omega\varepsilon_z\varepsilon_0}\partial_x - \omega\mu_y\mu_0 & \partial_x\frac{i}{\omega\varepsilon_z\varepsilon_0}\partial_y \\ -\partial_y\frac{i}{\omega\varepsilon_z\varepsilon_0}\partial_x & \partial_y\frac{i}{\omega\varepsilon_z\varepsilon_0}\partial_y - \omega\mu_x\mu_0 \end{bmatrix}\bar{\mathbf{H}} = \mathbf{R_E}\bar{\mathbf{H}} \\ \partial_z\bar{\mathbf{H}} &= \begin{bmatrix} -\partial_y\frac{i}{\omega\mu_z\mu_0}\partial_y - \omega\varepsilon_x\varepsilon_0 & \partial_y\frac{i}{\omega\mu_z\mu_0}\partial_x \\ -\partial_x\frac{i}{\omega\mu_z\mu_0}\partial_y & \partial_x\frac{i}{\omega\mu_z\mu_0}\partial_x - \omega\varepsilon_y\varepsilon_0 \end{bmatrix}\bar{\mathbf{E}} = \mathbf{R_H}\bar{\mathbf{E}} \end{aligned}, \tag{34.37}$$

where

$$\bar{\mathbf{H}} = \begin{bmatrix} H_x \\ H_y \end{bmatrix} \text{ and } \bar{\mathbf{E}} = \begin{bmatrix} E_x \\ E_y \end{bmatrix}. \tag{34.38}$$

Equations 34.37 can be transformed into

$$\begin{aligned} \partial_z^2\bar{\mathbf{E}} &= \mathbf{R_H}\mathbf{R_E}\bar{\mathbf{E}} \\ \partial_z^2\bar{\mathbf{H}} &= \mathbf{R_E}\mathbf{R_H}\bar{\mathbf{H}} \end{aligned}. \tag{34.39}$$

Both equations can be solved in the base, which simplifies the matrices $\mathbf{R_H R_E}$ and $\mathbf{R_E R_H}$ into diagonal forms. The following analysis leads to the solution of Equation 34.39 and enables the characteristic value of the problem to be found. The solution corresponds to the wavenumber of the mode, as well as to the characteristic vectors, which determine the distribution of the electromagnetic field within the structure.

The equation for the electric field rewritten in the new virtual base (after diagonalization) is

$$\partial_z^2 \hat{\mathbf{E}} = \mathbf{R_H R_E} \hat{\mathbf{E}}, \tag{34.40}$$

where $\hat{\mathbf{E}}$ stand for electric field in the new, virtual base and can be defined as

$$\hat{\mathbf{E}} = \mathbf{T}_\mathbf{E}^{-1} \bar{\mathbf{E}}, \tag{34.41}$$

where the matrix $\mathbf{T_E}$ diagonalizes $\mathbf{R_H R_E}$:

$$\mathbf{T}_\mathbf{E}^{-1} \mathbf{R_H R_E T_E} = \boldsymbol{\Gamma}_\mathbf{E}^2. \tag{34.42}$$

Here $\boldsymbol{\Gamma}_\mathbf{E}^2$ is a diagonal matrix, which allows a solution of Equation 34.39 in the well-known form of a standing wave:

$$\hat{\mathbf{E}}(z) = \mathbf{A} \cosh\left(i\boldsymbol{\Gamma}_\mathbf{E} z\right) + \mathbf{B} \sinh\left(i\boldsymbol{\Gamma}_\mathbf{E} z\right). \tag{34.43}$$

For further derivation, we can use values of the electric field equal to $\hat{\mathbf{E}}_0$ and $\hat{\mathbf{E}}_d$ on the extreme borders of the uniform region along the z direction. The thickness of the considered uniform layer equals d (Figure 34.10).

We can therefore relate the solution to the values of the field at the opposite borders of the uniform region:

$$\hat{\mathbf{E}}(z) = \hat{\mathbf{E}}_\mathbf{0} \cosh\left(i\boldsymbol{\Gamma}_\mathbf{E} z\right) + \left(\hat{\mathbf{E}}_\mathbf{j} \sinh^{-1}\left(i\boldsymbol{\Gamma}_\mathbf{E} d\right) - \hat{\mathbf{E}}_\mathbf{0} \tanh^{-1}\left(i\boldsymbol{\Gamma}_\mathbf{E} d\right)\right) \sinh\left(i\boldsymbol{\Gamma}_\mathbf{E} z\right). \tag{34.44}$$

The solution for the magnetic field can be found in a similar manner.

Using the solution for $\hat{\mathbf{E}}$ and $\hat{\mathbf{H}}$ and substituting to Equation 34.37, one can find the relation connecting electric and magnetic fields within the layer:

$$\begin{bmatrix} \hat{\mathbf{H}}_0 \\ -\hat{\mathbf{H}}_d \end{bmatrix} = \begin{bmatrix} \mathbf{y}_1 & \mathbf{y}_2 \\ \mathbf{y}_2 & \mathbf{y}_1 \end{bmatrix} \begin{bmatrix} \hat{\mathbf{E}}_0 \\ \hat{\mathbf{E}}_d \end{bmatrix}, \tag{34.45}$$

where

$$\mathbf{y}_1 = \left(\mathbf{T}_\mathbf{E}^{-1} \mathbf{R_H T_H}\right)^{-1} \boldsymbol{\Gamma}_\mathbf{E} \tanh^{-1}\left(i\boldsymbol{\Gamma}_\mathbf{E} d\right), \tag{34.46}$$

$$\mathbf{y}_2 = -\left(\mathbf{T}_\mathbf{E}^{-1} \mathbf{R_H T_H}\right)^{-1} \boldsymbol{\Gamma}_\mathbf{E} \sinh^{-1}\left(i\boldsymbol{\Gamma}_\mathbf{E} d\right). \tag{34.47}$$

The relation between the fields from the ith and $(i-1)$th layers is given in the formulas:

$$\hat{\mathbf{E}}_0^{(i)} = \left(\mathbf{T}_\mathbf{E}^{(i)}\right)^{-1} \mathbf{T}_\mathbf{E}^{(i-1)} \hat{\mathbf{E}}_d^{(i-1)} = \mathbf{t}_\mathbf{E}^{(i)} \hat{\mathbf{E}}_d^{(i-1)}, \tag{34.48}$$

$$\hat{\mathbf{H}}_0^{(i)} = \left(\mathbf{T}_\mathbf{H}^{(i)}\right)^{-1} \mathbf{T}_\mathbf{H}^{(i-1)} \hat{\mathbf{H}}_d^{(i-1)} = \mathbf{t}_\mathbf{H}^{(i)} \hat{\mathbf{H}}_d^{(i-1)}. \tag{34.49}$$

Using Equations 34.45, 34.48, and 34.49, the relationship between the electric and magnetic fields, for the interface separating layers $i+1$ and i, takes an iterative formula:

$$\hat{\mathbf{H}}_d^{(i)} = \mathbf{Y}^{(i)}\hat{\mathbf{E}}_d^{(i)}, \tag{34.50}$$

$$\mathbf{Y}^{(i)} = -\left(\mathbf{y}_2^{(i)}\left(\mathbf{t}_{\mathbf{H}}^{(i)}\mathbf{Y}^{(i-1)}\left(\mathbf{t}_{\mathbf{E}}^{(i)}\right)^{-1} - \mathbf{y}_1^{(i)}\right)^{-1}\mathbf{y}_2^{(i)} + \mathbf{y}_1^{(i)}\right). \tag{34.51}$$

Propagating by this procedure from the bottom limit of the simulated structure to the upper limit, the characteristic equation can be determined. However, from the numerical point of view, it is more efficient and precise to perform the procedure starting from the bottom limit to the center, and then to start again from the upper limit down to the center. The center does not correspond to the geometrical center but to the plane at which the electromagnetic field reaches or it is close to its maximum, which is termed as the "matching interface." One can write the relation (Equation 34.50) for the top-down and bottom-up algorithms as:

$$\hat{\mathbf{H}}_d^{(m)} = \mathbf{Y}_{\mathrm{up}}^{(m)}\hat{\mathbf{E}}_d^{(m)}, \tag{34.52}$$

$$\hat{\mathbf{H}}_d^{(l)} = \mathbf{Y}_{\mathrm{down}}^{(l)}\hat{\mathbf{E}}_d^{(l)}. \tag{34.53}$$

m counts layers from the bottom and l from the top.

Since the magnetic field is continuous in the matching interface in the real base, one can write the relation

$$\bar{\mathbf{H}}_d^{(m)} = \bar{\mathbf{H}}_d^{(l)}. \tag{34.54}$$

Transforming back Equation 34.54 to the virtual base:

$$\mathbf{T}_{\mathbf{H}}^{(m)}\hat{\mathbf{H}}_d^{(m)} = \mathbf{T}_{\mathbf{H}}^{(m)}\hat{\mathbf{H}}_d^{(l)}, \tag{34.55}$$

using Equations 34.52 and 34.53:

$$\mathbf{T}_{\mathbf{H}}^{(m)}\mathbf{Y}_{\mathrm{up}}^{(m)}\hat{\mathbf{E}}_d^{(m)} = \mathbf{T}_{\mathbf{H}}^{(m)}\mathbf{Y}_{\mathrm{down}}^{(l)}\hat{\mathbf{E}}_d^{(l)}, \tag{34.56}$$

and finally using relation (Equation 34.41) to lead us to an equation that includes the fields in the real base:

$$\left(\mathbf{T}_{\mathbf{H}}^{(m)}\mathbf{Y}_{\mathrm{up}}^{(m)}\left(\mathbf{T}_{\mathbf{E}}^{(m)}\right)^{-1} - \mathbf{T}_{\mathbf{H}}^{(m)}\mathbf{Y}_{\mathrm{down}}^{(l)}\left(\mathbf{T}_{\mathbf{E}}^{(l)}\right)^{-1}\right)\bar{\mathbf{E}} = \mathbf{Y}\bar{\mathbf{E}} = 0. \tag{34.57}$$

Equation 34.57 is also the characteristic value equation. The solution of the equation determines the characteristic values, which correspond to the complex wavenumber. As before, the imaginary part of the characteristic value relates to the modal gain of the propagating wave. Equation 34.57 has a nontrivial solution only if matrix **Y** is singular, which imposes the condition that one of the eigenvalues of **Y** is 0. This condition is useful in numerical searches for solutions to Equation 34.57.

The intensity distributions of lateral modes calculated using the vectorial approach are very close to those depicted in Figure 34.11. The main difference between the solutions lies in the different number of modes (Table 34.3). In the vectorial solution, there are no modes of shorter wavelengths than LP_{22}. This may be associated with diffraction losses, which are more severe for higher order modes but are not included in the scalar model.

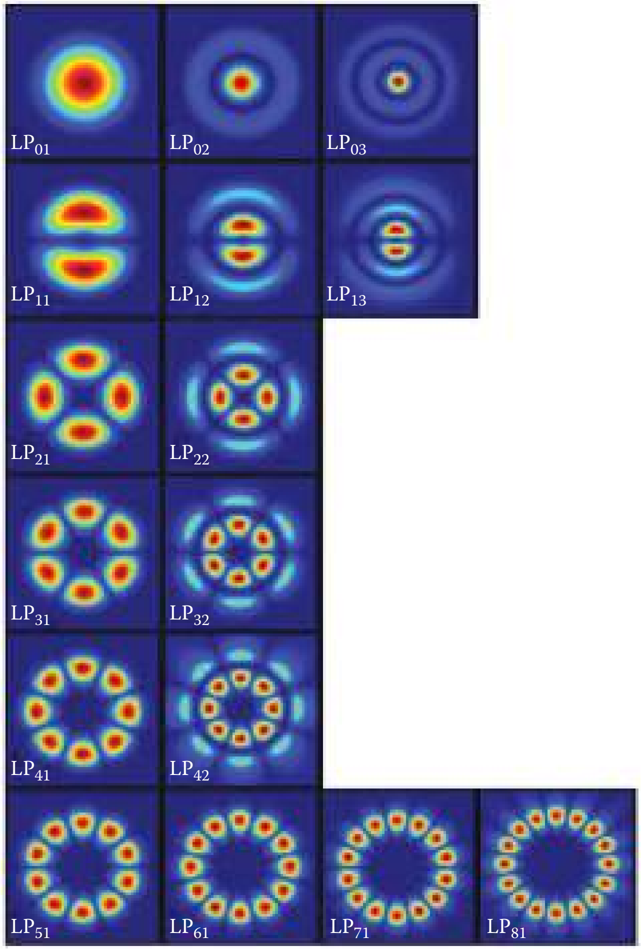


FIGURE 34.11 Distributions of lateral modes in the VCSEL structure detailed in Table 34.1, where $R = 4$ μm, calculated using the scalar two-dimensional approach.

TABLE 34.3 Resonant Wavelengths of Successive Lateral Modes Calculated Using the Three-Dimensional Vectorial Model

Mode Number	Wavelength (nm)
LP_{01}	979.3014
LP_{11}	978.8108
LP_{21}	978.1691
LP_{02}	977.9616
LP_{31}	977.3912
LP_{12}	976.9431
LP_{41}	976.4811
LP_{22}	975.8031

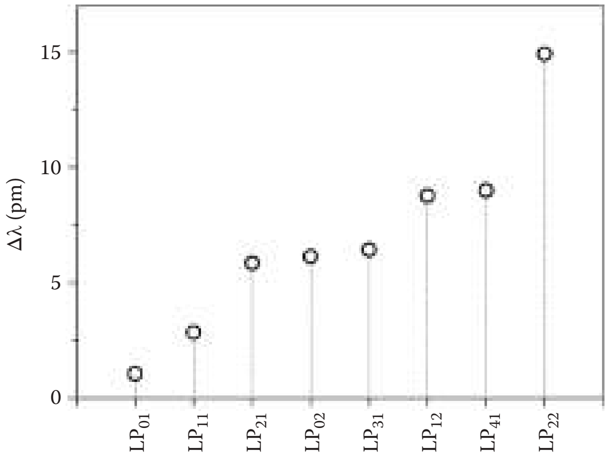


FIGURE 34.12 Difference in resonant wavelengths of successive lateral modes calculated using scalar and vectorial models.

Figure 34.12 illustrates the difference (Δλ) with respect to the wavelength of modes calculated using scalar and vectorial models. The vectorial model provides solutions with shorter wavelengths. However, the difference is on a scale of picometers. The monotonic increase in Δλ shows that the plane wave assumption used in the scalar model is more accurate in the case of lower order modes.

34.9 Impact of Oxide Aperture Diameter and Position

Finally, we shall analyze the influence of oxide aperture on the emitted wavelength of the VCSEL considered in this study. Although the structure is a very simplified design, the general tendencies presented in Figures 34.13 and 34.14 are applicable to real-life oxide-confined VCSEL structures.

Reducing the aperture (Figure 34.13) size leads to a characteristic blue shift, since a narrower aperture introduces stronger interaction between the mode and the oxide layer, which has a low refractive index. The wider the aperture diameter, the closer the values for the wavelengths determined using each approach.

The location of the oxide layer has a significant effect on the VCSEL parameters (Figure 34.14). If it is positioned at the antinode of the standing wave, the interaction between the optical field and carriers within the active region is intensified, since the modes are strongly confined within the aperture. This configuration reduces the threshold current and broadens the spectrum of emission, as numerous lateral modes are excited. If the oxide layer is in the node position, the number of modes is reduced, which narrows the

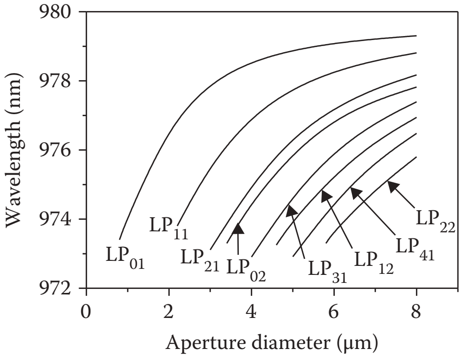


FIGURE 34.13 Dependence of the lateral mode wavelengths on aperture diameter.

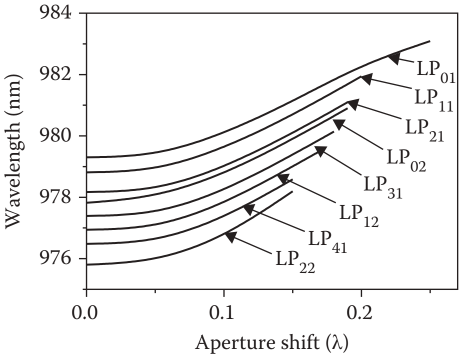


FIGURE 34.14 Dependence of lateral mode wavelengths on aperture shift from the antinode position relative to the wavelength of the standing wave.

emitted spectrum. However, interaction between the mode and the oxide layer is weaker and the threshold current increases. Shifting the oxide layer from the antinode position leads to an increase in the wavelength. This can be explained by weaker interaction between the field and the oxidized layer, which leads to an increase in the effective index of the mode and eventually causes redshift.

34.10 Conclusion

Scalar one- and two-dimensional optical approaches represent very efficient tools for VCSEL designers. Their calculation algorithms, requiring short calculation times and minor memory storage, enable one to perform rapid calculations using even PCs. However, the assumptions that make them so efficient introduce simplifications, which may produce inaccurate results in more sophisticated structures. Most significantly, the basic scalar assumption of plane optical wave propagation is violated in the case of aperture sizes comparable to the radiation wavelength and for higher order modes.

Vectorial methods are much more complicated, since they are based on the solution of six coupled equations instead of only one (in the simplified scalar approach). They give more exact results at the price of significantly longer computational time and a considerably more complicated algorithm. Vectorial methods may not be the optimal approach to modeling a standard VCSEL. However, analysis of VCSELs with in-built refractive index steps in lateral directions or subwavelength patterning certainly requires the use of vectorial methods.

References

1. Melngailis, I. Longitudinal injection plasma laser of InSb. *Appl Phys Lett.* 6, 59–60 (1965).
2. Soda, H. et al. GaInAsP/InP surface emitting injection lasers. *Jap J Appl Phys.* 18, 2329–2330 (1979).
3. Ogura, M. et al. GaAs/$Al_xGa_{1-x}As$ multilayer reflector for surface emitting laser diode. *Jap J Appl Phys.* 22, L112–L114 (1983).
4. MacDougal, M. H. et al. Ultralow threshold current vertical-cavity surface-emitting lasers with AlAs oxide-GaAs distributed Bragg reflectors. *IEEE Photon Technol Lett.* 7, 229–231 (1995).
5. Tai, K. et al. Use of implant isolation for fabrication of vertical cavity surface emitting lasers. *Electron Lett.* 25, 1644–1645 (1989).
6. Gadallah, A.-S. et al. High-output-power single-higher-order transverse mode VCSEL with shallow surface relief. *IEEE Photon Technol Lett.* 23, 1040–1042 (2011).

7. Zhou, D. et al. High-power single-mode antiresonant reflecting optical waveguide-type vertical-cavity surface-emitting lasers. *IEEE J Quantum Electron*. 38, 1599–1606 (2002).
8. Kapon, E. et al. Long-wavelength VCSELs: Power-efficient answer. *Nat Photon*. 3, 27–29 (2009).
9. Danner, A. J. et al. Progress in photonic crystal vertical cavity lasers. *IEICE Trans Electron*. E88, 944–950 (2005).
10. Chang-Hasnain, C. J. et al. High-contrast gratings for integrated optoelectronics. *Adv Opt Photon*. 4, 379–440 (2012).
11. Saleh, B. E. A., Teich, M. C. *Fundamentals of Photonics*. New York, NY: John Wiley & Sons (1991).
12. Bienstman, P. R. et al. Comparison of optical VCSEL models on the simulation of oxide-confined devices. *IEEE J Quant Electron*. 37, 1618–1631 (2001).
13. Kuszelewicz, R. *COST 268 Modeling Exercise*. Paris, France: CNET (1998).
14. Bergmann, M. J. et al. Optical-field calculations for lossy multiple-layer $Al_xGa_{1-x}N/In_xGa_{1-x}N$ laser diodes. *J Appl Phys*. 84, 1196–1203 (1998).

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

VII

Photodetectors and Modulators

35 Photodetector Fundamentals *Prasanta Basu* 283
Introduction • Basic Working Principle • Important Characteristics • Classification of PDs • Different PD Configurations • Improved Devices • Materials

36 P-N Junction Photodiodes *Weida Hu* 307
Introduction • Basic Equations and Models • Optical Generation Models • Characteristic Input Parameters • Simulation Examples of P-N Junction Photodiodes • Summary

37 Quantum Well Infrared Photodetectors *Kwong-Kit Choi* 337
Introduction • Theoretical Model of QWIPs • Material Absorption Modeling • Detector QE Modeling • EM Design of Resonator-QWIPs • Conclusion

38 Optical Modulators *Dominic F.G. Gallagher and Dmitry Labukhin* 363
Introduction • QCSE-Based Modulators • Electro-Optic Effect Modulators • Plasma-Effect Modulators • Si Electroabsorption versus Electrorefractive Effect • Resonant Modulators

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

35

Photodetector Fundamentals

Prasanta Basu

35.1 Introduction 283
35.2 Basic Working Principle 284
Conversion of Light into Current • Absorption in Semiconductors
35.3 Important Characteristics 287
Wavelength Selectivity • Quantum Efficiency • Responsivity • Response Time • Noise Equivalent Power and Detectivity
35.4 Classification of PDs 290
35.5 Different PD Configurations 290
Photoconductive Detectors • p-n Photodiode • p-i-n Photodiode • Avalanche Photodiodes • Schottky Barrier PD • Metal-Semiconductor-Metal PD • Phototransistors • QWIP and QDIP • Single-Photon Detectors
35.6 Improved Devices 301
Waveguide PDs • RCE Structure • Incorporation of Heterojunctions and Quantum Nanostructures
35.7 Materials 303

35.1 Introduction

Photodetectors (PDs) convert light energy into electrical energy, which usually manifests as a photocurrent. They are a kind of transducer. The electrical signal is useful in electronic imaging, providing a means for viewing images much faster than the earlier photographic imaging. PDs also find important applications in present-day fiber-optic communication systems. In this case, PDs receive the stream of weak optical pulses at the far end of the optical fiber originally transmitted by the transmitter. The role of the detector is to convert these optical pulses into electrical pulses, which are then processed for use in a telephone, a computer, or other terminals at the receiving end (Bhattacharyya 1996; Deen and Basu 2012; Kumar and Deen 2014).

Other important applications of PDs include automatic door opener, use in remote control elements in TV, VCR, air conditioners, etc., in CCDs in a video camera, and in astronomy, research, and defense.

This chapter gives an introduction to the working principle of photodetection by semiconductors, the material properties, different types, and their structures, important parameters pertaining to PDs, and materials and makes a list of the application areas.

35.2 Basic Working Principle

The conversion of optical energy into electrical energy by a PD involves three basic processes. First, a bunch of incident photons is absorbed by the semiconducting material to create excess electron–hole pairs (EHPs). The carriers are then transported across the absorption or transit region. In the third stage, the transported carriers are collected to generate photocurrent. In this section, we consider the basic principle, structure, and the process of absorption (Bhattacharyya 1996; Deen and Basu 2012).

35.2.1 Conversion of Light into Current

The process of creation of excess EHPs and their collection is illustrated in Figure 35.1. In Figure 35.1a, a photon of energy $\hbar\omega \geq E_g$, incident on the semiconductor, is absorbed by an electron in the valence band (VB), which then moves up in the conduction band (CB), leaving a hole in the VB. An excess EHP, in excess of the thermal equilibrium values, is therefore created. To collect these carriers, a reverse-biased p-n junction is employed. As shown in Figure 35.1b, the electric field in the depletion layer drives the excess electrons and excess holes in the opposite directions to be collected by the terminals and thus to produce current. The reverse saturation current in the absence of illumination is called the dark current. The photogenerated current is proportional to the flux of incident photons or the power of incident light (Kaiser 2013; Senior and Jamro 2009).

As shown in Figure 35.1b, absorption may also take place outside the depletion layer. The excess EHPs generated therein move by the slower process of diffusion and are collected by the terminals. The diffusion process limits the speed of operation and distorts the electrical pulses.

35.2.2 Absorption in Semiconductors

The process of absorption is now discussed in some detail in Figure 35.2. In Figure 35.2a, a photon of energy $\hbar\omega \geq E_g$ is shown incident on the semiconductor. The photon lifts an electron from the VB to a higher lying state in the CB. The energetic electron then relaxes to the bottom of the CB via phonon emission. A photon thus creates an excess EHP. The pair ultimately recombines to give rise to emission (luminescence). The same processes are illustrated in the E-k diagrams of a direct gap semiconductor in Figure 35.2b and of an indirect gap semiconductor in Figure 35.2c. The photon momentum $\hbar\mathbf{k}_{\text{photon}}$ is usually negligible and the momentum ($\mathbf{k}$) conservation can be written as $\mathbf{k}_e = \mathbf{k}_h + \mathbf{k}_{\text{photon}} \approx \mathbf{k}_h$, where $\mathbf{k}_e(\mathbf{k}_h)$ denotes the wave vector for electron (hole). Accordingly, the absorption in a direct gap semiconductor is indicated by a vertical

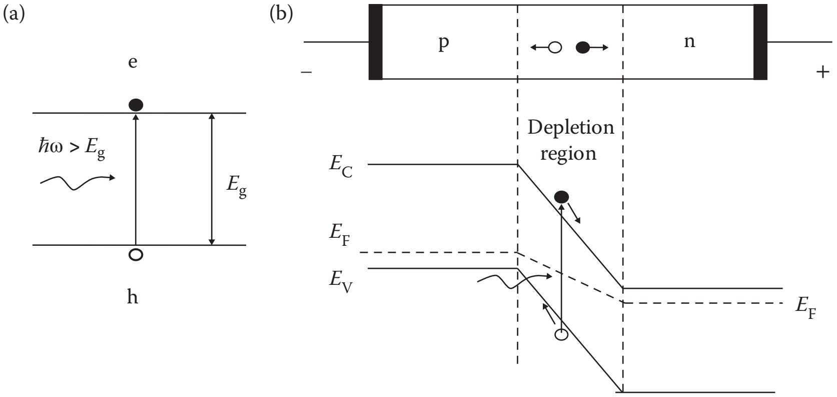


FIGURE 35.1 Conversion of optical energy into electrical energy: (a) absorption of photons and generation of excess EHPs; (b) collection of carriers by using a reverse-biased p-n junction.

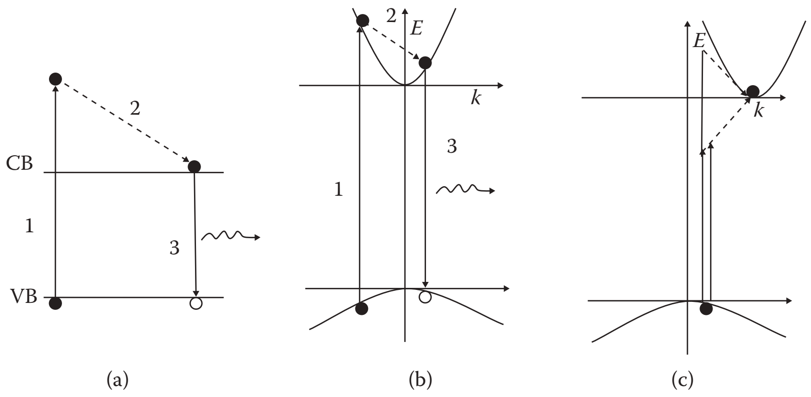


FIGURE 35.2 (a) Absorption (1), relaxation (2), and recombination (3) processes in semiconductors. CB (VB) denotes conduction (valence) band edges; (b) absorption (1), relaxation (2), and recombination (3) processes shown in the E-k diagram of a direct gap semiconductor; (c) the same processes in the E-k diagram of an indirect gap semiconductor. Dashed lines correspond to phonon scattering processes.

line as shown in Figure 35.2b. However, for indirect gap semiconductor, an extra momentum is needed to place the electron in the CB, which is provided by, among other agencies, phonons. In Figure 35.2c, the indirect absorption is indicated by both phonon absorption and emission. The indirect absorption has a very low probability and orders of magnitude being lower than in direct gap semiconductors (Basu 1997).

The variation of light intensity inside a semiconductor, $I(z)$, is governed by the following expression known as the Lambert's law:

$$I(z) = I_0 \exp(-\alpha z), \tag{35.1}$$

where I_0 is the intensity at the surface ($z = 0$), and α is the absorption coefficient of the material, which is a function of photon energy or wavelength.

The absorption coefficient for direct gap semiconductors is much higher than for indirect gap semiconductors.

35.2.2.1 Absorption in Direct Gap Semiconductors

As mentioned already, the absorption process in a direct gap semiconductor is indicated by a vertical line in the E-k diagram. The absorption coefficient in direct gap semiconductors is calculated by using the first-order time-dependent perturbation theory and the Fermi golden rule. The expression for absorption coefficient is of the form (Basu 1997)

$$\alpha(\hbar\omega) = A\left(\hbar\omega - E_g\right)^{1/2}\left[f_c - f_v\right), \hbar\omega \geq E_g, \tag{35.2}$$

where the prefactor A contains the fundamental constants, material parameters such as permittivity, etc., and $f_{c(v)}$ denotes the Fermi occupation probabilities of electrons in the CB (VB).

35.2.2.2 Absorption in Indirect Gap Semiconductors

Since a photon and a phonon are involved in the absorption process, the second-order perturbation theory is used to calculate the absorption coefficient in indirect gap semiconductors. The expression takes the

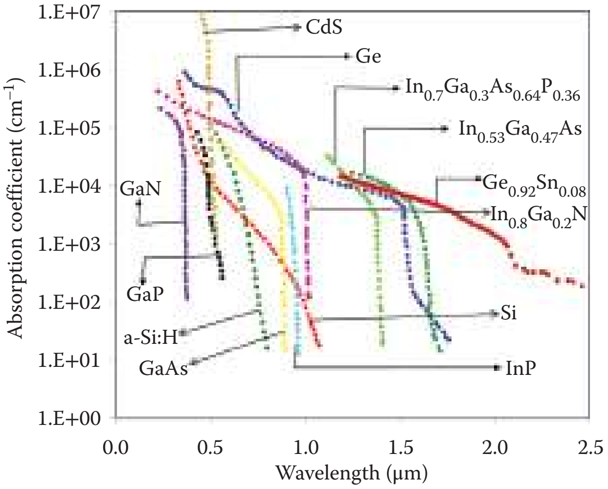


FIGURE 35.3 Absorption spectra of some common semiconductors.

following form at 0 K when a single type of phonon of energy $\hbar\omega_p$ is involved (Basu 1997):

$$\alpha(\hbar\omega) = A_a\left(\hbar\omega - E_g + \hbar\omega_p\right)^2 + A_e\left(\hbar\omega - E_g - \hbar\omega_p\right)^2, \tag{35.3}$$

where the first and second terms correspond, respectively, to the absorption followed by phonon absorption and that followed by phonon emission, and the prefactors A's are distinguished by the corresponding suffixes. There is, in general, a possibility of involvement of different types of phonons. Figure 35.3 shows typical absorption coefficients of some representative semiconductors.

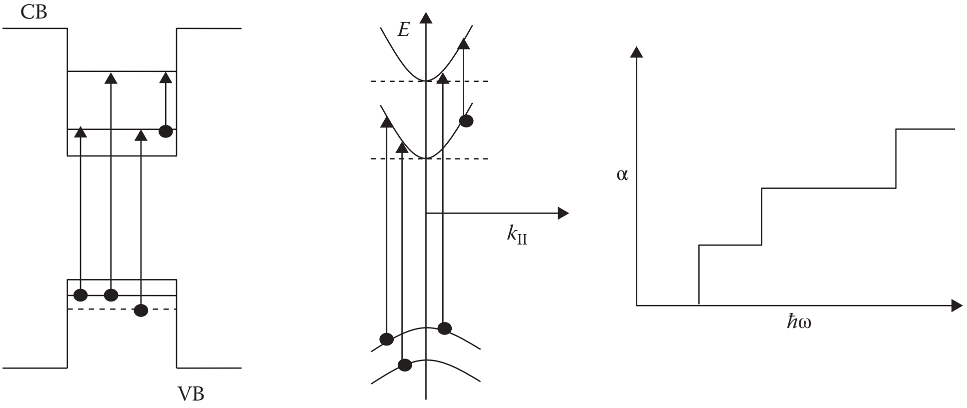
35.2.2.3 Absorption in Quantum Nanostructures

Recent detectors make increasing use of quantum nanostructures such as quantum wells (QWs), quantum wires (QWRs), and quantum dots (QDs), rather than the bulk semiconductors. In QWs, for example, the thickness of the layer of interest is comparable to the de Broglie wavelength of the particle. A QW is realized by sandwiching a thin layer of GaAs between two layers of a higher gap material such as $Al_xGa_{1-x}As$. The typical band diagram of the QW is shown in Figure 35.4a. Due to very small thickness d (<de Broglie wavelength) of GaAs, the motion of electrons and holes is quantized, and assuming that the barrier heights are infinite, the quantized energy levels, or subbands as they are called, are expressed as follows:

$$E_n = \frac{\hbar^2}{2m_{\text{eff}}}\left(\frac{n\pi}{d}\right)^2, \quad n = 1, 2, 3, \ldots. \tag{35.4}$$

The effective mass, m_{eff}, is to be replaced by the electron effective mass, m_e, heavy hole effective mass, m_{hh}, and light hole effective mass, m_{lh}, in order to calculate the respective subband energies (Basu 1997).

The absorption processes in a QW are shown in Figure 35.4b. Here, transitions can occur in several ways. An electron from a valence subband (both lh and hh) can absorb a photon and can come to one of the subbands in the CB. This is a typical band-to-band absorption; however, the absorption threshold, that is, the photon energy marking the sharp rise in the absorption coefficient, increases from the threshold in bulk material that equals the bandgap (Equation 35.2). In addition, an electron (or a hole) in a lower subband (i) can absorb a photon and may move up to a higher subband (j) in the same band. This is a representative intersubband transition. The absorption processes in QWs are governed by some selection rules.



(a) (b) (c)

FIGURE 35.4 (a) Band diagram and quantized energy levels or subbands in a QW structure; (b) optical absorption processes in a QW; both band-to-band and intersubband processes are shown; (c) idealized absorption spectra for QWs. α and $\hbar\omega$ denote the absorption coefficient and the photon energy, respectively. k_{II} is the wave vector along the QW layer plane. CB (VB) denotes conduction (valence) band edges.

The idealized band-to-band absorption coefficient in the QW of a direct gap semiconductor takes the following form (Basu 1997):

$$\alpha(\hbar\omega) = K\sum_{n} H\left[\hbar\omega - E_{\mathrm{nn}}\right), \tag{35.5}$$

where the prefactor K contains fundamental constants, effective (reduced) masses of electrons and holes, and the permittivity of the QW material. In Equation 35.5, H is a Heaviside step function, and E_{nn} is the energy difference between nth hole and nth electron subband.

Typical absorption spectra for QWs obtained from Equation 35.5 are shown in Figure 35.4c, which indicates that the spectra are staircase like.

The idea can be extended to the cases of QWRs and QDs, in which the motion of the particles is inhibited along two and three directions, respectively.

The absorption spectra get modified when excitonic effects are included. Excitons are complexes formed by the EHP bound by the mutual Coulomb interaction.

35.3 Important Characteristics

In this section, some of the important characteristics of PDs are discussed.

35.3.1 Wavelength Selectivity

Different materials are used for the detection of photons at different wavelengths. The wavelength range covered by a particular material is determined by the cutoff wavelength, above which the absorption by it is negligibly small. The cutoff wavelength is determined by the expression $\hbar\omega = E_{\mathrm{g}}$, where E_{g} is the bandgap, from which it is simple to write

$$\lambda_c = \frac{hc}{E_{\mathrm{g}}} = \frac{1.24}{E_{\mathrm{g}}(\mathrm{eV})}\,\mu\mathrm{m}. \tag{35.6}$$

For a bandgap of 1 eV, the cutoff wavelength is 1.24 μm.

35.3.2 Quantum Efficiency

Quantum efficiency (QE), sometimes called the internal quantum efficiency (IQE), is the ratio of the generated EHPs and the number of incident photons. The higher the QE is, larger is the sensitivity of the PD and it is therefore an important figure of merit.

More practical measure of the sensitivity is the external QE (EQE), which is the ratio of EHPs collected at the terminals to the number of incident photons. EQE is less than IQE since a part of the incident photon flux is reflected at the surface and some EHPs are lost by recombination at the surface or by nonradiative processes. EQE is a function of wavelength.

Let I_{pc} be the photocurrent, P_i be the incident power, and $\hbar\omega$ be the photon energy. By definition,

$$\eta = \frac{\text{electron generation rate}}{\text{photon incidence rate}} = \frac{I_{\mathrm{pc}}/e}{P_i/\hbar\omega} = \frac{I_{\mathrm{pc}}}{P_i}.\frac{\hbar\omega}{e} = \frac{1.24}{\lambda\,(\mu\mathrm{m})}\frac{I_{\mathrm{pc}}}{P_i}. \tag{35.7}$$

Equation 35.7 may also be written as a product of three factors as follows:

$$\eta = \zeta(1 - e^{-\alpha d})(1 - R). \tag{35.8}$$

In Equation 35.8, ζ is the fraction of the photogenerated carriers that reach the outer circuit without recombination, the second term within parentheses denotes the fraction of light absorbed in the absorption region of width d, and the last term represents the amount of light transmitted into the semiconductor, R being the reflection coefficient.

35.3.3 Responsivity

Responsivity is defined as the ratio of the primary photocurrent without gain, I_{pc}, and the incident optical power and is expressed as

$$\mathfrak{R} = \frac{I_{\mathrm{pc}}}{P_l} = \eta(\lambda)\left[\frac{e}{(hc/\lambda)}\right]. \tag{35.9}$$

In PDs, responsivity, or sensitivity as it is alternately called, may be increased by increasing the absorption region thickness and minimizing reflection from its surface. The latter can be accomplished by coating the surface with an antireflection (AR) coating of refractive index $n_{\mathrm{AR}} = \sqrt{n_{\mathrm{air}} n_{\mathrm{SC}}}$, where the subscripts refer to AR material, air, and semiconductor, respectively.

35.3.4 Response Time

The speed of response of PDs is limited by three main factors, as discussed next.

35.3.4.1 Diffusion Time of Carriers

The time taken by carriers to diffuse a distance d is given by (Streetman and Banerjee 2016)

$$\tau_{\mathrm{diff}} = \frac{d^2}{2D}, \tag{35.10}$$

where D is the diffusion constant. Diffusion is a slower process than drift. The carriers that are optically generated within a diffusion length from the depletion layer of a photodiode move to the depletion layer by this slow process. The photogenerated current pulse due to a narrow optical pulse shows “tails” due to the slow diffusion process.

35.3.4.2 Drift Time through Depletion Region

The drift velocity of both electrons and holes in a semiconductor attains a saturation value in the presence of very high fields (Streetman and Banerjee, 2016). Usually, a high electric field exists in the depletion layer of a reverse-biased junction. The longest transit time of carriers in the depletion layer of width W may therefore be written as follows:

$$\tau_{\text{drift}} = \frac{W}{v_{\text{sat}}}, \tag{35.11}$$

where v_{sat} is the saturation velocity of the carrier, which is different for electrons and holes. The maximum speed of operation in a PD is limited by transit time.

35.3.4.3 Junction Capacitance Effect

The electrical equivalent circuit of a PD using a p-i-n diode as an example is shown in Figure 35.5. It includes a photogenerated current source (I_{ph}), the depletion layer capacitance (C_{J}), a series resistance (R_{S}) due to bulk resistance and ohmic contacts, a parasitic capacitance (C_{P}), and a load resistance (R_{L}). The electrical 3-dB bandwidth may be expressed as

$$f_{\text{3dB}} = \frac{1}{2\pi(C_{\text{J}} + C_{\text{P}})(R_{\text{L}} + R_{\text{S}})}. \tag{35.12}$$

35.3.5 Noise Equivalent Power and Detectivity

The term *noise* refers to unwanted electrical signal that masks the signal to be detected. Several noise sources, such as thermal, shot, dark current, and flicker noise, are present in PDs and associated circuits and make it difficult to detect weak signals. The minimum detectable signal corresponds to an rms output signal equal to that generated by noise. The signal-to-noise ratio (SNR) is then unity. The noise equivalent power (NEP), a measure of the minimum detectable signal, is defined as the power of sinusoidally modulated monochromatic optical signal that generates the same rms output power in the ideal noise-free detector as the noise signal produced in a real detector.

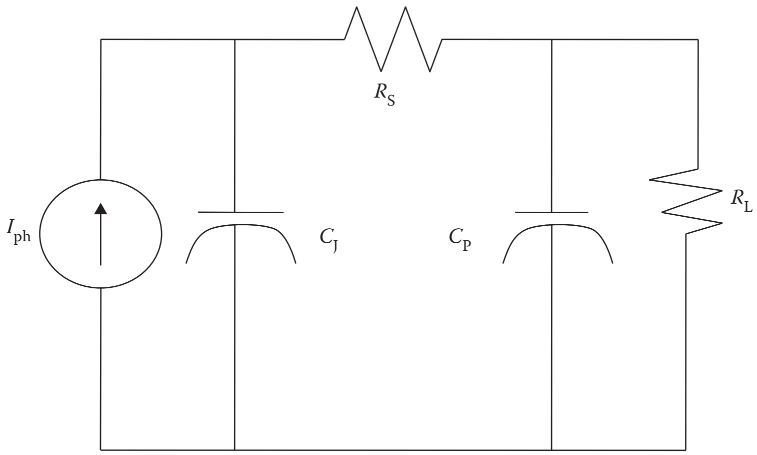


FIGURE 35.5 Electrical equivalent circuit of a photodiode. I_{ph} is the photogenerated current, and C_{J} and C_{p} are, respectively, the junction and parasitic capacitances. R_{S} and R_{L} are, respectively, the series and load resistances.

A unit NEP^* is defined as

$$NEP^* = \frac{NEP}{(A\Delta f)^{1/2}}, \tag{35.13}$$

where $A$is the area of the detector and Δf the bandwidth. The commonly used parameter is, however, the specific detectivity D^*, which is the reciprocal of NEP^*.

35.4 Classification of PDs

PDs can be classified in different ways. It may be an intrinsic or an extrinsic type. In the former, the absorption is band to band and the excess EHPs created produce a photocurrent. In the extrinsic types, the photon energy is less than the bandgap energy, and an impurity or deep-level defect in the forbidden gap is involved. An electron from this deep level is lifted by photon absorption process to the CB. Alternately, an electron in the VB may come to the deep level by photon absorption leaving behind a hole in the VB. In both the cases, the excess carriers created produce photocurrent. The absorption may also take place between two confined states in the CB or VB in a QW, QWR, or QD. Such processes give rise to the intersubband transitions.

PDs can be of two types: with or without internal gain. Photoconductors, avalanche PDs (APDs), and phototransistors show internal gain, that is, the primary photocurrent is multiplied. On the other hand, p-n or rather p-i-n PDs do not have any gain.

In another classification scheme, PDs may be classified as having either vertically illuminated or horizontally illuminated structures. In the former type, the absorbing layer thickness is low to reduce the transit time and to be useful for high-speed operation, though sacrificing the QE. However, a high packing density of arrays of PDs can be achieved, also facilitating coupling of the PDs with fibers. The low QE may be increased by enclosing the active layer within a resonant cavity structure, which is realized by using multilayer dielectrics (Unlu et al. 2001).

The simplest example of a horizontally illuminated PD is an edge-illuminated detector. However, due to diffraction, light spreads outside the high-field region and slow diffusion process leads to tails in the impulse response. The problem may be solved by using a waveguide structure, confining the light into the high-field region covering the optical waveguide. The speed of the devices is limited by the capacitance due to a thin intrinsic absorbing layer. A traveling wave detector may solve the limitations due to capacitance.

Another way of classifying the PDs is based on the type of absorbing layer. Various possibilities exist that include (1) bulk semiconductor, (2) QWs/multiple QWs (MQWs), (3) QWRs, (4) QDs, (5) strained quantum nanostructures, and (6) quantum nanostructures of different band alignments (see Basu 1997) such as type I, type II, type III, etc., as the active absorbing region.

35.5 Different PD Configurations

In this section, we shall discuss the structure and working principles of photoconductors, p-i-n and APDs, and PDs involving intersubband transitions. PDs relying on extrinsic processes will not be addressed. Detailed discussions about different types are given in many text books and reviews (see, for example, Bhattacharyya 1996; Bowers and Burrus 1987; Bowers and Wey 1995; Deen and Basu 2012; Kaiser 2013; Senior and Jamro 2009)

35.5.1 Photoconductive Detectors

The simplest form of PD is a photoconductor, in which the conductivity of the semiconductor is changed by absorption of photons. The changed photoconductivity produces photocurrent proportional to the incident optical power. In this section, the structure, principle, and noise analysis of such photoconductive detectors are presented.

35.5.1.1 Structure and Analysis

Consider a slab of semiconductor of length L, width W, and thickness d as shown in Figure 35.6. A voltage V is applied across the length and the slab is illuminated by light at the top. The current I flowing through the photoconductor may be expressed as

$$I = eWd(n\mu_e + p\mu_h)\left(\frac{V}{L}\right) = eWd\left(\frac{V}{L}\right)[(n_0\mu_e + p_0\mu_h) + (\Delta n\mu_e + \Delta p\mu_h)] = I_D + I_{ph}, \tag{35.14}$$

where I_D is the dark current, I_{ph} is the photocurrent, n_0 and p_0 are the equilibrium electron and hole concentrations, respectively, Δn (Δp) is the excess electron (hole) concentration, and μ_e (μ_h) is the electron (hole) mobility.

The average generation rate r_g within the photoconductor is expressed as follows:

$$r_g = \frac{\eta I_0 WL}{\hbar\omega WLd} = \frac{\eta I_0}{\hbar\omega d}, \tag{35.15}$$

where η is the efficiency, I_0 the light intensity at the surface, and $\hbar\omega$ the photon energy. Under steady state, the generation rate equals the recombination rate so that $r_g = r_r = \Delta n/\tau = \Delta p/\tau$, where τ is the recombination lifetime. Thus, one obtains

$$\Delta n = \Delta p = r_g\tau. \tag{35.16}$$

Substituting Equation 35.16 in Equation 35.14, the photocurrent is obtained as follows:

$$I_{ph} = \frac{Wd}{L} r_g\tau e(\mu_e + \mu_h)V. \tag{35.17}$$

The photoconductive gain, defined as the ratio of the rate of flow of electrons from the device to the rate of generation of EHPs within the device, takes the following form:

$$G = \frac{I_{ph}}{e}\frac{1}{r_g WLd} = \frac{\tau(\mu_e + \mu_h)V}{L^2}. \tag{35.18}$$

Equation 35.17 has been used to arrive at the last equality in Equation 35.18. The variation in gain with voltage V may be understood if we assume that $\mu_e >> \mu_h$, which is usually the case. The transit time of electrons through the device is $\tau_{tre} = L^2/\mu_e V$ and from Equation 35.18 $G = \tau/\tau_{tre}$. At low bias, $\tau < \tau_{tre}$ and the gain is <1. With larger V, electrons move faster, $\tau > \tau_{tre}$, and the gain becomes larger than unity.

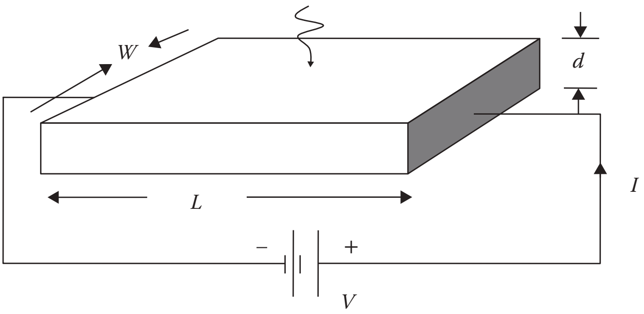


FIGURE 35.6 Schematic of a photoconductive PD. L, W, and d are the length, width, and depth of the photoconductor, respectively. V is the applied voltage and I is the current.

At still larger bias, the velocities of both electrons and holes are saturated. The gain is again reduced below unity. By following these arguments, it is easily concluded that higher values of gain may be achieved by increasing V, or by decreasing L, or by increasing the recombination lifetime.

35.5.1.2 Noise in Photoconductors

There are three main sources of noise in a photoconductive detector. The first is thermal or Johnson noise that arises due to random motion of carriers in the device. The mean squared noise current due to a conductor having finite resistance R is given as follows:

$$\left\langle i_{\mathrm{th}}^2 \right\rangle = \frac{4k_{\mathrm{B}}TB}{R}, \tag{35.19}$$

where B is the bandwidth of the device. In developing the noise equivalent circuit, the noisy photoconductor is replaced by a noiseless photoconductor and an ideal noise current or voltage generator.

Fluctuations in generation and recombination rates give rise to fluctuations in current and hence generate noise. The mean squared generation-recombination (GR) current is expressed as follows:

$$\left\langle i_{\mathrm{GR}}^2 \right\rangle = \frac{4eGI_0B}{1+\omega^2\tau^2}, \tag{35.20}$$

where I_0 is the steady photogenerated current. GR noise is also termed as shot noise. Since shot noise decreases as ω^{-2}, the thermal noise is the dominating noise source at high frequencies.

The third kind of noise, the flicker noise, has its origin in the presence of surface and interface defects and traps in the bulk material, and the mean squared current is expressed as:

$$\left\langle i_f^2 \right\rangle \propto \frac{1}{f}. \tag{35.21}$$

Equation 35.21 indicates that flicker noise is of importance at very low frequencies, less than 1 KHz.

In the noise equivalent circuit, the noise current sources appear in parallel. Assuming that the noise sources are uncorrelated and also that flicker noise is negligible, one may write for the SNR as

$$\frac{S}{N} = \frac{i_{\mathrm{ph}}^2}{\left\langle i_{\mathrm{th}}^2 \right\rangle + \left\langle i_{\mathrm{GR}}^2 \right\rangle}. \tag{35.22}$$

Let us assume that the optical signal incident on the detector is sinusoidally modulated by a modulating signal of angular frequency ω_{m}, and its power variation has the form $P_1 \exp(j\omega_{\mathrm{m}}t)$. The rms optical power is $P_1/\sqrt{2}$. The rms photocurrent is expressed as

$$i_{\mathrm{ph}} = \frac{eP_1\eta}{\sqrt{2}\hbar\omega}\frac{\tau}{\tau_{\mathrm{tr}}}\frac{1}{\left[1+\omega_{\mathrm{m}}^2\tau^2\right)^{1/2}}. \tag{35.23}$$

In arriving at Equation 35.23, the expression for responsivity, Equation 35.9, has been used to relate photocurrent with power and the gain is replaced by the ratio τ/τ_{tr}. The SNR as given by Equation 35.22 may now be written by using Equations 35.21 through 35.23 as

$$\frac{S}{N} = \frac{\eta P_1^2}{8B\hbar\omega P_0}\frac{1}{1+\frac{k_BT}{eG}\left[1+\omega_{\mathrm{m}}^2\tau^2\right)\frac{1}{RI_0}}. \tag{35.24}$$

35.5.2 p-n Photodiode

The simplest PD without internal gain is a reverse-biased p-n junction, shown schematically in Figure 35.7 along with the electric field profile. Both depletion and diffusion regions are shown. Photons are absorbed in the absorption region ($\sim 1/\alpha(\lambda)$). In case of weak absorption, the absorption region may cover the whole device, a part of which may contain the depletion region. Note that EHPs are generated in both the depletion and diffusion regions. While the electric field in the depletion region sweeps the photogenerated carriers quickly, the EHPs created in the diffusion region move slowly and thus limit the response of the PD. To improve the response, it is to be ensured that all photons are absorbed in the depletion region. Therefore, its width is made as large as possible by using lower doping.

35.5.3 p-i-n Photodiode

An efficient way to overcome the limitations of a p-n PD is to use a p-i-n configuration. Near the absorption edge, the weak absorption coefficient makes the absorption region quite long. The depletion region must match this, and to achieve this, the n region is low-doped to make it intrinsic (i). To make a low-resistance contact, a heavily doped n layer (n^+) is added. The resulting structure, a p-i-n structure, is shown in Figure 35.8 including the electric field profile, depletion and absorption regions. As shown, in this structure, the EHPs created are subject to the depletion layer field.

The speed of response of the p-i-n photodiode depends on (1) the transit time of carriers across the depletion layer as given by Equation 35.11 and (2) the RC time constant, where C is the junction capacitance.

The junction capacitance may be used either by using a mesa structure or by using a planar geometry with selective diffusion in the contact region (Bandyopadhyay and Deen 2001). The capacitance may also be reduced by increasing the width of the i layer. This increases the absorption and hence the QE, but increases the transit time, thereby decreasing the speed of response. To have high values of both QE and

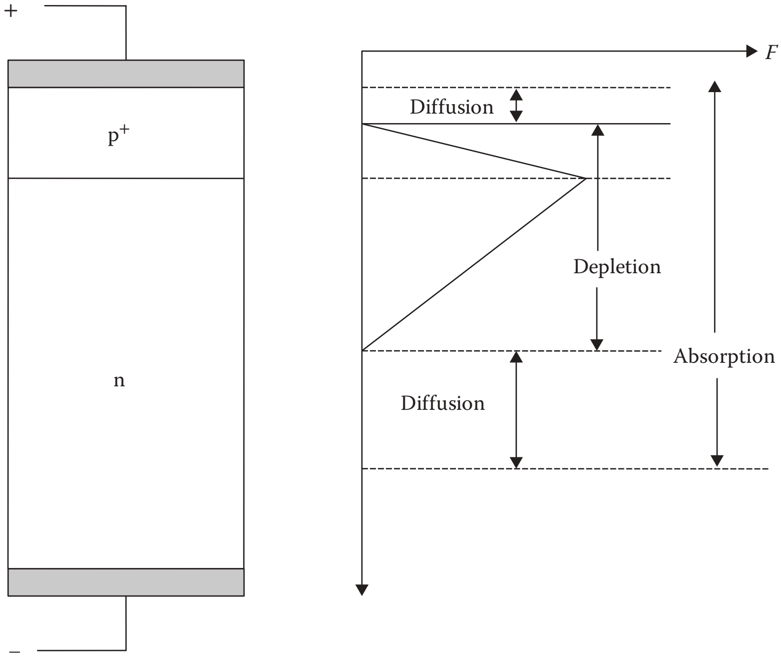


FIGURE 35.7 p-n PD structure (left) and electric field profile across the depth of the device (right). The depletion, diffusion, and absorption regions are shown.

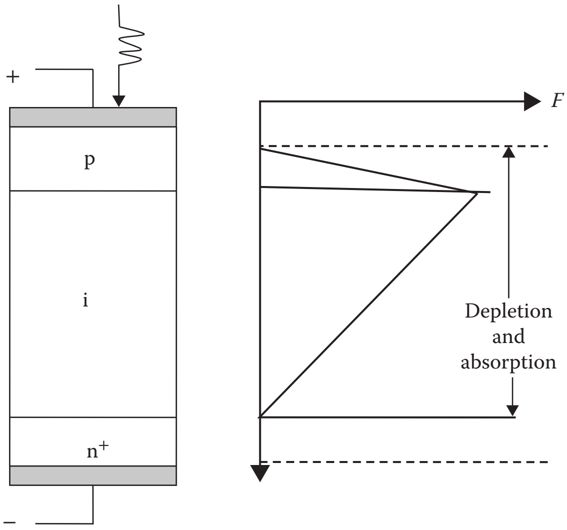


FIGURE 35.8 The p-i-n photodiode structure (left) and absorption and depletion regions and electric field (F) profile in the device (right).

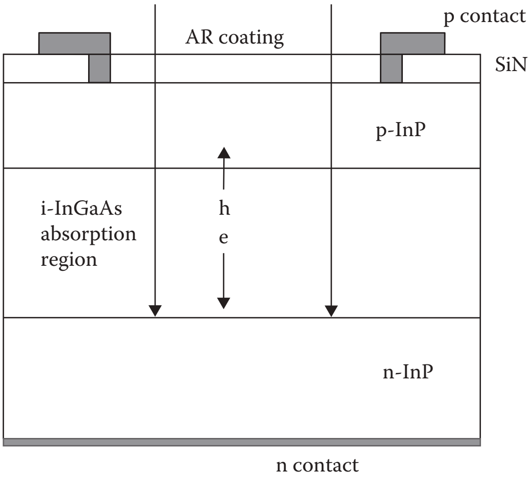


FIGURE 35.9 A cross-sectional view of a vertically illuminated p-i-n photodiode using p-InP/i-InGaAs/n-InP and SiN as antireflection (AR) coating.

speed (or bandwidth), the usual methods are (1) to employ an edge-illuminated structure or (2) to use a resonant-cavity-enhanced (RCE) structure.

A typical structure for a vertically illuminated p-i-n PD is shown in Figure 35.9. The n and p layers may be InP and the i layer is an undoped (or unintentionally doped) InGaAs layer. Since InP is transparent to 1.3–1.55 μm, the incident light is absorbed in the InGaAs layer, in which the electric field appears. Thus, there is no recombination or diffusion of the carriers. This leads to a fast response. The area of the absorbing layer must be optimized for giving high-speed operation (Bandyopadhyay and Deen 2001; Beling and Campbell 2009).

Edge-illuminated and RCE structures are presented in subsections 35.6.1 and 35.6.2, respectively.

35.5.4 Avalanche Photodiodes

APDs possess internal gain and as such are used to detect very weak optical signals. The working principle of an APD is discussed first and then some of the structures for improved performance are presented. The noise performance of APDs is discussed next.

35.5.4.1 Principles

The internal gain arises due to impact ionization in the presence of a high electric field in the depletion layer. A primary electron gets enough kinetic energy to break a covalent bond, thereby creating one secondary EHP. The primary and secondary carriers are accelerated by the depletion field and may again create tertiary EHPs. The multiplication process gives rise to an increased photocurrent. A primary hole can also initiate the ionization process. By this process, one initial electron or a hole may generate M extra EHPs, where M is the multiplication gain of the PD. In general, the ionization coefficients for electrons (α) and for holes (β) are different. The structure of an APD is shown in Figure 35.10, in which the electric field profile is also included.

In order to cause impact ionization, the carriers must possess a minimum energy, called the ionization threshold, which must exceed the bandgap energy. The ionization thresholds for electrons and holes are different. This leads to different values of ionization coefficients α and β, for electrons and holes, respectively. The ionization coefficient is the reciprocal of the average distance traversed by a carrier in the direction of the electric field to create an EHP. In its motion, a carrier may lose or gain energy via carrier–phonon interaction, which becomes more dominant with rise in temperature, thereby reducing the ionization probability. The breakdown voltage in a p-n junction therefore increases with temperature.

Defects in the device can lead to microplasma effect, in which breakdown at a lower voltage than in the remaining part of the junction occurs in a small area. This needs careful device processing and fabrication.

One of the important performance parameters for a PD used in optical communication is the maximum bit rate of operation. The detector must convert faithfully the incident optical pulses into electrical pulses. This is governed by how quickly the detector current responds to a pulse of photons. At low bias, carriers may be generated in the undepleted region and the response of an APD to an optical pulse may produce

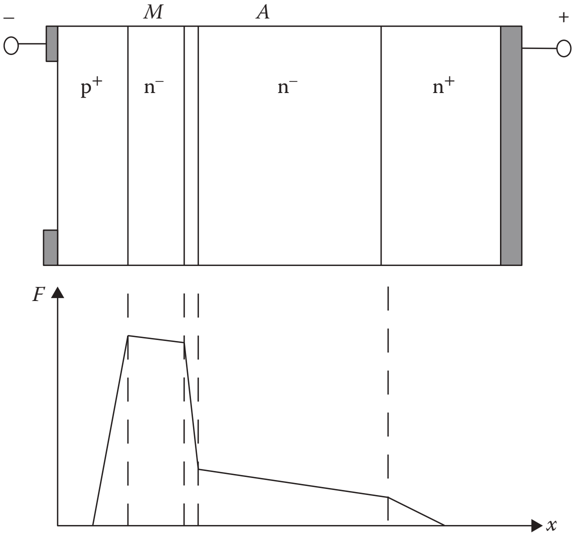


FIGURE 35.10 Structure of an APD (top) and the electric field (F) profile within it (bottom). M and A denote the multiplication and absorption regions, respectively.

a diffusion tail in the current pulse as in a p-i-n diode. When the bias is high enough to fully deplete the junction, the following three factors contribute to the response time:

1. Transit time of the carriers across the absorption layer of width W, given by Equation 35.11
2. The time taken by the carriers to create EHPs by avalanche multiplication
3. The RC time constant, where R is the device and load resistance and C is the junction capacitance

For low gain, the response time is determined by the transit time and the RC time constant. These two limit the value of bandwidth. When the gain is large, avalanche multiplication time becomes dominant and the bandwidth is reduced. APDs are therefore characterized by a constant gain bandwidth product.

35.5.4.2 Separate Absorption Grading and Multiplication and Separate Absorption, Grading, Charge Sheet, and Multiplication APDs

PDs for optical communication at a wavelength of 1.55 μm have $In_{0.53}Ga_{0.47}As$ as the absorbing layer. The usual p-n junction shows a high dark current due to the low bandgap of InGaAs. Furthermore, at high reverse field in the depletion region, the breakdown may be due to tunneling as opposed to the avalanche process. Therefore, the p-n junction is made using InP. The design involves separate absorption and multiplication (SAM). In an SAM structure, a sufficiently high electric field exists to cause avalanche multiplication in InP, and at the same time the field is low enough to prevent tunneling in the lower bandgap InGaAs layer in which optical absorption takes place. However, the photogenerated holes in the InGaAs layer while flowing toward InP layers find a band offset ΔE_v at the InGaAs/InP interface. Holes are thus trapped at the heterointerface, giving rise to a slower response. Incorporation of one or more layers of intermediate gap InGaAsP in between the absorption and multiplication regions eliminates this problem. The structure is called a separate absorption grading and multiplication (SAGM) structure.

The presence of quaternary and ternary layers in SAGM structures leads to another problem. The ionization rate of holes is higher than that of electrons in InP, but is lower in both the InGaAs and InGaAsP layers. This may give rise to unwanted multiplication in the ternary and quaternary layers. As the ionization ratio k approaches unity there, higher noise and lower speed are expected. The doping and thickness of the multiplication layer should be accurately controlled in order to have the increased gain–bandwidth product. A planar SAGM APD that avoids premature edge breakdown employs a partial charge sheet layer in between the multiplication layer and the grading region. The structure is known as a separate absorption, grading, charge sheet, and multiplication (SAGCM) structure. The structure consists of a lightly (or unintentionally) doped wide gap multiplication region where the field is high and an adjacent doped charge layer or a field control region. The presence of the charge (C) layer offers an extra degree of freedom such that the electric field profiles in the active region and in the periphery of the device can be controlled independently.

A cross-sectional view of an SAGCM APD using InP multiplication, InP charge, InGaAsP grading, and InGaAs absorption layer is shown in Figure 35.11. APDs used in telecommunications are reviewed by Campbell (2007).

35.5.4.3 Noise in APDs

Three important noise sources are operational in p-i-n PDs and in APDs. They are (1) thermal noise, (2) dark current noise, and (3) shot noise. In addition, APDs are characterized by an excess noise due to the random nature of the avalanche multiplication process. The statistical variation of multiplication is responsible for multiplication excess noise. The excess noise factor F depends on the ratio $k = \alpha/\beta$ and is expressed as

$$F = kM + (1-k)\left(2 - \frac{1}{M}\right) = M\left\{1 - \left[(1-k)\left(\frac{M-1}{M}\right)^2\right]\right\}. \tag{35.25}$$

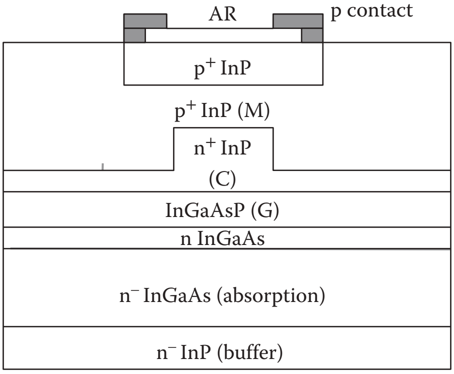


FIGURE 35.11 Cross-sectional view of an SAGCM APD. Antireflection (AR), multiplication (M), charge sheet (C), and graded (G) regions are shown.

The excess noise factor therefore reduces with large value of k. The total dark current in an APD is given by

$$I_\mathrm{d} = I_\mathrm{s} + I_\mathrm{dm} M, \tag{35.26}$$

where I_s is the unmultiplied surface current, and I_dm is the unmultiplied dark current. Let I_ph0 be the primary DC photocurrent, I_bk the photo-induced background current, and B the bandwidth. The mean square shot noise current is then

$$\langle I_M \rangle^2 = 2e\left[I_\mathrm{s} + \left(I_\mathrm{dm} + I_\mathrm{bk} + I_\mathrm{ph0}\right) M^2 F\right] B. \tag{35.27}$$

The SNR of APD is written as

$$\mathrm{SNR} = \frac{M^2 I_\mathrm{ph0}^2}{\langle I_M \rangle^2 + \left(4k_\mathrm{B}TB/R\right)} = \frac{M^2 I_\mathrm{ph0}^2}{2e\left[I_\mathrm{s} + \left(I_\mathrm{dm} + I_\mathrm{bk} + I_\mathrm{ph0}\right) M^2 F\right] B + \left(4k_\mathrm{B}TB/R\right)}. \tag{35.28}$$

In the above expression, the thermal noise contribution $\left(4k_\mathrm{B}TB/R\right)$ has been included. Note that by putting $M = 1$ and $F = 1$, we may obtain the SNR for p-i-n PD.

It is easy to assess the relative contributions of the shot noise (first term in the denominator of Equation 35.28) and of the thermal noise term (second term in the denominator), by dividing the numerator and denominator by M^2. The shot noise increases with M as does F. However, the thermal noise decreases with M as M^{-2}. It appears therefore that there exists an optimum value of M for which SNR can be a maximum. The variation of M is effected by change of reverse bias and may be expressed by the following empirical expression:

$$M \approx \frac{1}{\left\{1 - \left(\frac{V_\mathrm{a} - I_\mathrm{ph0}R}{V_\mathrm{B}}\right)^n\right\}}, \tag{35.29}$$

where V_a is the bias voltage, R the effective resistance, V_B the breakdown voltage, and n a constant that depends on the design of the device.

35.5.5 Schottky Barrier PD

In a Schottky barrier (SB) PD, the rectifying junction is formed between a metal and a semiconductor, rather than by a p-n junction. A schematic structure of an SB PD is shown in Figure 35.12, which indicates that the active or absorbing layer is quite narrow. This leads to short transit time and hence higher bandwidth, but to a lower QE.

35.5.6 Metal-Semiconductor-Metal PD

A metal-semiconductor-metal (MSM) PD is shown schematically in Figure 35.13. The schematic layer structure consists of a thin undoped layer grown on a semi-insulating substrate. Metal electrodes are then deposited on the active layer as interdigitated fingers. Each set of fingers forms a Schottky contact with the semiconductor and is connected to a large pad for connection to the external circuit. Both contacts in the MSM PD are on the same side of the substrate. Since the active layer thickness is small, the QE is less than in a p-n PD. Moreover, the reflections from the metal and semiconductor surfaces lead to smaller responsivity. However, the small transit time and very low capacitance greatly enhance the bandwidth.

A contact between a metal and a semiconductor may be either ohmic or rectifying. To obtain a rectifying contact, a certain combination of metal and a given semiconductor is needed. In such a junction, potential barrier develops due to the difference between the work functions of the metal and semiconductor. Another difference between a p-n junction and an SB diode is that both electrons and holes contribute to the current in a p-n junction, whereas SB diodes are majority carrier devices. The MSM PD consists of two junctions,

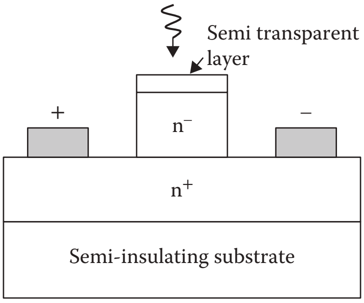


FIGURE 35.12 Schematic diagram of an SB PD. Shaded regions represent metallic contacts.

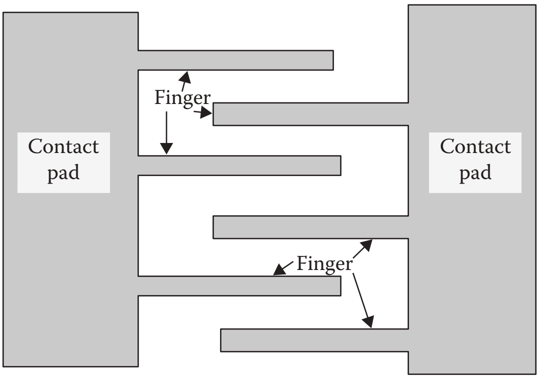


FIGURE 35.13 Top view of an MSM PD showing fingers and contact pads for connection to external circuits.

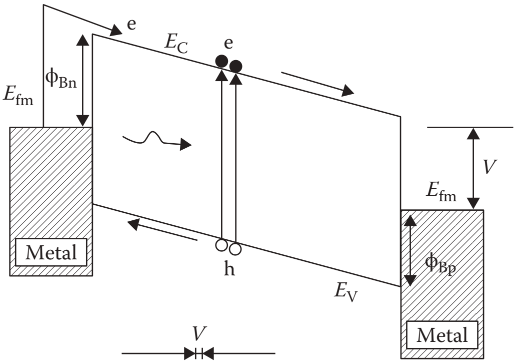


FIGURE 35.14 Band diagram of an MSM PD showing directions of flow of electrons and holes. E_c and E_v are conduction and valence band edges, respectively, E_{fm} is the Fermi level of the metal, ϕ_{Bn} and ϕ_{Bp} are the barrier heights at the n and p sides, respectively, and V is the applied bias between two metallic contacts. The PD is represented by two rectifying contacts shown at the bottom.

both of which may be rectifying or one junction ohmic and the other rectifying. The band diagram of an MSM PD is shown in Figure 35.14, in which the potential barrier ϕ_b is due to the work function difference. The two electrodes are connected serially, back-to-back, and the device contains two diodes as shown in Figure 35.14.

The dark current in the PD is due to the flow of thermally generated EHPs and by carriers from the metal surmounting the potential barrier ϕ_b. Under illumination, excess EHPs are generated and flow of electrons and holes is in opposite directions under the influence of the external bias V, as shown in Figure 35.14.

As stated already in connection with SB PDs, MSM PDs make a compromise between QE and bandwidth. The use of multiple fingers increases the QE. However, to increase the bandwidth, the effective absorption layer thickness needs to be reduced. Introduction of a highly doped layer at a certain depth may restrict the electric field within a certain absorption region and may provide a solution to the bandwidth problem.

Current improved technology may produce very thin fingers with narrow spacings between fingers. This reduces the transit time, but the speed of response is governed by the RC time constant. The following simple analytical expression for the capacitance may be used to calculate the time constant (Averine et al. 2001):

$$C = 0.226NL\varepsilon_0 \left[\varepsilon_S + 1\right) \left[6.5\theta^2 + 1.08\theta + 2.37\right). \tag{35.30}$$

In Equation 35.30, L is the length of a finger, ε_0 is the permittivity of free space, ε_S is the relative permittivity of the semiconductor, and $\theta = (D + t)/t$ is the finger-to-period ratio, where D is the width of the metal finger and t is the gap spacing between the metal fingers.

The EQE is given by

$$\eta = (1 - R)\frac{t}{t + D}\overset{\perp}{1} - \exp(-\alpha d)\big], \tag{35.31}$$

where R is the reflection coefficient, α the absorption coefficient, and d the thickness of the absorption layer of the semiconductor. The QE may be increased by replacing the metal electrodes by transparent conducting indium tin oxide (ITO), thereby reducing R. It can be increased by reducing the finger width at the cost of increased resistance however.

35.5.7 Phototransistors

Phototransistors may be a bipolar junction transistor (BJT) or metal-oxide-semiconductor field-effect transistor (MOSFET). The BJT may be either p-n-p or n-p-n type. It may be used in two-terminal (2T) configuration, with electrical contacts to the emitter and collector with the base floating, or in three-terminal (3T) configuration as in a normal BJT.

We first consider the 2T configuration. The structure is akin to a heterojunction bipolar transistor in which the emitter region is a higher bandgap material. The emitter is transparent to the incident light and thus acts as a window layer; the incident light falls directly on the base and is primarily absorbed in the base, base-collector depletion region, and the collector region. The heterojunction also increases the current gain. The base-collector junction is reverse biased so that the transistor acts as a reverse-biased PD.

Under no illumination the usual collector saturation current flows. However, when light is incident, the photogenerated electrons cross the depletion region in the base-collector junction and are collected at the collector terminal. The photogenerated holes travel in the opposite direction to the emitter terminal, thus reducing the emitter injection efficiency and the current gain.

The incident light is absorbed in the base and the base-collector depletion layer. The collector current for the (2T) configuration is given by (Chand et al. 1985):

$$I_c = I_{ph}(1 + h_{fe}), \tag{35.32}$$

where I_{ph} is the photogenerated current and h_{fe} the common emitter current gain of the BJT. A dark current also contributes to the collector current.

The currently used phototransistors are heterojunction bipolar transistors with larger gap and smaller gap semiconductors forming the emitter and base layers, respectively [see Basu et al. (2015) and Chang et al. (2016)] for work on GeSn-based heterojunction phototransistors and references therein for work on III-V compounds and alloy-based devices).

As stated already, a p-i-n diode is formed by the base, collector, and subcollector of a heterojunction phototransistor (HPT). Under optical illumination in a 2T-HPT, the photocurrent generated within the p-i-n diode is injected across the emitter-base junction and then amplified through transistor action. In this way, both the photocurrent in the p-i-n structure and the dc current gain determine the final collector photocurrent and the optical gain. When the incident optical power is low, the dc current gain is low, thereby making the optical gain low. This situation is improved in a 3T-HPT, in which a dc bias is applied to the base. The bias shifts the operating point to a higher collector current, making the dc current gain and optical gain higher.

A detailed analysis of the 3T-HPT has been presented by Frimel and Roenker (1997a,b) using a thermionic emission-diffusion model. The theory is too lengthy to be included here.

Useful references for 3T-HPTs for III–V compound-based materials as well as recently developed GeSn alloys may be found in Basu et al. (2015) and Chang et al. (2016).

In an MOS phototransistor, light incident at the source-gate region changes the surface potential under the gate oxide. This change causes a change in the current flowing through the channel.

35.5.8 QWIP and QDIP

It has been mentioned in Section 35.4 that the absorbing layer in a PD may contain quantum nanostructures. QWs and MQWs, in particular, are incorporated in the intrinsic (i) layer of a p-i-n PD, an APD, or in the base of a hetero phototransistor. The absorption in MQWs is from a valence subband to a conduction subband.

Quantum well infrared photodetectors (QWIPs) have QWs as the absorbing region, but the transition occurs from one subband to another in the same band, as shown in Figure 35.11. The intersubband

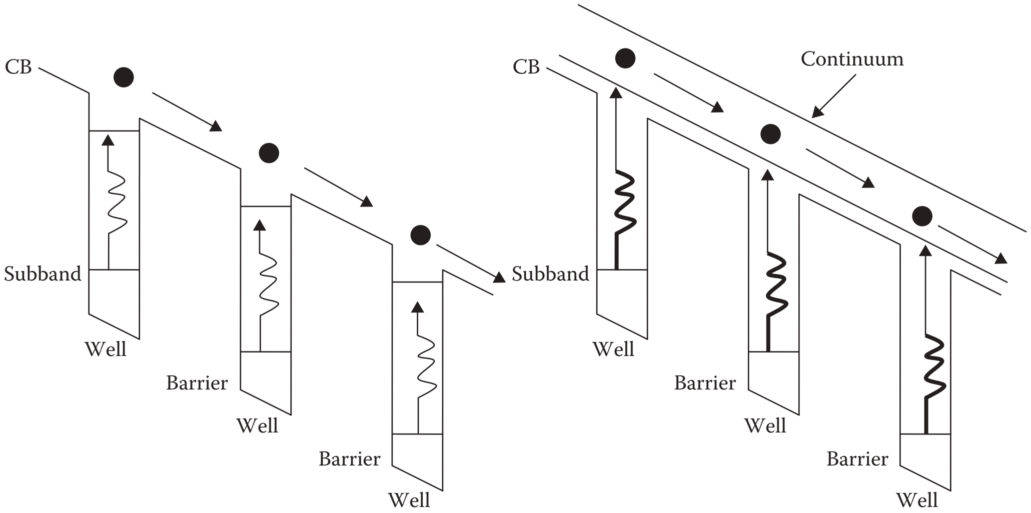


FIGURE 35.15 Intersubband absorption in QWIPs: (a) a bound-to-bound transition and (b) a bound-to-continuum absorption. The movement of electrons (solid balls) is shown in the figures.

transition energy is lower than the energy involved in band-to-band transitions in a QW. The QWIPs therefore work at the mid-infrared range (3–20 μm). Equation 35.4 may be used to calculate the energy difference between two subbands corresponding to the wavelength to be detected (see Equation 35.6). The width of the QW is then calculated from the value of the energy difference.

The types of transitions involved for CB are illustrated in Figure 35.15. The MQWs form the intrinsic layer of a p-i-n structure, in which the reverse bias causes the band bending. As shown in Figure 35.15a, the transition takes place from the lowest subband to the upper subband that lies close to the CB edge in the barrier layer. The electron lifted to the upper subband tunnels to the continuum states in the barrier and is drifted by the built-in electric field. In Figure 35.15b, an electron from the lone lowest subband absorbs the MIR photon and goes to the continuum states in the barrier and is then transported (see Schneider and Liu 2007).

Instead of QWs, multiple stacks of QDs can act as the absorbing layer, and the PD is called a quantum dot infrared photodetector (QDIP) (Barve et al. 2010).

35.5.9 Single-Photon Detectors

Single-photon emitters and detectors are needed for quantum information processing systems. Semiconductor-based single-photon detectors are APDs that operate with reverse bias exceeding the breakdown voltage. In this regime, the gain is enormously high. An incoming photon creates a charge, the number of which is multiplied (avalanche process) until it saturates at a current typically limited by an external circuit. In order that the single-photon APD should respond to a subsequent optical pulse, the saturated avalanche current must be terminated by lowering the bias voltage below the breakdown voltage (Eisaman et al. 2011).

35.6 Improved Devices

The PD structures such as p-i-n, APDs, etc., may show improved performances if some structural or materials modifications are introduced, some of which are discussed briefly in this section.

35.6.1 Waveguide PDs

In all the earlier discussions, vertically illuminated PD structures have been considered. For good PD performance, both the QE and speed of operation must be as high as possible. The absorbing layer in vertically illuminated devices must be very thin in order to reduce transit time; however, the thin layer cannot absorb the incident photons, thereby reducing the QE. To achieve high QE and high speed, edge-illuminating devices are used (Beling and Campbell 2014). The advantage of using waveguide structure has been pointed out in Section 35.4. The waveguide confines the otherwise spreading light into the high-field region. (Beling and Campbell 2014).

The waveguide photodetector (WPD) structure employs waveguides with absorbing layers either embedded in the waveguide or grown adjacent to the waveguide. The edge-coupled structure has been realized in p-i-n, APD, MSM, or uni traveling wave configuration.

The simplest type of edge-illuminated WPD structure is shown in Figure 35.16. The incident light falls directly into the edge of the absorbing layer. Here the direction of propagation of light is normal to the direction of transit of the photogenerated carriers. Note that the incident light may spread outside the absorbing region, making the optical confinement factor low and effective absorption less than in the bulk. However, the absorption length is quite large (~mm) so that responsivity and QE are enhanced. At the same time, the photogenerated carriers traverse a very short distance equal to the thickness of the absorbing layer making the transit time low. The increase in the capacitance due to the thin depletion layer may be arrested by reducing the area of the device. The state-of-the art technology for WPDs is discussed by Beling and Campbell (2014).

One drawback of the edge-coupled WPD is the misalignment of the optical field with the waveguide region. Use of lensed fiber or additional optics can increase the coupling efficiency. A useful way to improve coupling efficiency or to avoid the misalignment problem is to employ the evanescent waveguide structure as shown in Figure 35.17. In this configuration, the absorbing layer is placed adjacent to the waveguide core. The optical coupling is enhanced by inserting a transition layer between the waveguide and the detector. Since the waveguide mode has a smaller cross-section than that of a single-mode fiber, the input coupling efficiency is increased by using different approaches, for example, using a spot-size converter.

35.6.2 RCE Structure

The performance of a PD can be enhanced by placing the active device structure inside a Fabry–Pérot cavity. The cavity acts as a resonant structure providing wavelength selectivity and a large enhancement of the optical field, which in turn allows thinner PD layer. The device shows faster response and increased QE at the resonant wavelength. The wavelength selectivity and high-speed response make the devices the right candidates for wavelength division multiplexing applications (Unlu et al. 2001).

Photonic crystal structures have been employed to realize nanoscale PDs offering ultralow capacitance, thereby increasing the bandwidth and enabling ultralow power operation (Nozaki et al. 2013) (Figure 35.16).

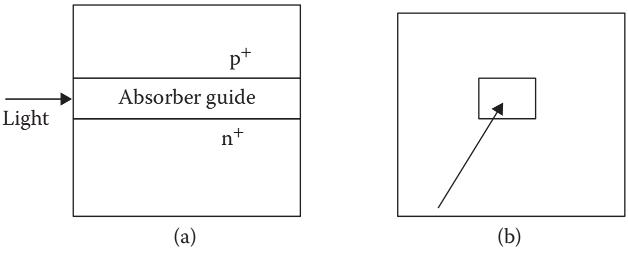


FIGURE 35.16 Schematic view of a WPD: (a) side view and (b) front view.

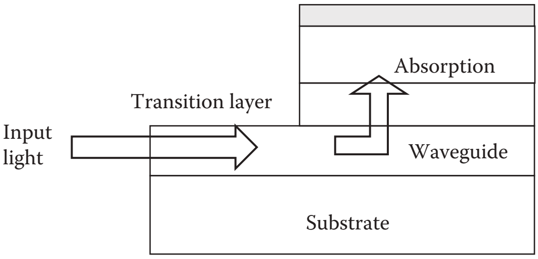

Absorption
Transition layer
Input light
Waveguide
Substrate

FIGURE 35.17 Schematic of an evanescently coupled WPD. Light from the waveguides couples into the absorption region via the transition layer.

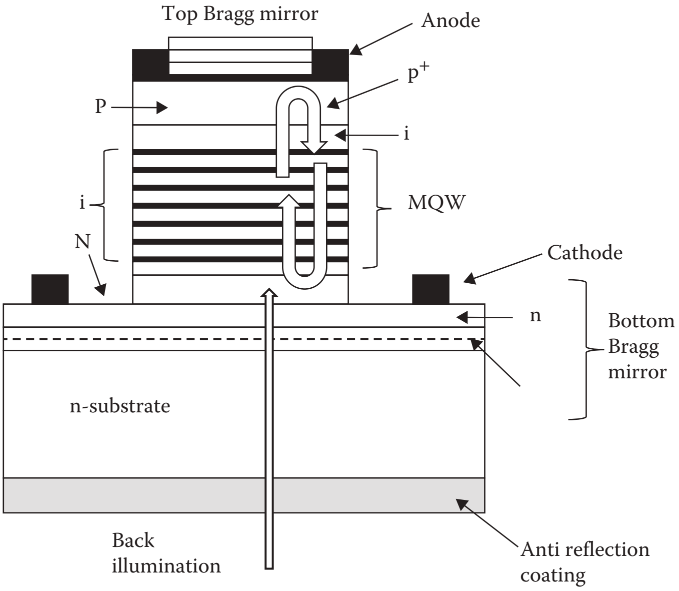

Top Bragg mirror
Anode
p+
P
i
i
MQW
N
Cathode
n
Bottom Bragg mirror
n-substrate
Back illumination
Anti reflection coating

FIGURE 35.18 Schematic diagram of a RCE PD. MQW, multiple quantum well.

35.6.3 Incorporation of Heterojunctions and Quantum Nanostructures

The active materials for detection are usually grown on a suitable substrate leading to heterojunctions. However, heterojunctions and MQWs and stacks of QDs are also used in the active regions of PDs. The heterostructures extend or control the range of wavelength to be detected. For example, in Ge_xSi_{1-x}/Si heterostructures or MQWs, the absorption takes place in the lower gap alloy material, and the heterobarrier prevents leakage of carriers, thus reducing the dark current.

35.7 Materials

The wavelength range covered by a PD depends on the active material, the bandgap of which determines the maximum wavelength (cutoff wavelength). Table 35.1 includes a few important materials and the corresponding wavelength range.

TABLE 35.1 Photodetector Materials and the Wavelength Range Covered

Material	Bandgap (eV)	Wavelength Range (μm)	Substrate
GaAs	1.42	<0.87	GaAs
$In_{0.53}Ga_{0.47}As$	0.81	1.3–1.6	InP
Si	1.12	0.2–1.12	Si
Ge	0.66	<1.8	Si
$Si_{1-x}Ge_x$	1.12–0.66	<1.8	Si
$Si_{1-x}Ge_x$/Si QWs	1.12–0.66	2–12	Si
$Ge_{1-x}Sn_x$ ($x > 0.08$)	0.58– ∼ 0	1.2–2.2	Si
GaAlAsSb/GaSb	0.60	0.75–1.9	GaSb
InGaAlAs/InGaAs/InP	–	0.5–2.1	InP
HgCdTe/CdTe	–	0.8–2.0	CdTe
$Ga_{1-x}Al_xN$	0.38–0.22	0.36–<0.2	Sapphire

Acknowledgments

The author is thankful to Dr. Rikmantra Basu of NIT Delhi, and Dr. Bratati Mukhopadhyay and Vedatrayee Chakraborty of Institute of Radio Physics and Electronics for their help in preparation of the manuscript.

References

Averine S. V., Chan Y. C., and Lam Y. L. (2001) Geometry optimization of interdigitated Schottky-barrier metal-semiconductor-metal photodiode structures. *Solid State Electron.*, vol. 45, 441–446.

Bandyopadhyay A. and Deen M. J. (2001) Photodetectors for optical fiber communications, Ch. 5 in *Photodetectors and Fiber Optics*, Nalwa H. S (ed) San Diego, CA: Academic Press

Barve A J., Lee S. J., Noh S. K., and Krishna S. (2010) Review of current progress in quantum dot infrared photodetectors. *Laser Photon. Rev.*, vol. 6, 738–750.

Basu P. K. (1997) *Theory of Optical Processes in Semiconductors: Bulk and Microstructures*. Oxford, UK: Oxford University Press.

Basu R., Chakraborty V., Mukhopadhyay B., and Basu P. K. (2015) Predicted performance of Ge/GeSn hetero-phototransistors on Si substrate at 1.55 μm. *Opt. Quant. Electron.*, vol. 47, no. 2., 387–399.

Beling A. and Campbell J. C. (2009) InP-based high-speed photodetectors. *J. Lightwave Technol*, vol. 27, no. 3, 343–355.

Beling A. and Campbell J. C. (2014) High-speed photodiodes. *IEEE J. Sel. Top. Quantum Electron*, vol. 20, no. 6, 3804507.

Bhattacharyya P. (1996) *Semiconductor Optoelectronic Devices*, 2nd edn. Upper Saddle Point, NJ: Prentice-Hall.

Bowers J. E. and Burrus, C. A., Jr. (1987) Ultrawide-band long-wavelength p-i-n photodetectors. *J. Lightwave Technol.*, vol. 5, no. 10, 1339–1350.

Bowers J. E. and Wey Y. G. (1995) High speed photodetectors, Ch. 17 in *Handbook of Optics, Fundamentals, Techniques and Design*, Vol. 1, Bass M (ed.). Pennsylvania, NY: McGraw-Hill.

Campbell J. C. (2007) Recent advances in telecommunications avalanche photodiodes. *J. Lightwave Technol*, vol. 25, no. 1, 109–121.

Chand N., Houston P. A., and Robson P. N. (1985) Gain of a heterojunction bipolar phototransistor. *IEEE Trans. Electron Dev*, vol. 32, no. 3, 622–627.

Chang G. E., Basu Rikmantra, Mukhopadhyay B. and Basu P. K. (2016) Design and modeling of GeSn-based heterojunction phototransistors for communication applications. *IEEE J. Sel. Topi. Quantum Electronics*, vol. 22, no. 6, 1–9.

Deen M. J. and Basu P. K. (2012) *Silicon Photonics: Fundamentals and Applications*, Chichester: John Wiley.

Eisaman M. D., Fan J., Migdall A., and Polyakov S. V. (2011) Single-photon sources and detectors. *Rev. Sci. Instruments*, vol. 82, 071101.

Frimel S. M. and Roenker K. P. (1997a) A thermionic-field-diffusion model for Npn bipolar heterojunction phototransistors. *J. Appl. Phys.*, vol. 82, 1427–1437.

Frimel S. M. and Roenker K. P. (1997b) Gummel-Poon model for Npn heterojunction bipolar phototransistor. *J. Appl. Phys.*, vol. 82, 3581–3592.

Kaiser G. (2013) *Optical Fiber Communication*, 5th edn. New Delhi; McGraw Hill Education (India) Pvt. Ltd.

Kumar S. and Deen M. J. (2014) *Fiber Optic Communications: Fundamentals and Applications*. Hoboken, NJ: Wiley.

Nozaki K., Matsuo S., Takeda K., Sato T., Kuramochi E., and Notomi M. (2013) InGaAs nanophotodetectors based on photonic crystal waveguide including ultracompact buried heterostructure. *Opt. Express*, vol. 21, no. 16, 19022–19028.

Schneider H. and Liu H. C. (2007) *Quantum Well Infrared Photodetectors: Physics and Applications*. Berlin: Spriner Verlag.

Senior M. and Jamro M. Y. (2009) *Optical Fiber Communications: Principles and Practice*. Harlow, England: Pearson Prentice-Hall.

Streetman B. G. and Banerjee S.K. *Solid State Electronic Devices*, 5th edn. Harlow, England: Pearson Education Ltd.

Unlu M. S., Ulu G., and Gokkavas M. (2001) Resonant cavity enhanced photodetectors, Ch. 2 in *Photodetectors and Fiber Optics*, Nalwa H. S (ed) San Diego, CA: Academic Press

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

36

P-N Junction Photodiodes

36.1 Introduction.. 307
36.2 Basic Equations and Models.. 309
36.3 Optical Generation Models 309
Optical Beam Absorption • Raytracing • Transfer Matrix Method • FDTD Method • Beam Propagation Method
36.4 Characteristic Input Parameters... 315
Minority Carrier Lifetimes • Interface Traps and Fixed Charges • Absorption Coefficients
36.5 Simulation Examples of P-N Junction Photodiodes 320
Physical Mechanism and Performance Parameters • Infrared Photodiodes • Visible Photodiodes • Ultraviolet Photodiodes • Two-Color Heterojunction Photodiodes
36.6 Summary... 334

Weida Hu

36.1 Introduction

Semiconductor p-n junction photodiodes have been available for half a century. By applying an electric field across the p-n junction, the light-generated carriers create an electrical current flow, thus the p-n junction photodiodes can convert optical signals into electrical signals (Shi and Li, 2014). Sensing applications in infrared, visible, and ultraviolet spectral bands promote theoretical research of these photodiodes. Numerical simulations, containing structural details such as layer thicknesses, doping profiles, and trap concentrations, provide key insights into device design and reliability degradation mechanisms, which could effectively reduce costly and time-consuming testing when developing and characterizing a new semiconductor device or technology. However, simulations of p-n junction photodiodes involve many important physical models, which play a decisive role in the calculations of accurate results. This chapter covers some basic equations and models used in the simulations of p-n junction photodiodes. Another key point is the accurate knowledge of device material parameters. Because of the complexity of an actual device structure, different device and material parameters are needed for various designs and structures. A number of these characteristic input parameters required for accurate modeling when building a basic theoretical framework are explained in this chapter.

First, this chapter presents several basic equations and models in detail. Second, principles and characteristics of several common optical generation models are reviewed and compared. Then, a number of important characteristic input parameters are introduced. Several p-n junction photodiode simulation examples are listed in the end of this chapter.

Figure 36.1 presents the general process simulation of p-n junction. After building the device structure, the device model could be defined by adding mesh generation and doping for different region. Different physical models are added in the optical/electrical simulation step, and then the desired information could be achieved. Figure 36.2 examples the results of some of the important characteristics of p-n junction.

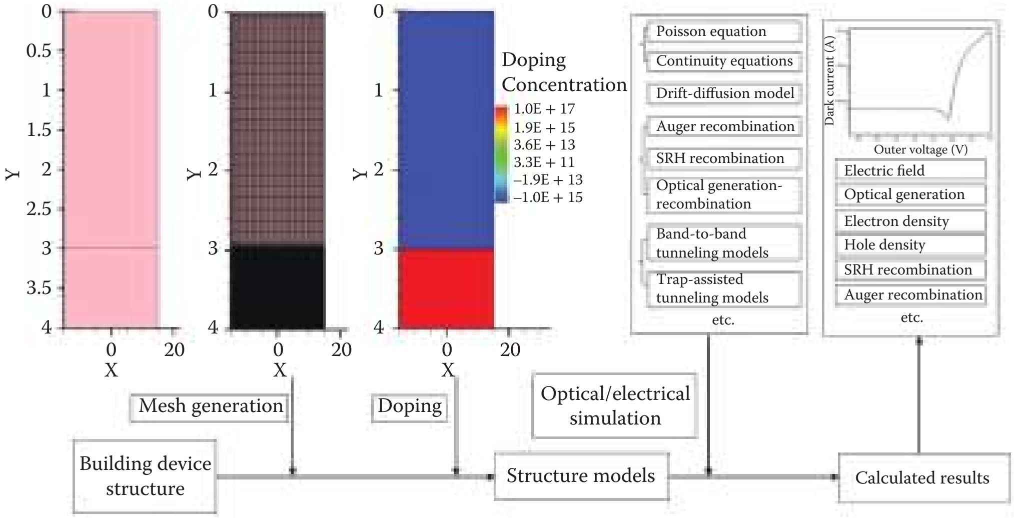

0
0.5
1
1.5
2
2.5
3
3.5
4
Y
0
20
X
Doping
Concentration
1.0E + 17
1.9E + 15
3.6E + 13
3.3E + 11
−1.9E + 13
−1.0E + 15
Poisson equation
Continuity equations
Drift-diffusion model
Auger recombination
SRH recombination
Optical generation-recombination
Band-to-band tunneling models
Trap-assisted tunneling models
etc.
Dark current (A)
Outer voltage (V)
Electric field
Optical generation
Electron density
Hole density
SRH recombination
Auger recombination
etc.
Mesh generation
Doping
Optical/electrical simulation
Building device structure
Structure models
Calculated results

FIGURE 36.1 The flow chart of numerical simulations of p-n junction photodiodes.

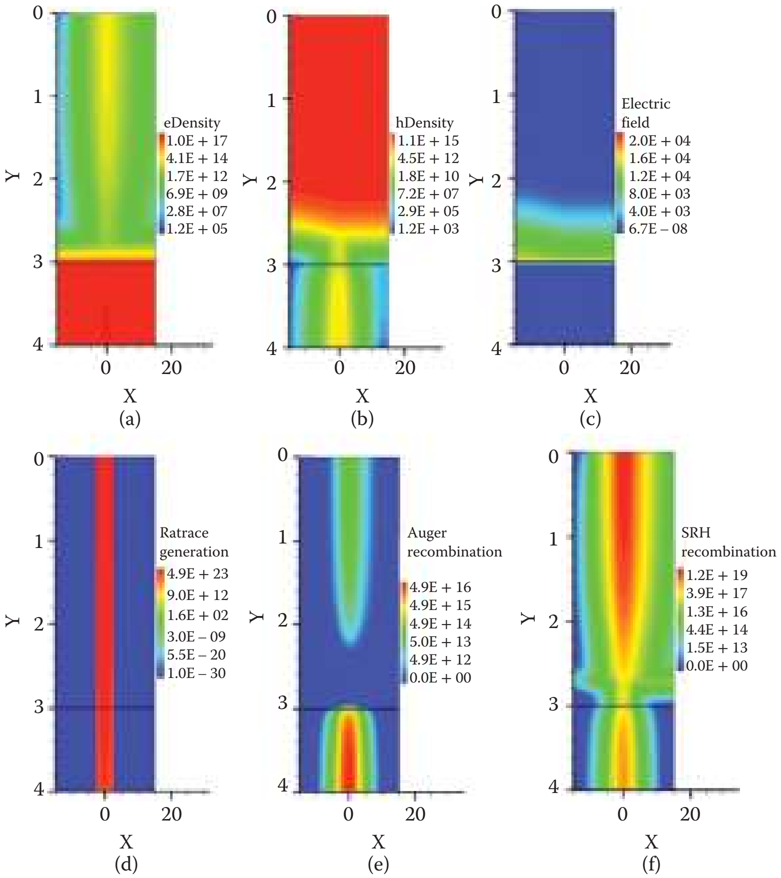


FIGURE 36.2 Important output/plots for a two-dimensional numerical simulation: (a) electron density (eDensity), (b) hole density (hDensity), (c) electric field, (d) raytrace generation, (e) Auger recombination, and (f) Shockley–Read–Hall (SRH) recombination.

36.2 Basic Equations and Models

Physical phenomena in semiconductor devices are very complicated and, for different devices and materials, models of different levels of complexity are needed in simulations. Different partial differential equations and models determine different microscopic physical mechanisms in a simulation, and can depend on different materials and device structures. This chapter introduces a number of widely used basic equations and models.

Depending on the p-n junction photodiodes under investigation and the level of modeling accuracy required, there are three governing equations for charge transport in semiconductor devices: *the Poisson equation* and *the electron and hole continuity equations*. The Poisson equation is as follows:

$$\nabla\varepsilon \cdot \nabla\psi = -q\left(p - n + N_{D^+} - N_{A^-}\right) \tag{36.1}$$

where ε is the electrical permittivity, ψ is the electrostatic potential, q is the elementary electronic charge, n and p are the electron and hole densities, N_{D^+} is the number of ionized donors, and N_{A^-} is the number of ionized acceptors. The electron and hole continuity equations are written as:

$$\frac{1}{q}\nabla\overrightarrow{J_n} + \left(G^{\text{opt}} - R\right) = 0 \tag{36.2}$$

$$\frac{1}{q}\nabla\overrightarrow{J_p} - \left(G^{\text{opt}} - R\right) = 0 \tag{36.3}$$

where R is the net electron–hole recombination rate, $\overrightarrow{J_n}$ the electron current density, and $\overrightarrow{J_p}$ is the hole current density.

Otherwise, *the drift-diffusion model* is widely used for the simulation of carrier transport in p-n junction photodiodes and is defined by the basic semiconductor equations, where current densities for electrons and holes are given by

$$\overrightarrow{J_\text{n}} = -nq\mu_\text{n}\nabla\phi_\text{n} \tag{36.4}$$

$$\overrightarrow{J_\text{p}} = -pq\mu_\text{p}\nabla\phi_\text{p} \tag{36.5}$$

where μ_n and μ_p are the electron and hole mobilities, ϕ_n and ϕ_p are the electron and hole quasi-Fermi potentials, respectively.

Additionally, *the thermodynamic model*, which is defined by a basic set of partial differential equations, and the lattice heat flow equations have been used in the simulations. The relations are generalized to include the temperature gradient as a driving term:

$$\overrightarrow{J_\text{n}} = -nq\mu_\text{n}\left(\nabla\phi_\text{n} + P_n\nabla T\right) \tag{36.6}$$

$$\overrightarrow{J_p} = -pq\mu_p\left(\nabla\phi_p + P_p\nabla T\right) \tag{36.7}$$

where P_n and P_p are the absolute thermoelectric powers.

36.3 Optical Generation Models

In the numerical simulation of the photoelectric detector, different structures and devices have different requirements for optical absorption models, including photoelectric incident position, power density, incident angle, reflection, refraction, and so on. Therefore, it is very important to select suitable optical

TABLE 36.1 Comparison of Some Common Optical Generation Models

	Optical Beam	Raytracing	Transfer Matrix	FDTD	Beam Propagation
Layered media	×	√	√	√	√
Anisotropy on horizontal direction	×	√	×	√	√
Geometrical optics	×	√	×	√	√
Diffraction	×	×	×	√	×
Reflection	×	√	√	√	√
Oblique incidence light	×	√	×	√	√

Note: FDTD = finite-difference time-domain.

generation models for different devices. The application conditions of some common optical generation models are listed in Table 36.1, and the characteristics of each model are introduced in this chapter.

Optical beam method is a simple optical absorption simulation method using Beer's law, and is only applicable to very simple optical process. Raytracing is capable of simulating a wide variety of optical effects such as reflection, refraction, and scattering, but its accuracy is severely limited, and cannot handle complex optical problems such as diffraction. Transfer matrix is very suitable for layered media, but requires media to be transversely isotropic. The FDTD method belongs in the general class of grid-based differential numerical modeling methods (finite difference methods). The time-dependent Maxwell's equations (in partial differential form) are discretized using central-difference approximations to the space and time partial derivatives, and could deal with more complex optical effects such as diffraction. Beam propagation method (BPM) is a quick and easy method of solving for fields in integrated optical devices, and a higher accuracy could be achieved. However, the BPM relies on the slowly varying envelope approximation, and is inaccurate for the modeling of discretely or fastly varying structures. What's more, BPM has some problems when dealing with scattering and diffraction problems.

36.3.1 Optical Beam Absorption

The optical beam absorption method supports the simulation of photogeneration using Beer's law. Multiple vertical photon beams can be defined to represent the incident light. The following equations describe useful relations for the photogeneration problem:

$$E_{\mathrm{ph}} = \frac{hc}{\lambda} \tag{36.8}$$

$$J_0 = \frac{P_0}{E_{\mathrm{ph}}} \tag{36.9}$$

where J_0 denotes the optical beam intensity (number of photons that cross an area of 1 cm^2 per 1 s) incident on the semiconductor device, P_0 is the incident wave power per area (W/cm^2), λ is the wavelength (cm), h is Planck's constant (J · s), c is the speed of light in a vacuum (cm/s), and E_{ph} is the photon energy that is approximately equal to $\frac{1.24}{\lambda\,(\mu\mathrm{m})}$ in eV.

The optical beam absorption model computes the optical generation rate along the z-axis taking into account that the absorption coefficient varies along the propagation direction of the beam according to

$$G^{\mathrm{opt}}(z,t) = J_0 F_t(t)\, F_{xy} \cdot \alpha(\lambda, z) \cdot \exp\left(-\left|\int_{z_0}^{z} \alpha(\lambda, z)\,\mathrm{d}z\right|\right) \tag{36.10}$$

where t is the time, $F_t(t)$ is the beam time behavior function: for a Gaussian pulse, it is equal to 1 for t in $[t_{\min}, t_{\max}]$ and shows a Gaussian distribution decay outside the interval with the standard deviation σ_t; F_{xy} is equal to 1 inside the illumination window and zero otherwise; z_0 is the coordinate of the semiconductor surface; and $\alpha(\lambda, z)$ is the nonuniform absorption coefficient along the z-axis.

The optical beam absorption method supports a wide range of window shapes as well as arbitrary beam time behavior functions. It is also not limited to beams propagating along the z-axis.

36.3.2 Raytracing

The optical generation method of raytracing supports the simulation of photogeneration in two dimensions (2Ds) and three dimensions (3Ds) for arbitrarily shaped structures. The calculation of refraction, transmission, and reflection follows geometric optics. A plane wave can be partitioned and each partition is represented by a 1D ray of light. Raytracing can approximate the behavior of a plane wave on a device by following such rays.

The optical generation along a ray, when the propagation is thought to be in the z-axis for a nonuniform absorption coefficient $\alpha(\lambda, x, y, z)$, can be written as a photon beam generation:

$$G^{\text{opt}}(z,t) = J\left(x,y,z_0\right)\alpha(\lambda,z)\cdot\exp\left(-\left|\int_{z_0}^{z}\alpha(\lambda,z)\,\mathrm{d}z\right|\right) \tag{36.11}$$

where $J(x, y, z_0)$ is the beam spatial variation of intensity over a window where rays enter the device, and z_0 is the position along the ray where absorption begins.

When a ray passes through a material boundary, the raytracer uses a recursive algorithm: It starts with a source ray and builds a binary tree that tracks the transmission and reflection of the ray. A reflection/transmission process occurs at interfaces with refractive index differences.

An incident ray impinges on the interface of two different refractive index (n_1 and n_2) regions, resulting in a reflected ray and a transmitted ray. The angles involved in raytracing are governed by Snell's law using

$$n_1\sin\left(\theta_1\right) = n_2\sin\left(\theta_2\right) \tag{36.12}$$

Absorption of photons occurs when there is an imaginary component (extinction coefficient), κ, to the complex refractive index. To convert the absorption coefficient to the necessary units, the following formula is used for power/intensity absorption:

$$\alpha(\lambda)\left[\text{cm}^{-1}\right] = \frac{4\pi\kappa}{\lambda} \tag{36.13}$$

In the complex refractive index model, the refractive index is defined element wise. In each element, the intensity of the ray is reduced by an exponential factor defined by $\exp(-\alpha L)$ where L is the length of the ray in the element. Therefore, the photon absorption rate in each element is

$$G^{\text{opt}}\left(x,y,z,t\right) = I\left(x,y,z\right)\left(1 - e^{(-\alpha L)}\right) \tag{36.14}$$

where $I(x, y, z)$ is the rate intensity of the ray in the element. After all of the photon absorptions in the elements have been computed, the values are interpolated onto the neighboring vertices and are divided by its sustaining volume to obtain the final absorption rate. The absorption of photons occurs in all materials with a positive extinction coefficient for raytracing. Depending on the quantum yield, a fraction of this value is added to the carrier continuity equation as a generation rate so that correct accounting of particles is maintained.

36.3.3 Transfer Matrix Method

The propagation of plane waves through layered media could be calculated by using a transfer matrix approach.

In the underlying model of the optical carrier generation rate, monochromatic plane waves with arbitrary angles of incidence and polarization states penetrating a number of planar, parallel layers are assumed. Each layer must be homogeneous, isotropic, and optically linear. In this case, the amplitudes of forward and backward running waves $A_j^{\pm}$ and $B_j^{\pm}$ in each layer shown in Figure 36.3 are calculated with the help of transfer matrices.

These matrices are functions of the complex wave impedances Z_j given by $Z_j = n_j \cdot \cos\theta_j$ in the case of E polarization (TE) and by $Z_j = n_j/\cos\theta_j$ in the case of H polarization (TM). Here, n_j denotes the complex index of refraction and θ_j is the complex counterpart of the angle of refraction ($n_0 \cdot \sin\theta_0 = n_j \cdot \sin\theta_j$).

The transfer matrix of the interface between layers j and $j + 1$ is defined by

$$T_{j,j+1} = \frac{1}{2Z_j} \cdot \begin{bmatrix} Z_j + Z_{j+1} & Z_j - Z_{j+1} \\ Z_j - Z_{j+1} & Z_j + Z_{j+1} \end{bmatrix} \tag{36.15}$$

The propagation of the plane waves through layer j can be described by the transfer matrix:

$$T_j\left(d_j\right) = \begin{bmatrix} \exp\left(2\pi i n_j \cos\theta_j \frac{d_j}{\lambda}\right) & 0 \\ 0 & \exp(-2\pi i n_j \cos\theta_j \frac{d_j}{\lambda}) \end{bmatrix} \tag{36.16}$$

with the thickness d_j of layer j, and the wavelength λ of the incident light. The transfer matrices connect the amplitudes shown in Figure 36.3 as follows:

$$\begin{bmatrix} B_j^+ \\ A_j^+ \end{bmatrix} = T_{j,j+1} \begin{bmatrix} A_{j+1}^- \\ B_{j+1}^- \end{bmatrix} \tag{36.17}$$

$$\begin{bmatrix} A_j^- \\ B_j^- \end{bmatrix} = T_j\left(d_j\right) \begin{bmatrix} B_j^+ \\ A_j^+ \end{bmatrix} \tag{36.18}$$

It is assumed that there is no backward running wave behind the layered medium, and the intensity of the incident radiation is known. Therefore, the amplitudes $A_j^{\pm}$ and $B_j^{\pm}$ at each interface can be calculated with appropriate products of transfer matrices.

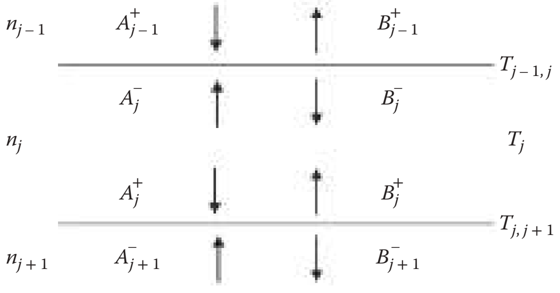


FIGURE 36.3 Wave amplitudes in a layered medium and transfer matrices connecting them.

For both cases of polarization, the intensity in layer j at a distance d from the upper interface $(j, j+1)$ is given by

$$I_{T(E,\mathrm{TM})}(d) = \frac{\Re(Z_j)}{\Re(Z_0)} \cdot \left\| T_j(d) \cdot \begin{pmatrix} A_j^- \\ B_j^- \end{pmatrix} \right\|^2 \tag{36.19}$$

with the proper wave impedances, $\Re(Z_j)$ means the real part of Z_j. If δ is the angle between the vector of the electric field and the plane of incidence, the intensities should be added according to

$$I(d) = aI_{\mathrm{TM}}(d) + I_{\mathrm{TE}}(d) \tag{36.20}$$

where $I_{\mathrm{TM}} = (1-a)I(d)$ and $I_{\mathrm{TE}} = aI(d)$ with $a = \cos^2\delta$.

36.3.4 FDTD Method

Light propagation through the sub-wavelength microstructure (such as nanoparticles, nanometal grating, photonic crystals, photo-trapping structures, and diffractive microlenses)-enhanced photodetectors is simulated by FDTD method based on a rigorous vector solution of Maxwell's equations. The computation area of the FDTD method is discretized into grid space where the electric field vector and the magnetic field vector are respectively arranged in the space and time step by 1/2, and the differential form of Maxwell's equations is rewritten as the central difference form with two-order accuracy.

In the calculation, these correspond to six time varying scalar equations in each Yee's grid (i.e., corresponding to the three spatial components of the electric and magnetic fields, as shown in Figure 36.4), along with the propagation of the electromagnetic field, the electromagnetic field is constantly updated in the set period until convergence. At this time, the corresponding electromagnetic field (light) has a stable distribution in the whole calculation area. At this point, the absorbed power density and the distribution of carrier density are calculated by assuming that each of the photons absorbed could excite an electron–hole pair.

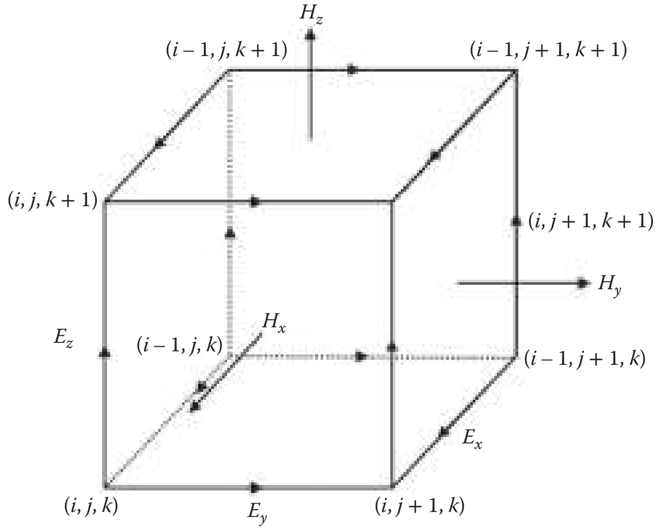


FIGURE 36.4 Structure of 3D Yee grid showing one cell of primary grid.

The differential forms of Maxwell equations are

$$\frac{\partial H}{\partial t} = -\frac{1}{\mu}\nabla \times E - \frac{\rho}{\mu}H \tag{36.21}$$

$$\frac{\partial E}{\partial t} = \frac{1}{\varepsilon}\nabla \times H - \frac{\sigma}{\varepsilon}E \tag{36.22}$$

Maxwell's curl equations can be discretized by the central difference method to achieve second-order accuracy. Equations 36.21 and 36.22 could be rewritten as formulas of electromagnetic field component:

$$\frac{\left.H_x\right|_{i,j,k}^{n+1/2} + \left.H_x\right|_{i,j,k}^{n-1/2}}{\Delta t} = \frac{1}{\mu_{i,j,k}}\left(\frac{\left.E_y\right|_{i,j,k+1/2}^{n} - \left.E_y\right|_{i,j,k-1/2}^{n}}{\Delta z} - \frac{\left.E_z\right|_{i,j+1/2,k}^{n} - \left.E_z\right|_{i,j-1/2,k}^{n}}{\Delta y} - \rho_{i,j,k}\cdot \left.H_x\right|_{i,j,k}^{n}\right) \tag{36.23}$$

$$\frac{\left.H_y\right|_{i,j,k}^{n+1/2} + \left.H_y\right|_{i,j,k}^{n-1/2}}{\Delta t} = \frac{1}{\mu_{i,j,k}}\left(\frac{\left.E_z\right|_{i+1/2,j,k}^{n} - \left.E_z\right|_{i-1/2,j,k}^{n}}{\Delta x} - \frac{\left.E_x\right|_{i,j,k+1/2}^{n} - \left.E_x\right|_{i,j,k-1/2}^{n}}{\Delta z} - \rho_{i,j,k}\cdot \left.H_y\right|_{i,j,k}^{n}\right) \tag{36.24}$$

$$\frac{\left.H_z\right|_{i,j,k}^{n+1/2} + \left.H_z\right|_{i,j,k}^{n-1/2}}{\Delta t} = \frac{1}{\mu_{i,j,k}}\left(\frac{\left.E_x\right|_{i,j+1/2,k}^{n} - \left.E_x\right|_{i,j-1/2,k}^{n}}{\Delta y} - \frac{\left.E_y\right|_{i+1/2,j,k}^{n} - \left.E_y\right|_{i-1/2,j,k}^{n}}{\Delta x} - \rho_{i,j,k}\cdot \left.H_z\right|_{i,j,k}^{n}\right) \tag{36.25}$$

$$\frac{\left.E_x\right|_{i,j,k}^{n+1} + \left.E_x\right|_{i,j,k}^{n}}{\Delta t} = \frac{1}{\varepsilon_{i,j,k}}\left(\frac{\left.H_z\right|_{i,j+1/2,k}^{n+1/2} - \left.H_z\right|_{i,j-1/2,k}^{n+1/2}}{\Delta y} - \frac{\left.H_y\right|_{i,j,k+1/2}^{n+1/2} - \left.H_y\right|_{i,j,k-1/2}^{n+1/2}}{\Delta z} - \sigma_{i,j,k}\cdot \left.E_x\right|_{i,j,k}^{n+1/2}\right) \tag{36.26}$$

$$\frac{\left.E_y\right|_{i,j,k}^{n+1} + \left.E_y\right|_{i,j,k}^{n}}{\Delta t} = \frac{1}{\varepsilon_{i,j,k}}\left(\frac{\left.H_x\right|_{i,j,k+1/2}^{n+1/2} - \left.H_x\right|_{i,j,k-1/2}^{n+1/2}}{\Delta z} - \frac{\left.H_z\right|_{i+1/2,j,k}^{n+1/2} - \left.H_z\right|_{i-1/2,j,k}^{n+1/2}}{\Delta x} - \sigma_{i,j,k}\cdot \left.E_y\right|_{i,j,k}^{n+1/2}\right) \tag{36.27}$$

$$\frac{\left.E_z\right|_{i,j,k}^{n+1} + \left.E_z\right|_{i,j,k}^{n}}{\Delta t} = \frac{1}{\varepsilon_{i,j,k}}\left(\frac{\left.H_y\right|_{i+1/2,j,k}^{n+1/2} - \left.H_y\right|_{i-1/2,j,k}^{n+1/2}}{\Delta x} - \frac{\left.H_x\right|_{i,j+1/2,k}^{n+1/2} - \left.H_x\right|_{i,j-1/2,k}^{n+1/2}}{\Delta y} - \sigma_{i,j,k}\cdot \left.E_z\right|_{i,j,k}^{n+1/2}\right) \tag{36.28}$$

Therefore, the electromagnetic field component can be solved by the Equations 36.23 through 36.28 for any given grid node at any moment.

For more detailed introduction of FDTD, see Chapter 52.

36.3.5 Beam Propagation Method

The BPM can be applied to find the light propagation and penetration into devices such as photodetectors. Its efficiency and relative accuracy are much better than the optical beam absorption method. The BPM solver is available for both 2D and 3D device geometries, where its computational efficiency compared with a full-wave approach becomes particularly apparent in 3Ds.

The BPM is based on the fast Fourier transform (FFT) and is a variant of the FFT BPM, which was developed by Feit and Fleck (1978).

The solution of the scalar Helmholtz equation is

$$\left[\nabla_t^2 + \frac{\partial^2}{\partial z^2} + k_0^2 n^2\left(x, y, z\right)\right] \phi\left(x, y, z\right) = 0 \tag{36.29}$$

At $z + \Delta z$ with $\nabla_t^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ and $n(x, y, z)$ being the complex refractive index can be written as:

$$\frac{\partial}{\partial z}\phi\left(x, y, z + \Delta z\right) = \mp i\sqrt{k_0^2 n^2 + \nabla_t^2}\,\phi\left(x, y, z\right) = \pm i\xi\phi\left(x, y, z\right) \tag{36.30}$$

In the paraxial approximation, the operator ξ reduces to

$$\xi = \sqrt{k_0^2 n_0^2 + \nabla_t^2} + k_0\delta n \tag{36.31}$$

where n_0 is taken as a constant reference refractive index in every transverse plane. By expressing the field ϕ at z as a spatial Fourier decomposition of plane waves, the solution to Equation 36.29 for forward-propagating waves reads

$$\phi\left(x, y, z + \Delta z\right) = \frac{1}{(2\pi)^2}\exp\left(ik_0\delta n\Delta z\right)\int_{-\infty}^{+\infty} d\vec{k_t}\exp\left(i\vec{k_t}\cdot\vec{r}\right)\exp\left(i\sqrt{k_0^2 n_0^2 \to \vec{k}_t^2}\,\Delta z\right)\tilde{\phi}\left(\vec{k}_t, z\right) \tag{36.32}$$

where $\tilde{\phi}$ denotes the transverse spatial Fourier transform. As can be seen from Equation 36.32, each Fourier component experiences a phase shift, which represents the propagation in a medium characterized by the reference refractive index n_0. The phase-shifted Fourier wave is then inverse transformed and given an additional phase shift to account for the refractive index inhomogeneity at each $\left(x, y, z + \Delta z\right)$ position. In the numeric implementation of Equation 36.32, an FFT algorithm is used to compute the forward and inverse Fourier transform.

For more detailed introduction of BPM, see Chapter Chapter 4.

36.4 Characteristic Input Parameters

The basic physical mechanism involved in numerical simulations of p-n junction photodiodes is very complicated and, depending on devices and use environments, different material parameters should be chosen for appropriate physical models. Some are related to bandgap such as effective mass, and another part of the parameters contributes to noise such as all of the recombination rate, while the absorption coefficient is dominant for the light absorption properties of device. In addition, the minority carrier lifetime of each region significantly influence the carrier transport. This chapter introduces some typical material parameters and related physical models (Table 36.2).

36.4.1 Minority Carrier Lifetimes

The minority carrier lifetime refers to the average survival time of minority carriers, and can be calculated as:

$$\tau = \frac{\Delta n}{U} \tag{36.33}$$

where Δn is the nonequilibrium carrier concentration and U is the carrier recombination rate.

TABLE 36.2 Parameters That Can Be Specified from Parameter File

Parameters	Units
Room-temperature bandgap, E_g	eV
Dielectric constant, ε_r	1
Density of states in conduction band, N_C	cm^{-3}
Density of states in valance band, N_V	cm^{-3}
Electron mobility, μ_e	$cm^2/V \cdot s$
Hole mobility, μ_h	$cm^2/V \cdot s$
Electron saturation velocity, v_{Se}	cm/s
Hole saturation velocity, v_{Sh}	cm/s
Electron effective mass (relative), m_e	1
Hole effective mass (relative), m_h	1
Electron Auger coefficient, R_{Ae}	cm^6/s
Hole Auger coefficient, R_{Ah}	cm^6/s
Electron SRH lifetime, τ_{Se}	s
Hole SRH lifetime, τ_{Sh}	s
Radiative coefficient, R_R	cm^3/s
Absorption spectra, A_1	cm^{-1}
Absorption spectra, A_2	cm^{-1}
Absorption spectra, E_1	eV
Absorption spectra, E_2	eV
Absorption spectra, P	1

Note: SRH, Shockley–Read–Hall.

The electron concentration n_0 and the hole concentration p_0 under the thermal equilibrium state fit the function:

$$n_0 p_0 = n_i^2 \tag{36.34}$$

where n_i is the intrinsic carrier concentration.

The generation-recombination rate of radiative recombination model is given by

$$U_R = r(np - n_i^2) \tag{36.35}$$

where r is the electron–hole recombination probability and n, p are the electron and hole concentration, respectively. The minority carrier lifetime is given by

$$\tau = \frac{1}{r(n_0 + p_0 + \Delta p)} \tag{36.36}$$

To the SRH recombination model, the generation-recombination rate can be written as

$$U_{SRH} = \frac{N_t r_n r_p (np - n_i^2)}{r_n \left(n + n_1\right) + r_p (p + p_1)} \tag{36.37}$$

with

$$n_1 = n_0 \exp\left(\frac{E_t - E_f}{k_0 T}\right) \tag{36.38}$$

$$p_1 = p_0 \exp\left(-\frac{E_t - E_f}{k_0 T}\right) \tag{36.39}$$

where r_n and r_p are the capture coefficient of electron and hole, N_t is the concentration of recombination center, E_t is the recombination center energy level, E_f is the Fermi energy level, and T is the lattice temperature. The SRH minority carrier lifetime can be calculated by

$$\tau = \frac{r_n\left(n+n_1\right)+r_p(p+p_1)}{N_t r_n r_p(n_0+p_0+\Delta p)} \tag{36.40}$$

When the surface recombination model is considered, a revised formula is used based on the SRH recombination model:

$$U_{\text{Surf}} = \frac{s_n s_p(np-n_i^2)}{s_p\left(n+n_1\right)+s_n(p+p_1)} \tag{36.41}$$

where s_n and s_p are the surface recombination velocity of electron and hole.

The rate of Auger recombination is given by

$$U_A = (\gamma_n n+\gamma_p p)(np-n_i^2) \tag{36.42}$$

where γ_n and γ_p are the auger recombination coefficient of electron and hole. The Auger recombination minority carrier lifetime can be written by

$$\tau = \frac{1}{\left(n_0+p_0+\Delta p\right)\left(\gamma_n n+\gamma_p p\right)} \tag{36.43}$$

36.4.2 Interface Traps and Fixed Charges

The following expressions for trap concentration versus the energy (E) define different types of distributions:

$$\begin{array}{ll}
N_0 \text{ for } E = E_0 & \text{Single-energy level} \\
N_0 \text{ for } E_0-0.5E_s < E < E_0+0.5E_s & \text{Uniform distribution} \\
N_0\exp\left(-\left|\dfrac{E-E_0}{E_S}\right|\right) & \text{Exponential distribution} \\
N_0\exp\left[-\dfrac{(E-E_0)^2}{2E_S^2}\right] & \text{Gaussian distribution} \\
\begin{cases} N_0 & \text{for} \quad E = E_1 \\ N_2 & \text{for} \quad E = E_2 \\ \cdots & \qquad \cdots \\ N_m & \text{for} \quad E = E_m \end{cases} & \text{User-defined table distribution}
\end{array} \tag{36.44}$$

Trapped electron and hole concentrations on one energy level are related to occupation probabilities for electrons (r_n) and holes (r_p) as follows:

$$\begin{array}{ll} n_{Dt} = N_{Dt} r_n & n_t = N_{Et} r_n \\ p_{At} = N_{At} r_p & p_t = N_{Ht} r_p \end{array} \tag{36.45}$$

where n_{Dt} is the electron concentration of the donor trap level, p_{At} is the hole concentration of the acceptor trap level, n_t is the electron concentration of the neutral electron trap level, and p_t is the hole concentration of the neutral hole trap level, N_{Dt} is the donor trap concentration, N_{At} is the acceptor trap concentration,

N_{Et} is the neutral electron trap concentration, N_{Ht} is the neutral hole trap concentration. It's obvious that $r_{\mathrm{n}} = 1 - r_{\mathrm{p}}$.

In consideration of the presence of traps, the Poisson equation is revised:

$$\nabla\varepsilon \cdot \nabla\psi = -q\left(p - n + N_{\mathrm{D}^+} - N_{\mathrm{A}^-} + \sum_{E_{\mathrm{t}}} N_{\mathrm{Dt}} - n_{\mathrm{Dt}} - \sum_{E_{\mathrm{t}}} N_{\mathrm{At}} - p_{\mathrm{At}} + \sum_{E_{\mathrm{t}}} p_{\mathrm{t}} - \sum_{E_{\mathrm{t}}} n_{\mathrm{t}}\right) \tag{36.46}$$

where E_{t} means trap energy levels.

The balance of the carrier flow to and from a trap level gives the following expression for the electron concentration of the trap level:

$$\frac{\mathrm{d}n_{\mathrm{t}}}{\mathrm{d}t} = v_{\mathrm{th}}^{n}\sigma_{\mathrm{n}}\mathrm{N}_{\mathrm{Et}}\left[\frac{n_1}{g_{\mathrm{n}}}r_{\mathrm{n}} - n(1 - r_{\mathrm{n}})\right] + v_{\mathrm{th}}^{p}\sigma_{\mathrm{p}}\mathrm{N}_{\mathrm{Et}}\left[\frac{p_1}{g_{\mathrm{p}}}(1 - r_{\mathrm{n}}) - pr_{\mathrm{n}}\right] \tag{36.47}$$

where $v_{\mathrm{th}}^{\mathrm{n}}$ and $v_{\mathrm{th}}^{\mathrm{p}}$ are the electron and hole thermal velocities, σ_{n} and σ_{p} are the electron and hole capture cross sections, and g_{n} and g_{p} are the electron and hole degeneracy factors, respectively.

In the steady state, the trap-assisted recombination process can be analyzed by standard SRH recombination; however, to the transient case, the recombination processes for electrons and holes are different, and the SRH form cannot be applied. The differential equations for the electron-related occupation probability at each energy level are

$$\frac{\mathrm{d}r_{\mathrm{n}}}{\mathrm{d}t} + \left[v_{\mathrm{th}}^{\mathrm{n}}\sigma_{\mathrm{n}}\left(n + \frac{n_1}{g_{\mathrm{n}}}\right) + v_{\mathrm{th}}^{\mathrm{p}}\sigma_{\mathrm{p}}\left(p + \frac{p_1}{g_{\mathrm{p}}}\right)\right]r_{\mathrm{n}} = v_{\mathrm{th}}^{\mathrm{n}}\sigma_{\mathrm{n}}n + v_{\mathrm{th}}^{\mathrm{p}}\sigma_{\mathrm{p}}\frac{p_1}{g_{\mathrm{p}}} \tag{36.48}$$

This equation is solved self-consistently with the transport and Poisson equations.

The trap total recombination term in the continuity equations is the sum for all levels:

$$R_{\mathrm{trap}} = \sum_{E_{\mathrm{t}}} R_{\mathrm{Dt}} + \sum_{E_{\mathrm{t}}} R_{\mathrm{Et}} + \sum_{E_{\mathrm{t}}} R_{\mathrm{At}} + \sum_{E_{\mathrm{t}}} R_{\mathrm{Ht}} \tag{36.49}$$

At nonheterointerface vertices, bulk traps are considered only from the region with the lowest bandgap. In cases where the bulk traps from other regions are important, assume appropriate value and solve nonlinear Poisson equation by changing the numeric damping of the trap charge, which properly considers bulk traps from all adjacent regions.

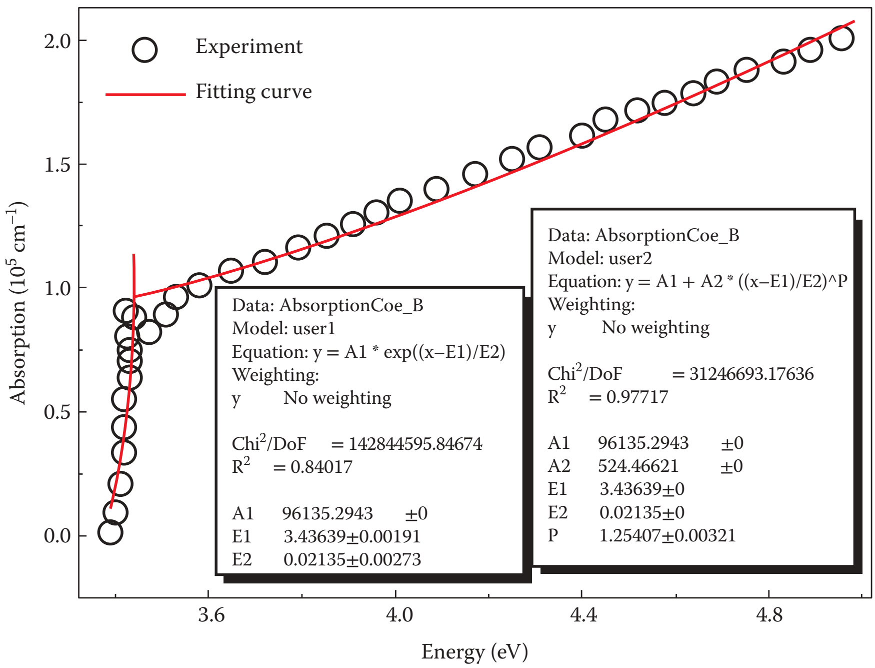
36.4.3 Absorption Coefficients

A widely used absorption coefficient model that fits all the materials is

$$\alpha\left(E_{\mathrm{ph}}\right) = \begin{cases} \alpha_1 \exp\left[\dfrac{\left(E_{\mathrm{ph}} - E_1\right)}{E_2}\right] & E_{\mathrm{ph}} < E_1 \\ \alpha_1 + \alpha_2 \exp\left[\dfrac{\left(E_{\mathrm{ph}} - E_1\right)}{E_2}\right]^{p} & E_{\mathrm{ph}} \geq E_1 \end{cases} \tag{36.50}$$

where E_{ph} is the photon energy, and $\alpha_1, \alpha_2, E_1, E_2$, and p could be specified in the parameter file.

The abovementioned parameters can be obtained by fitting the data acquired in the experiment. A fitting result (Wang et al., 2015) for GaN and AlGaN are shown in Figures 36.5 and 36.6.



Data: AbsorptionCoe_B
Model: user1
Equation: y = A1 * exp((x–E1)/E2)
Weighting:
y No weighting

Chi^2/DoF = 142844595.84674
R^2 = 0.84017

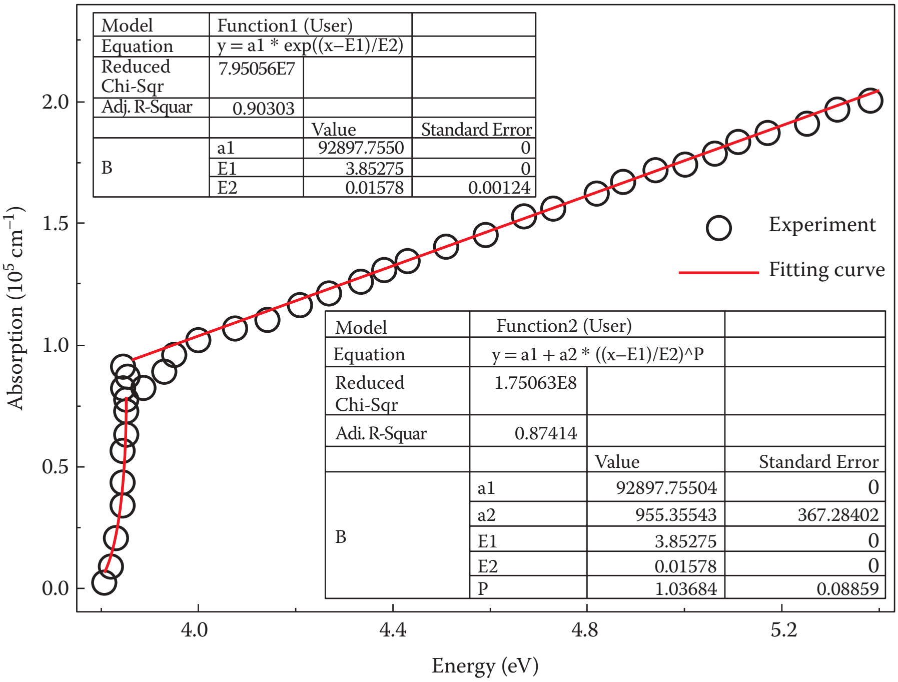
A1 96135.2943 ±0
E1 3.43639±0.00191
E2 0.02135±0.00273

Data: AbsorptionCoe_B
Model: user2
Equation: y = A1 + A2 * ((x–E1)/E2)^P
Weighting:
y No weighting

Chi^2/DoF = 31246693.17636
R^2 = 0.97717

A1 96135.2943 ±0
A2 524.46621 ±0
E1 3.43639±0
E2 0.02135±0
P 1.25407±0.00321

FIGURE 36.5 Fitting results of absorption spectrum for GaN at room temperature.



Model	Function1 (User)		
Equation	y = a1 * exp((x–E1)/E2)		
Reduced Chi-Sqr	7.95056E7		
Adj. R-Squar	0.90303		
		Value	Standard Error
B	a1	92897.7550	0
	E1	3.85275	0
	E2	0.01578	0.00124

Model	Function2 (User)		
Equation	y = a1 + a2 * ((x–E1)/E2)^P		
Reduced Chi-Sqr	1.75063E8		
Adi. R-Squar	0.87414		
		Value	Standard Error
B	a1	92897.75504	0
	a2	955.35543	367.28402
	E1	3.85275	0
	E2	0.01578	0
	P	1.03684	0.08859

FIGURE 36.6 Fitting results of absorption spectrum for $Al_{0.15}Ga_{0.85}N$ at room temperature.

An absorption model (Rajkanan et al., 1979) is implemented for silicon and materials with similar bandgap structure:

$$\alpha(T) = \sum_{i=1}^{2}\sum_{j=1}^{2} C_i A_j \left\{ \frac{\left[hf - E_{gj}(T) + E_{pi}\right]^2}{e^{-\frac{E_{pi}}{kT}} - 1} + \frac{\left[hf - E_{gj}(T) - E_{pi}\right]^2}{1 - e^{-\frac{E_{pi}}{kT}}} \right\} + A_d \left[hf - E_{gd}(T)\right]^{1/2} \tag{36.51}$$

where

$$E_g(T) = E_g(0) - \frac{\beta T^2}{T + \gamma} \tag{36.52}$$

and T is the temperature. Parameter values used in this formula can be changed for some materials with bandgap structures similar to silicon or for silicon itself.

Instead, the complex refractive index model is adopted to replace the absorption coefficient model in some numerical simulation commercial software. The complex refractive index $\tilde{n}$ can be written as

$$\tilde{n} = n + i \cdot k \tag{36.53}$$

where n refers to the refractive index, and k to the extinction coefficient. The numerical conversion of the extinction coefficient k and the absorption coefficient α is

$$\alpha = \frac{4\pi k}{\lambda} \tag{36.54}$$

For specific extinction coefficients of different materials, one can refer to the website http://refractiveindex.info/.

36.5 Simulation Examples of P-N Junction Photodiodes

In the following section, examples of some widely used p-n junction photodiodes are introduced with the calculation results and parameters shown graphically. First, examples about visible photodiodes using different materials and structure are given. Second, photodiodes simulations applied to ultraviolet and infrared wave band are introduced. Finally, we modeled a two-color heterojunction photodiode, in which more complex mechanism and structure are included. All the absorption models involved in examples are based on Equation 36.50.

36.5.1 Physical Mechanism and Performance Parameters

In order to obtain comprehensive understanding of photoresponse (including dark current) mechanisms of p-n junction photodiodes and improve the device performance, these key parameters, such as electron field (when the effect of electric field intensity becomes dominant, the dark current will be increased and the photoresponse starts reducing); different kinds of junction types (see Figure 36.7a and b), which have a great effect on technological requirements, surface leakage current, and so on; dark current (see Figure 36.7c through f), which will influence our device performance; as well as responsivity and quantum efficiency (QE) should be taken into consideration. In addition, surface leakage current is another consideration of p-n junction photodiodes. A detailed description of the surface leakage current of HgCdTe infrared photodiode is shown in Figure 36.8.

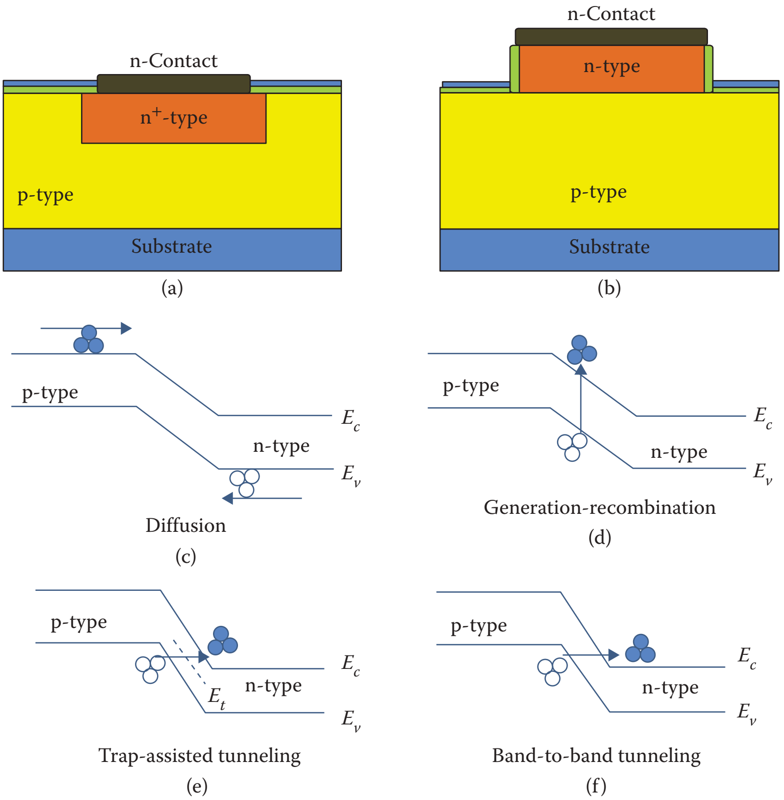

n-Contact
n$^+$-type
p-type
Substrate
(a)
n-Contact
n-type
p-type
Substrate
(b)
p-type
E_c
n-type
E_v
Diffusion
(c)
p-type
E_c
n-type
E_v
Generation-recombination
(d)
p-type
E_c
E_t
n-type
E_v
Trap-assisted tunneling
(e)
p-type
E_c
n-type
E_v
Band-to-band tunneling
(f)

FIGURE 36.7 Schematic structures of two types of main structure of p-n junction for (a) planar junction type, (b) mesa junction type, and the main dark current components of the bulk p-n junction: (c) diffusion, (d) generation-recombination, (e) trap-assisted tunneling, and (f) direct band-to-band tunneling.

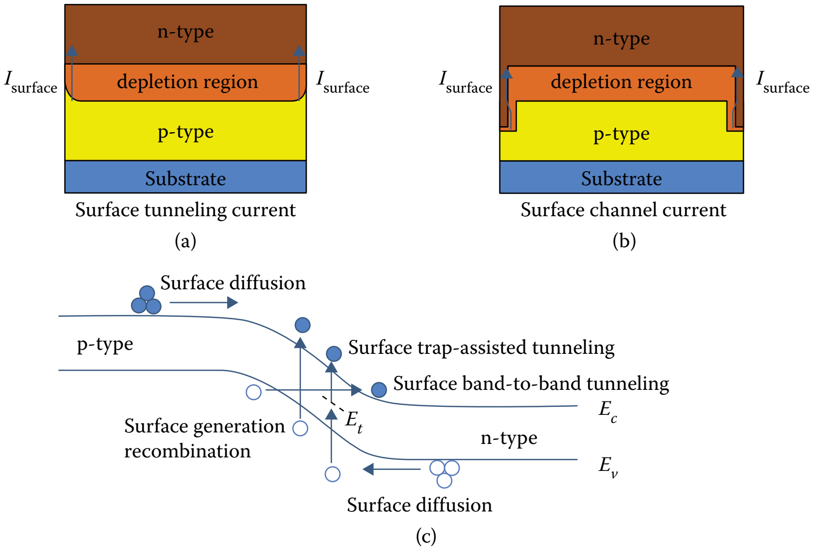

n-type
depletion region
p-type
Substrate
$I_{surface}$
$I_{surface}$
Surface tunneling current
(a)
n-type
depletion region
p-type
Substrate
$I_{surface}$
$I_{surface}$
Surface channel current
(b)
Surface diffusion
p-type
Surface trap-assisted tunneling
Surface band-to-band tunneling
E_c
E_t
Surface generation recombination
n-type
E_v
Surface diffusion
(c)

FIGURE 36.8 Surface leakage current of HgCdTe infrared photodiode: (a) surface tunneling current, (b) surface channel current, (c) the main dark current components of the surface-induced p-n junction surface diffusion current, surface generation-recombination, surface trap-assisted tunneling, and surface direct band-to-band tunneling.

The responsivity, quantum efficiency (QE) and cross-talk are index parameters measuring the ability of the device to acquire the target radiation signal. Responsivity is the light current formed by unit radiation power. The calculation formula is

$$R = \frac{I_0}{P_{\text{opt}}} \tag{36.55}$$

where R is the responsivity, I_0 is the optical current under zero bias, and P_{opt} is the power of incident light.

QE means the efficiency of the device with which the incident photons of light are converted to current:

$$\eta = \left(\frac{I_{\text{p}}}{q}\right)\left(\frac{P_{\text{opt}}}{hv}\right)^{-1} = \left(\frac{I_0}{P_{\text{opt}}}\right)hv = R\frac{1.24}{\lambda} = \frac{1.24R}{\lambda} \tag{36.56}$$

where λ is the wavelength of incident light and I_{p} is the light excitation current when the wavelength is λ (corresponding to the photon energy hv) and the power of incident light is P_{opt}.

Cross-talk is when the radiation signal in a unit produces a response signal R_0, the adjacent unit also appears in the response signal R_1. Cross-talk can be used to measure the image quality detection. The cross-talk is defined by:

$$C = \frac{R_1}{R_0} \times 100\% \tag{36.57}$$

36.5.2 Infrared Photodiodes

36.5.2.1 InP/In$_{0.53}$Ga$_{0.47}$As/InP PIN Photodiode

For an InP/In$_{0.53}$Ga$_{0.47}$As/InP PIN photodiode (Wang et al., 2008), the thickness of the absorption layer (T_{abs}) as a key structural parameter can determine, to a large extent, the magnitude of responsivity. The following discussion is concentrated on the effect of T_{abs} on the responsivity, and the following simulations are performed only at room temperature. Figure 36.9 shows the 2D structure of the device. The structure consists of a 0.5-μm heavily doped n-type InP buffer layer, a lightly doped n-type InGaAs absorption layer, and a 0.2-μm heavily doped p-type InP cap layer. The thickness of the InGaAs absorption layer (T_{abs}) is set as 1.5, 3.5, and 5.5 μm. Figure 36.10 presents the energy band of InP/InGaAs PIN device structure, respectively.

Here, the thickness of the InGaAs absorption layer is set as 1.5 μm. Vertical distributions of electric field with varying T_{abs} from 1.5 to 5.5 μm in 2-μm steps are investigated as shown in Figure 36.11; the dark and

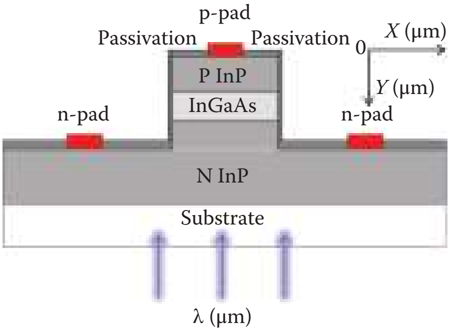


FIGURE 36.9 Schematic cross section of InP/InGaAs PIN photodiode.

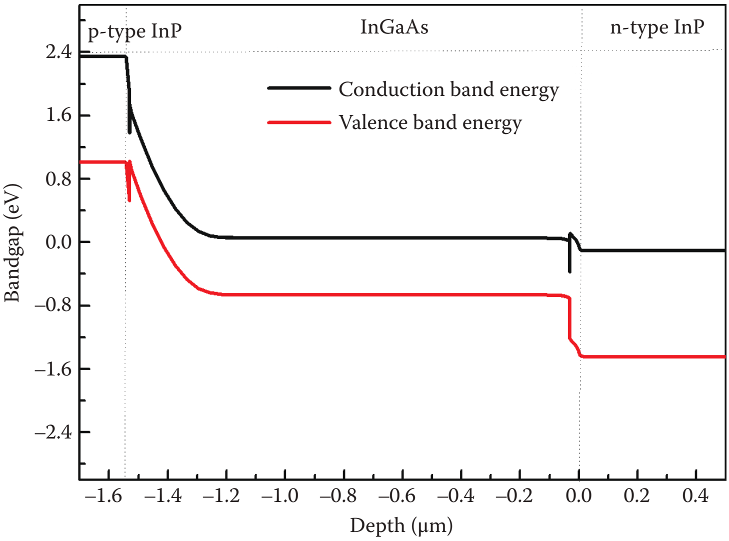


FIGURE 36.10 Energy band of the InP/InGaAs PIN photodiode.

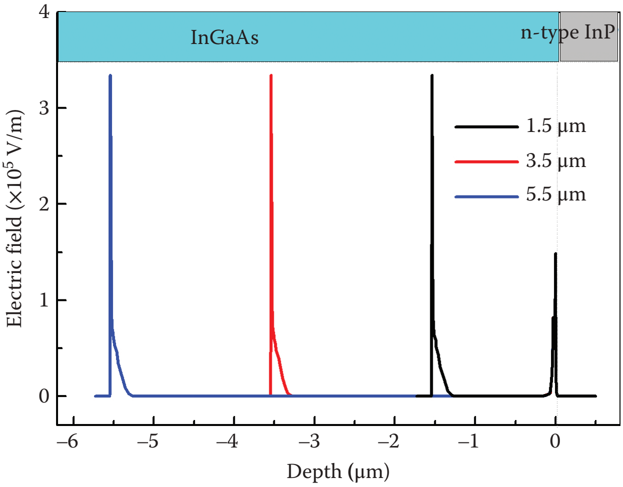


FIGURE 36.11 Vertical distributions of electric field with T_{abs} increasing from 1.5 to 5.5 μm in 2-μm steps.

light current characteristics with the absorption layer (T_{abs}) increasing from 1.5 to 5.5 μm in 1-μm steps are shown in Figure 36.12; the responsivity and QE, as a function of T_{abs} for different incident wavelengths, are shown in Figure 36.13. For clarity, the fitting parameters used in the simulation are listed in Table 36.3.

36.5.2.2 HgCdTe-Based P-N Junction Photodiode

Another example is the HgCdTe-based p-n junction photodiode infrared detector (Hu et al., 2008, 2010; Liang et al., 2014; Qiu and Hu, 2015; Yin et al., 2009; Wang et al., 2011a). The thickness of the n-region of the device has been chosen as a key structural parameter. This section discusses the effect of the thickness of the n-region on responsivity and QE of the device. The following simulations are performed at 77 K. Figure 36.14 presents a schematic diagram of the device where the white region represents the depletion region which is a space charge layer of a PN junction with positive charges in the n-type region and negative

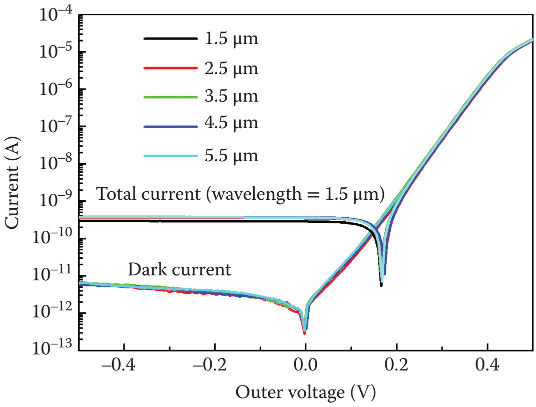


FIGURE 36.12 Dark and light current characteristics with the absorption layer (T_{abs}) increasing from 1.5 to 5.5 µm in 1-µm steps.

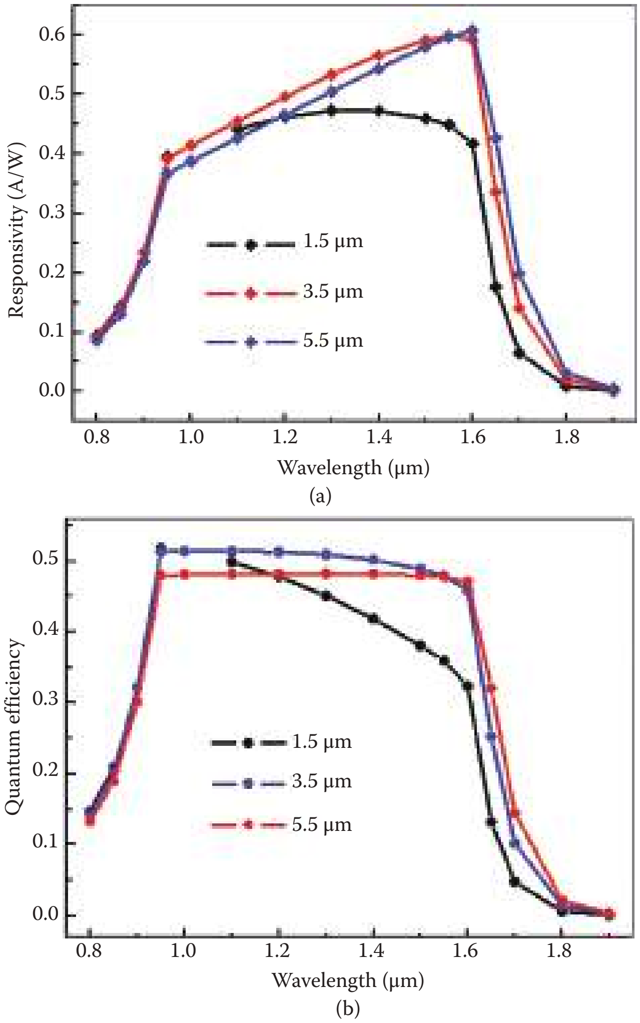


FIGURE 36.13 (a) Responsivity and (b) quantum efficiency for InP/InGaAs/InP PIN photodiode with the absorption layer (T_{abs}) increasing from 1.5 to 5.5 µm in 2-µm steps.

TABLE 36.3 InP/InGaAs/InP PIN Photodiode Simulation Parameter List

Parameters	Units	$In_{0.53}Ga_{0.47}As$	InP
Room-temperature bandgap, E_g	eV	0.78	1.34
Dielectric constant, ε_r	1	13.9	12.4
Density of states in conduction band, N_C	cm^{-3}	2.75×10^{17}	5.66×10^{17}
Density of states in valance band, N_V	cm^{-3}	7.62×10^{18}	2.03×10^{19}
Electron mobility, μ_e	$cm^2/V.{\cdot}s$	12,000	4,730
Hole mobility, μ_h	$cm^2/V.{\cdot}s$	450	151
Electron saturation velocity, v_{Se}	cm/s	8.9×10^6	2.6×10^7
Hole saturation velocity, v_{Sh}	cm/s	8.9×10^6	2.6×10^7
Electron effective mass (relative), m_e	1	0.0489	0.08
Hole effective mass (relative), m_h	1	0.45	0.861
Electron Auger coefficient, R_{Ae}	cm^6/s	3.2×10^{-28}	3.7×10^{-31}
Hole Auger coefficient, R_{Ah}	cm^6/s	3.2×10^{-28}	8.7×10^{-30}
Electron SRH lifetime, τ_{Se}	s	1×10^{-6}	1×10^{-9}
Hole SRH lifetime, τ_{Sh}	s	1×10^{-6}	1×10^{-11}
Radiative coefficient, R_R	cm^3/s	1.43×10^{-10}	2×10^{-11}
Absorption spectra, A_1	cm^{-1}	6687.08	6045.319
Absorption spectra, A_2	cm^{-1}	618.99	4008.099
Absorption spectra, E_1	eV	0.776	1.363
Absorption spectra, E_2	eV	0.012	0.013
Absorption spectra, P	1	1.013	0.65

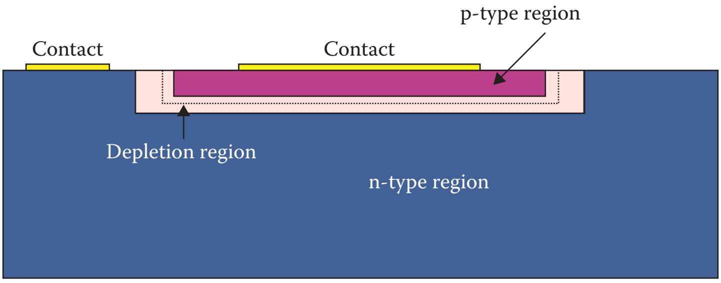


FIGURE 36.14 Schematic cross section of HgCdTe p-n junction photodiode.

charges in the p-type region. Figure 36.15 shows the energy band of the device with different thicknesses (d). Vertical distributions of electric field with the thickness of device (d) increasing from 3 to 7 in 2-μm steps under reverse bias of 200 mV are shown in Figure 36.16. Dark current and photocurrent with varying d from 3 to 7 μmin 2-μm steps under reverse bias are shown in Figure 36.17. The vertical distributions of the electric field of the device with three different conditions are very close. The responsivity and QE, as a function of d, for different incident wavelengths are shown in Figure 36.18. The responsivity and QE of the device with thickness of 5 μm are greater than those devices with thickness of 3 μm, because more photons can be absorbed by the material as the absorber region becomes thicker. However, the responsivity and QE of the device with thickness of 7 μm is smaller than that of the device with the thickness of 5 μm, because the recombination rate of the device with thickness of 7 μm increases, causing that photon-generated carriers to recombine before they arrive at the contacts. For clarity, the fitting parameters used in the simulations are listed in Table 36.4.

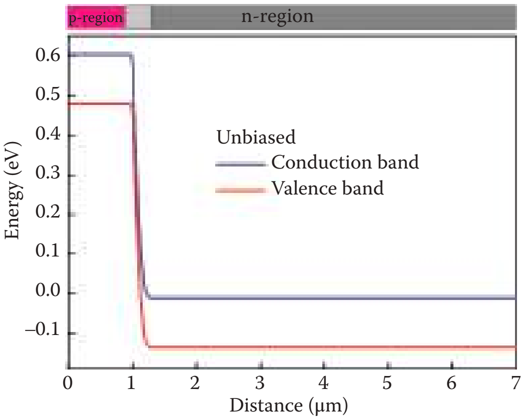


FIGURE 36.15 Energy band of HgCdTe p-n junction photodiode.

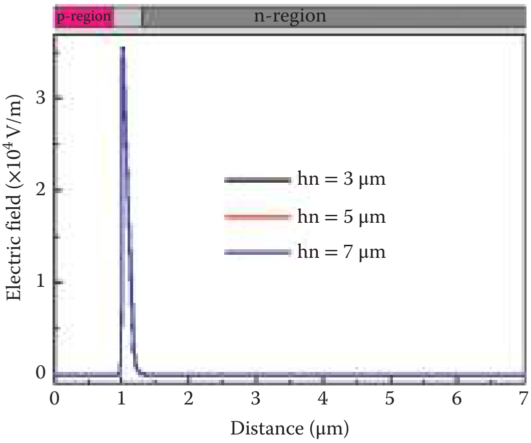


FIGURE 36.16 Vertical distributions of electric field with the height of n-region (hn) increasing from 3 to 7 μm in 2-μm steps.

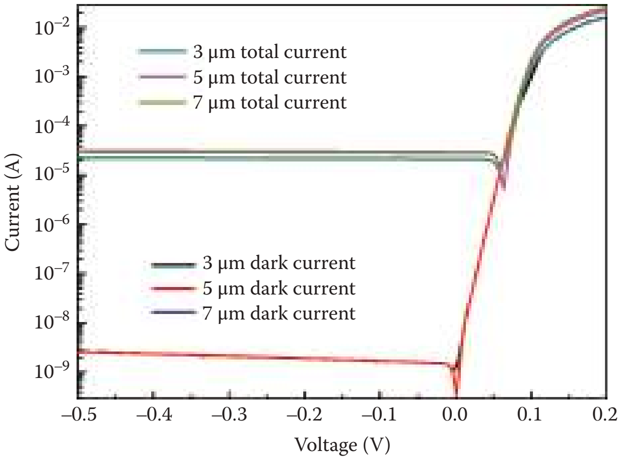


FIGURE 36.17 Dark and light current characteristics with the height of n-region increasing from 3 to 7 μm in 2-μm steps.

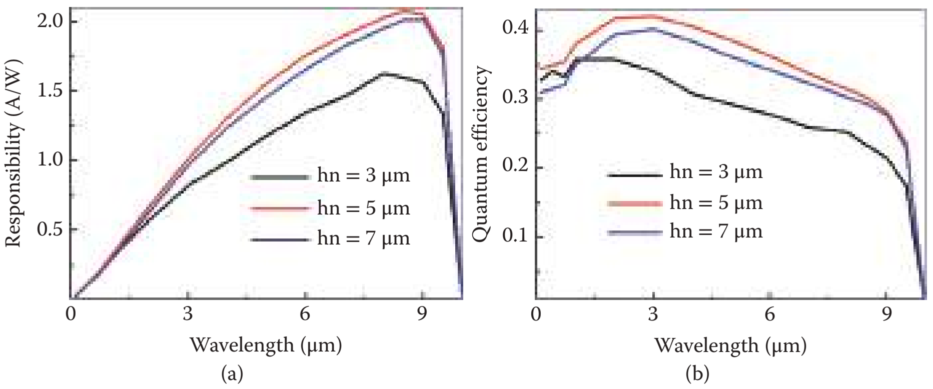


FIGURE 36.18 (a) Responsivity and (b) quantum efficiency for HgCdTe PN photodiode with the height of n-region (hn) increasing from 3 to 7 µm in 2-µm steps.

TABLE 36.4 List of the Key Parameters Used in the Simulation

Parameters	Units	HgCdTe
Bandgap	eV	0.12
Dielectric constant	1	17.4
Effective conduction band density of states	cm^{-3}	0.64×10^{19}
Effective valence band density of states	cm^{-3}	0.70×10^{19}
Electron mobility	$cm^2/V.{\cdot}s$	6,1367
Hole mobility	$cm^2/V.{\cdot}s$	610
Temperature exponent	1	0
Electron effective mass (relative)	1	1.09
Hole effective mass (relative)	1	1.15
Electron Auger coefficient	cm^6/s	2.3×10^{-25}
Hole Auger coefficient	cm^6/s	4.66×10^{-26}
Electron SRH lifetime	s	1×10^{-7}
Hole SRH lifetime	s	1×10^{-7}
Radiative recombination coefficient	cm^3/s	3.47×10^{-17}

36.5.3 Visible Photodiodes

36.5.3.1 Si P-N Junction Photodiode

In the Si p-n junction photodiode, the influence of p-region thickness, operating temperature, and doping is calculated. Figures 36.19 and 36.20 show a schematic diagram and the energy band of a silicon p-n junction device structure, respectively. Vertical distributions of electric field with varying doping concentration are investigated as shown in Figure 36.21. Figure 36.22 shows the light and dark current distribution as a function of different operating temperatures. The responsivity and QE as a function of incident wavelengths for different p-region absorption layer are shown in Figure 36.23. Parameters not described are all according to those given as (default) in parameter Table 36.5.

36.5.4 Ultraviolet Photodiodes

36.5.4.1 GaN/AlGaN PIN Photodiode

The structure of a visible blind GaN/AlGaN PIN photodiode (Wang et al., 2011b, 2014) consists of three epitaxial layers grown on a 2-inch-diameter sapphire substrate. The first layer is a 0.6-µm n^+-doped AlGaN

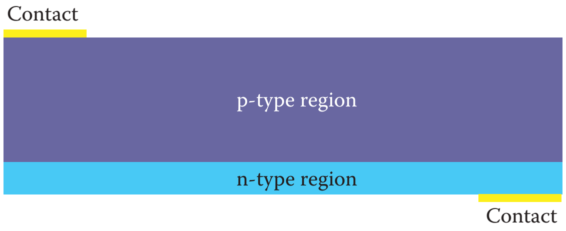


FIGURE 36.19 Schematic cross section of Si p-n junction photodetector.

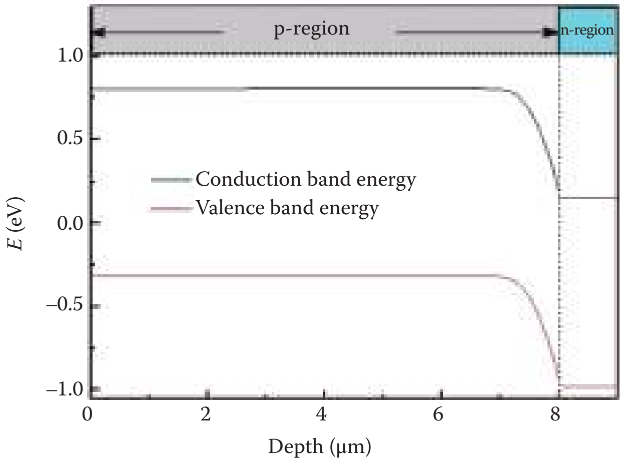


FIGURE 36.20 Energy band of Si p-n junction photodetector.

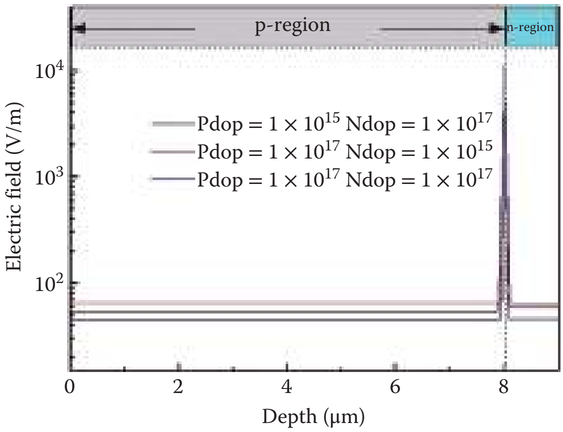


FIGURE 36.21 Vertical distributions of electric field with temperature increasing from 280 to 320 K in 20-K steps.

layer (n-layer) with a doping concentration of 3×10^{18} cm^{-3}. This layer is followed by a 0.1-μm unintentionally doped GaN absorption layer with an electron concentration of 5×10^{15} cm^{-3}, and a 0.15 μm p^{+}-doped GaN layer (p-layer) with a doping concentration of 3×10^{17} cm^{-3}. For simplicity, the thicknesses of the absorption layer and n-layer are defined as d_{abs} and d_{n}, respectively. The schematic cross section of

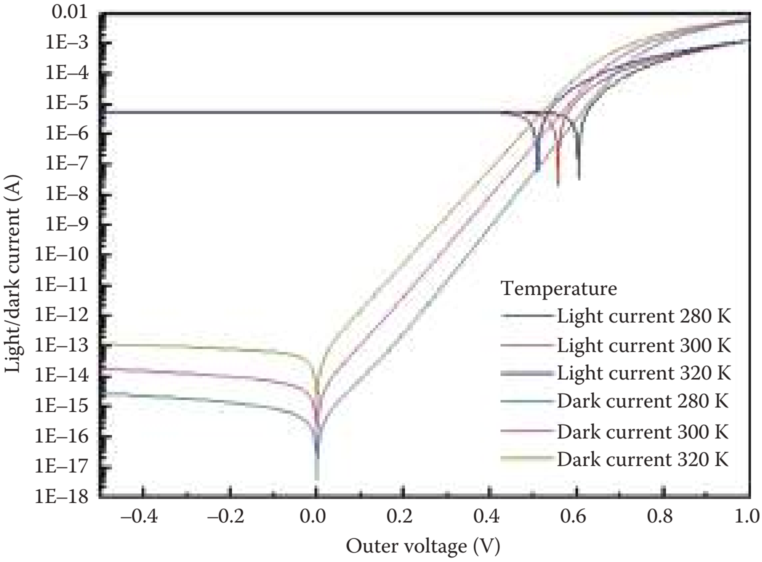
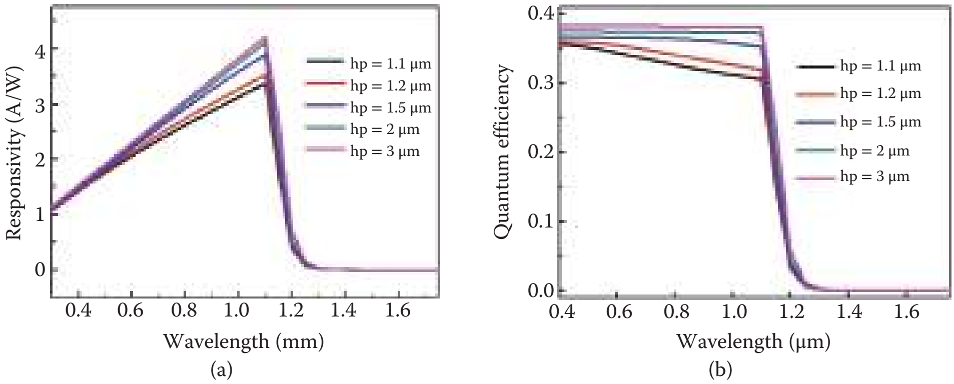


FIGURE 36.22 Dark and light current characteristics with temperature increasing from 280 to 320 K in 20-K steps.



(a)

(b)

FIGURE 36.23 (a) Responsivity and (b) quantum efficiency for Si PN photodiode with the height of p-region (hp) under 1.1, 1.2, 1.5, 2, and 3 μm.

GaN/AlGaN PIN photodiode is shown in Figure 36.24. Figure 36.25 shows energy band as a function of vertical distance at equilibrium state, and the d_{abs} is set as 0.1 μm. Figure 36.26 describes responsivity versus the absorption layers (d_{abs}) with incident wavelengths of 0.30, 0.32, 0.34, and 0.36 μm, respectively, and electric field profile in the vertical direction with the absorption layers (d_{abs}) under 0.05, 0.2, 0.4, 0.6, and 0.8 μm.

The dark and light current characteristics with a bias voltage are shown in Figure 36.27. The absorption layer (d_{abs}) is set as 0.2 μm. Figure 36.28 shows responsivity and QE for GaN/AlGaN/GaN PIN photodiode with the absorption layer (d_{abs}) under 0.2, 0.3, 0.4, 0.6, and 1.1 μm. For clarity, the fitting parameters used in the simulation are listed in Table 36.6.

TABLE 36.5 List of the Key Parameters Used in the Simulation

Parameters	Units	Silicon
Bandgap	eV	1.124
Temperature (default)	K	300
Dielectric constant		11.7
p-type doping level (default)	cm^{-3}	10^{15}
n-type doping level (default)	cm^{-3}	10^{17}
p-type layer thickness (default)	μm	8
n-type layer thickness	μm	1
Lattice heat capacity	$J/K.\cdot cm^3$	1.63
Electron affinity	eV	4.05
Electron's lifetime	s	10^{-5}
Hole's lifetime	s	10^{-6}
Electron effective mass		1.09
Hole effective mass		1.15
Electron mobility	$cm^2/V.\cdot s$	1.42×10^3
Hole mobility	$cm^2/V.\cdot s$	4.70×10^2

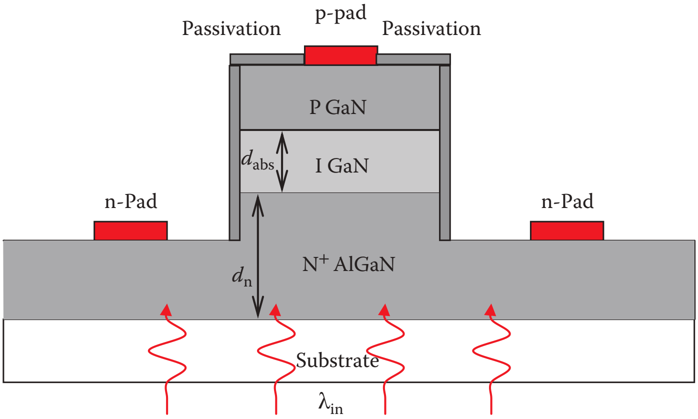

Passivation
p-pad
Passivation
P GaN
d_{abs}
I GaN
n-Pad
n-Pad
N^+ AlGaN
d_n
Substrate
λ_{in}

FIGURE 36.24 Schematic cross section of GaN/AlGaN/GaN PIN photodiode.

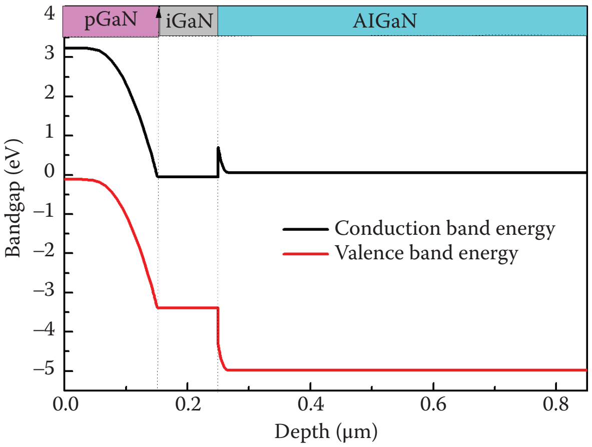
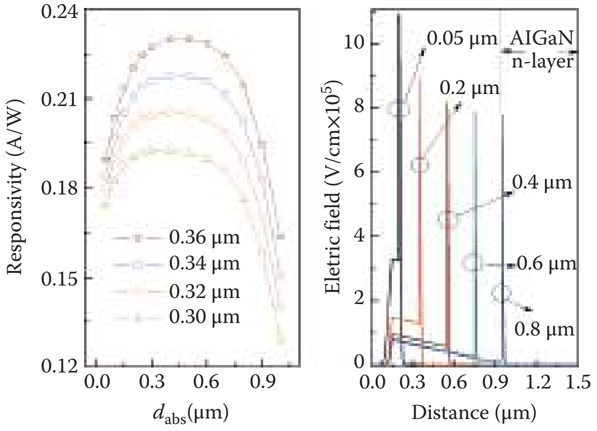


FIGURE 36.25 Energy band of GaN/AlGaN/GaN PIN photodiode.



(a) (b)

FIGURE 36.26 (a) Responsivity versus the absorption layers (d_{abs}) with incident wavelengths of 0.30, 0.32, 0.34, and 0.36 μm, respectively and (b) electric field profile in the vertical direction with the absorption layers (d_{abs}) changing from 0.05 to 0.8 μm.

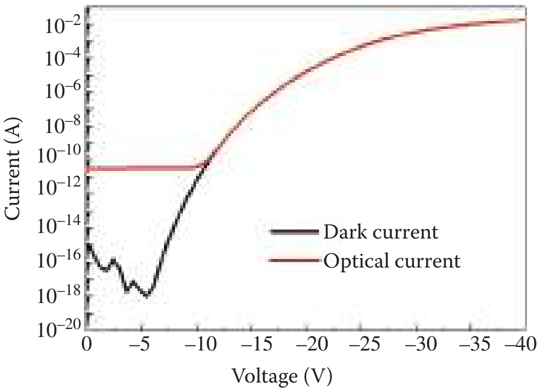


FIGURE 36.27 Dark and light current characteristics of GaN/AlGaN/GaN PIN photodiode.

36.5.5 Two-Color Heterojunction Photodiodes

36.5.5.1 Long-Wavelength/Mid-Wavelength Two-Color HgCdTe

There are two important parameters for the long-wavelength/mid-wavelength (LW/MW) two-color HgCdTe infrared focal plane array detector (Hu et al., 2009, 2010, 2011, 2012, 2013, 2014; Qiu et al., 2016): (1) absorption layer thickness and (2) opening dimensions of the groove (the length of the groove top, L_{top}; the length of the groove bottom, L_{bot}). In the simulation, the electrical cross-talk of both LW to MW ($C_{\lambda e\text{-LW-to-MW}}$) and MW to LW ($C_{\lambda e\text{-MW-to-LW}}$) contributes to the total spectral cross-talk dependent on thickness and band energy offset of the barrier layer (as shown in Figure 36.29). Figure 36.30 shows the spectral photoresponse for the HgCdTe two-color infrared detector with the proposed structure from the previous numerical simulations. Figure 36.31 shows the QE and cross-talk as a function of the thickness of the MW layer for the grooved n_1^+-p_1-P-p_2-n_2^+ HgCdTe two-color infrared detector. It is found that a minimum MW-to-LW cross-talk and maximum MW QE can be achieved with an MW layer of approximately 7 μm. To further optimize the groove structure, the QEs with different L_{top} and L_{bot} are calculated at cutoff wavelengths. It indicates that the MW QE at the MW cutoff wavelength of 4.8 μm increases with L_{bot} (no

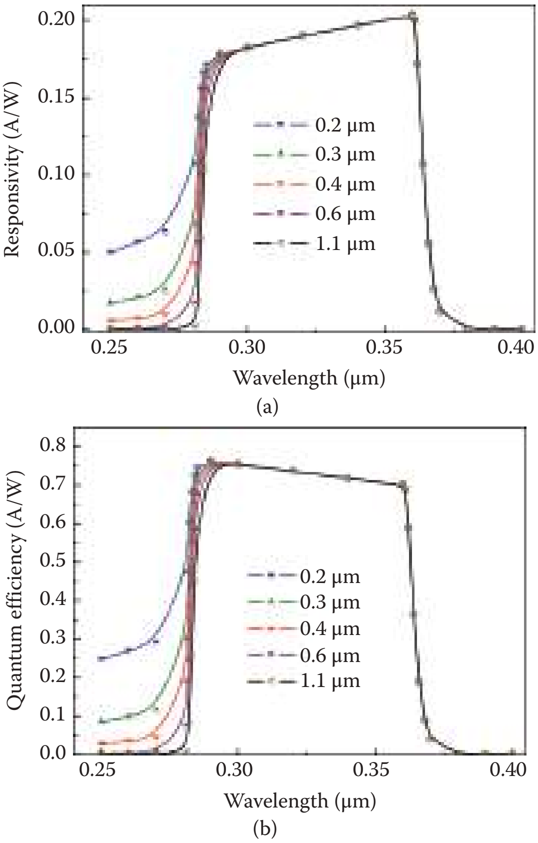


FIGURE 36.28 (a) Responsivity and (b) Quantum efficiency for GaN/AlGaN/GaN PIN photodiode with the absorption layer (Tabs) under 0.2,0.3,0.4,0.6, and 1.1 μm.

TABLE 36.6 List of the Key Parameters Used in the Simulation

Parameters	Units	$Al_{0.3}Ga_{0.7}N$	GaN
Band gap	eV	4.29	3.47
Dielectric constant	1	9.2	9.5
Effective conduction band density of states	cm^{-3}	3.09×10^{18}	2.65×10^{18}
Effective valence band density of states	cm^{-3}	1.03×10^{20}	2.5×10^{19}
Electron mobility	$cm^2/V.{\cdot}s$	800	1010
Hole mobility	$cm^2/V.{\cdot}s$	18	20
Saturation velocity	cm/s	1.5×10^{17}	1.5×10^{17}
Temperature exponent	1	0	0
Electron effective mass (relative)	1	0.25	0.222
Hole effective mass (relative)	1	2.2	1
Electron Auger coefficient	cm^6/s	2.1×10^{-29}	3.2×10^{-29}
Hole Auger coefficient	cm^6/s	6.5×10^{-30}	8.7×10^{-30}
Electron SRH lifetime	s	1×10^{-10}	1×10^{-9}
Hole SRH lifetime	s	1×10^{-10}	1×10^{-9}
Radiative recombination coefficient	cm^3/s	2×10^{-10}	2×10^{-10}

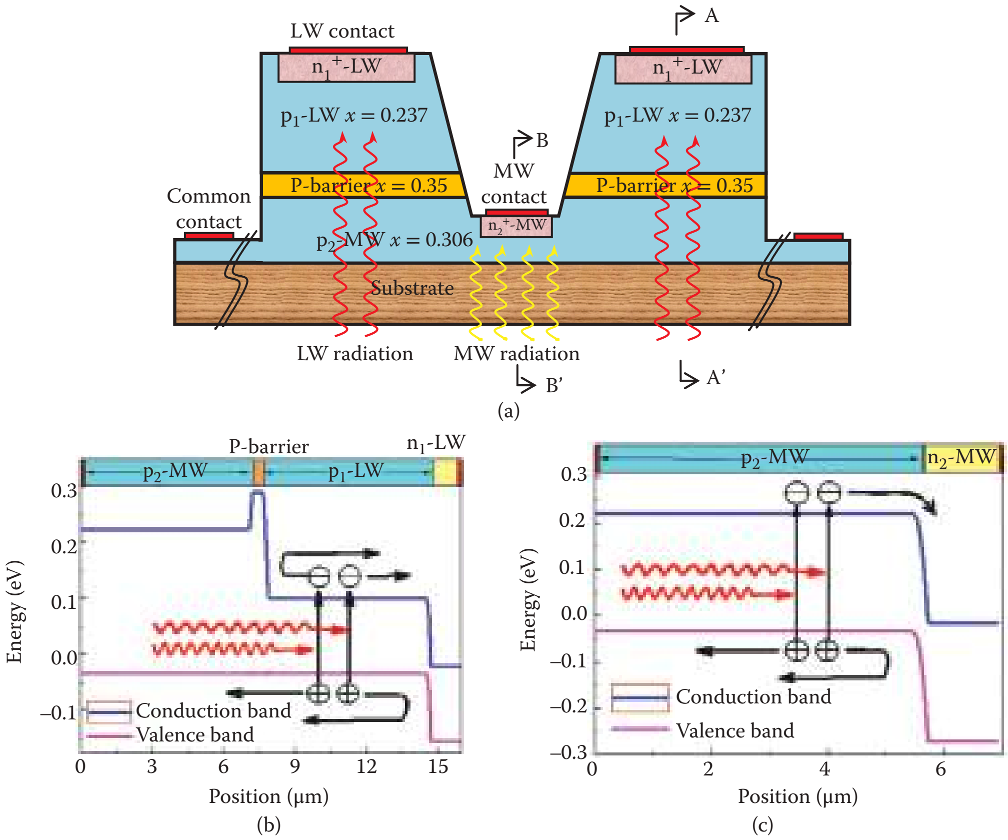

LW contact
n_1^+-LW
p_1-LW $x = 0.237$
P-barrier $x = 0.35$
p_2-MW $x = 0.306$
Common contact
MW contact
n_2^+-MW
Substrate
LW radiation
MW radiation
A
A'
B
B'
(a)
P-barrier
n_1-LW
p_2-MW
p_1-LW
Energy (eV)
0.3
0.2
0.1
0.0
–0.1
Conduction band
Valence band
0
3
6
9
12
15
Position (μm)
(b)
p_2-MW
n_2-MW
Energy (eV)
0.3
0.2
0.1
0.0
–0.1
–0.2
–0.3
Conduction band
Valence band
0
2
4
6
Position (μm)
(c)

FIGURE 36.29 (a) Schematic of grooved HgCdTe two-color infrared detector, (b) equilibrium energy band diagram cut at A-A′, and (c) equilibrium energy band diagram cut at B-B′.

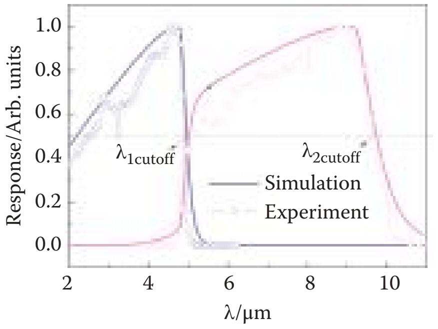


FIGURE 36.30 Experimental (dotted) and simulated (solid) spectral photoresponse of LW/MW HgCdTe two-color infrared detector with cutoff wavelengths of $\lambda_{1cutoff} = 4.8$ μm and $\lambda_{2cutoff} = 9.7$ μm for the LW and MW diodes respectively. A voltage of 0.01 V is used in simulation and experiment.

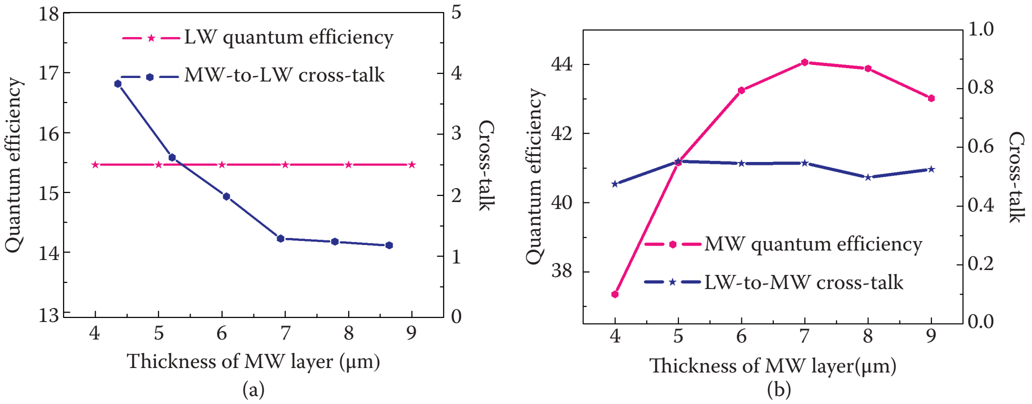


FIGURE 36.31 (a) LW quantum efficiency (QE) and MW-to-LW cross-talk and (b) MW QE and LW-to-MW cross-talk as a function of the thickness of MW layer for the infrared detector. The QE is calculated at cutoff wavelength.

TABLE 36.7 Material and Structural Parameters of the Two-color HgCdTe Photodiode at 77 K

Parameters	Value of LW (Units)	Value of MW (Units)	Value of Barrier (Units)
Cd molar fraction (*x*)	0.232	0.309	0.35
Doping density of n-region	1×10^{17} (cm^{-3})	1×10^{17} (cm^{-3})	–
Doping density of p-region	8×10^{15} (cm^{-3})	8×10^{15} (cm^{-3})	8×10^{15} (cm^{-3})
Thickness of n-region	1.2 (μm)	1.2 (μm)	–
Thickness of p-region	9 (μm)	5.8 (μm)	0.8 (μm)

dependence on L_{top}). When $L_{\text{bot}} \geq 14$ μm, the MW QE saturates at a maximum value as the photogenerated carriers with the diffuse length of ~10 μm are efficiently collected by the MW contact. The specific details of the detector are listed in Table 36.7.

36.6 Summary

With the rapid development of computational capabilities and algorithm, numerical simulation becomes a very important tool in both qualitative and quantitative studies of p-n junction photodiodes. In this chapter, physical equations, models, and parameters that play important roles in the numerical simulation have been introduced. First, for the construction of basic theoretical framework, the well-known Poisson and continuity equations are introduced. Meanwhile, the widespread used drift-diffusion model and thermodynamic model is also described. The results show the effects of the important device parameters such as structure size, doping concentration, and incident wavelength on the device performance such as dark current, photoresponsivity, QE, and so on. Second, we introduced and compared the characteristics of several common optical generation models in detail, which could achieve optoelectronic simulations of different precision and structural requirements. Third, the characteristic input parameters and widely used models in the simulations of typical p-n junction photodiodes are discussed and introduced. Different input parameters and models can be selected according to different optoelectronic materials, device structure, and microscopic physical mechanism. Finally, we introduced some key parameters and measures of diode simulation, and simulation examples of some typical p-n junction photodiodes are given. In addition, some important results are numerically extracted, and the parameter tables used in the simulations are provided.

References

Feit MD and Fleck JA (1978) Light propagation in graded-index optical fibers, *Applied Optics*, 17: 3990–3998.

Hu WD, Chen XS, Ye ZH, Chen YG, Yin F, Zhang B, and Lu W (2012) Polarity inversion and coupling of laser beam induced current in As-doped long-wavelength HgCdTe infrared detector pixel arrays: Experiment and simulation, *Applied Physics Letters*, 101: 181108.

Hu WD, Chen XS, Ye ZH, Feng AL, Yin F, Zhang B, Liao L, and Lu W (2013) Dependence of ion-implant-induced LBIC novel characteristic on excitation intensity for long-wavelength HgCdTe-based photovoltaic infrared detector pixel arrays, *IEEE Journal of Selected Topics in Quantum Electronics*, 19: 4100107.

Hu WD, Chen XS, Ye ZH, Lin C, Yin F, Zhang J, Li ZF, and Lu W (2010a) Accurate simulation of temperature-dependence of dark current in HgCdTe infrared detector assisted by analytical modeling, *Journal of Electronic Materials*, 39: 981–985.

Hu WD, Chen XS, Ye ZH, and Lu W (2010b) An improvement on short-wavelength photoresponse for heterostructure HgCdTe two-color infrared detector, *Semiconductor Science and Technology*, 25: 045028.

Hu WD, Chen XS, Ye ZH, and Lu W (2011) A hybrid surface passivation on HgCdTe long wave infrared detector with in-situ CdTe deposition and high-density hydrogen plasma modification, *Applied Physics Letters*, 99: 091101.

Hu WD, Chen XS, Yin F, Ye ZH, Lin C, Hu XN, Li ZF, and Lu W (2009) Numerical analysis of two-color HgCdTe infrared photovoltaic heterostructure detector, *Optical and Quantum Electronics*, 41: 699–704.

Hu WD, Chen XS, Yin F, Ye ZH, Lin C, Hu XN, Quan ZJ, Li ZF, and Lu W (2008) Simulation and design consideration of photoresponse for HgCdTe infrared photodiodes, *Optical and Quantum Electronics*, 40: 1255–1260.

Hu WD, Ye ZH, Liao L, Chen HL, Chen L, Ding RJ, He L, Chen XS, and Lu W (2014) A 128 × 128 long-wavelength/mid-wavelength two-color HgCdTe infrared focal plane array detector with ultra-low spectral crosstalk, *Optics Letters*, 39: 5130–5133.

Liang J, Hu WD, Ye ZH, Liao L, Li ZF, Chen XS, and Lu W (2014) Improved performance of HgCdTe infrared detector focal plane arrays by modulating light field based on photonic crystal structure, *Journal of Applied Physics*, 115:184504.

Qiu WC and Hu WD (2015) Laser beam induced current microscopy and photocurrent mapping for junction characterization of infrared photodetectors, *Science China-Physics Mechanics & Astronomy*, 58 (2): 1–13.

Qiu WC, Hu WD, Lin C, Chen XS, and Lu W (2016) Surface leakage current in 12.5 μm long-wavelength HgCdTe infrared photodiode arrays, *Optics Letters*, 41 (4): 828–831.

Rajkanan K, Singh R, and Shewchun J (1979) Absorption coefficient of silicon for solar cell calculations, *Solid-State Electronics*, 22: 793–795.

Shi M and Li ML (2014) *Semiconductor Devices Physics and Technology*, Soochow, China: Soochow University Press.

Wang XD, Chen XY, Hou LW, Wang BB, Xie W, and Pan M (2015) Role of n-type AlGaN layer in photoresponse mechanism for separate absorption and multiplication (SAM) GaN/AlGaN avalanche photodiode, *Optical and Quantum Electronics*, 47: 1357–1365.

Wang J, Chen XS, Hu WD, Wang L, Lu W, Xu FQ, Zhao J, Shi YL, and Ji RB (2011a) Amorphous HgCdTe infrared photoconductive detector with high detectivity above 200 K, *Applied Physics Letters*, 99: 113508.

Wang XD, Hu WD, Chen XS, Lu W, Tang HJ, Li T, and Gong HM (2008) Dark current simulation of InP/In0.53Ga0.47As/InP p-i-n photodiode, *Optical and Quantum Electronics*, 40: 1261–1266.

Wang XD, Hu WD, Chen XS, Xu JT, Li XY, and Lu W (2011b) Photoresponse study of visible blind GaN/AlGaN p-i-n ultraviolet photodetector, *Optical and Quantum Electronics*, 42: 755–764.

Wang XD, Hu WD, Pan M, Hou LW, Xie W, Xiu JT, Li XY, Chen XS, and Lu W (2014) Study of gain and photoresponse characteristics for back-illuminated separate absorption and multiplication GaN avalanche photodiodes, *Journal of Applied Physics*, 115:013103.

Yin F, Hu WD, Zhang B, Li ZF, Hu XN, Chen XS, and Lu W (2009) Simulation of laser beam induced current for HgCdTe photodiodes with leakage current, *Optical and Quantum Electronics*, 41: 805–810.

37

Quantum Well Infrared Photodetectors

37.1 Introduction 337

37.2 Theoretical Model of QWIPs 338

Dark Current • Photoconductive Gain • Photocurrent • Absorption Coefficient

37.3 Material Absorption Modeling 343

Material Parameters • Numerical Examples

37.4 Detector QE Modeling 347

Modeling Validation and Phase Coherence • Planar Solar Cell • InAs/GaInSb IR Photodiode • Edge Coupling for QWIPs • Quantum Grid Infrared Photodetector • Plasmonic-Enhanced QWIP • Photonic Crystal Slab QWIP

37.5 EM Design of Resonator-QWIPs 356

Resonator-QWIP Design • R-QWIP Realization

37.6 Conclusion 361

Kwong-Kit Choi

37.1 Introduction

Quantum well infrared photodetectors (QWIPs) are based on optical transitions among quantized states in the conduction band of the quantum wells (QWs). They are widely used in long-wavelength infrared (LWIR) detection such as in gas sensing (Hinnrichs and Guptab, 2008) and in Earth observation from space (Jhabvala et al., 2011). The popularity of QWIPs is gained from the advantages in material quality and uniformity, detector sensitivity and stability, and array availability and cost. Another unique advantage of this technology is its predictable optoelectronic properties. In the absence of the extrinsic effects imposed by impurities and surfaces, the observed photocurrent and dark current of a QWIP are intrinsic to the detector structure and are well understood from fundamental physics. More specifically, its dark current is found to agree with the familiar thermionic emission (TE) process, while the photocurrent can be quantitatively predicted from first principle calculations on QW energy levels and electromagnetic (EM) field distributions inside the detector volume. With the known dark and photocurrents, the detection characteristics and performance of a QWIP array placed in the focal plane of an optical lens system can readily be predicted with very few system parameters. But because of the arbitrariness of the QW materials and the detector geometries, the associated physical equations can only be solved by numerical means. In this chapter, we describe the theoretical framework and numerical modeling of QWIPs.

The role of photocurrent I_p and dark current I_d in affecting the sensitivity of a QWIP focal plane array (FPA) has been analyzed (Choi et al., 1998). Here, it suffices to note that while the signal S of a detector is directly proportional to I_p, the noise N is directly proportional to $(I_p + I_d)^{1/2}$, which can be derived from the Poisson statistics of charge collection. The signal-to-noise ratio of a QWIP is therefore proportional to

$$\left(\frac{S}{N}\right)_{\text{detector}} \propto \frac{I_{\text{p}}}{\sqrt{I_{\text{p}} + I_{\text{d}}}} = \left(\frac{I_{\text{p}}}{1 + I_{\text{d}}/I_{\text{p}}}\right)^{1/2} = \left(\frac{I_{\text{p}}}{1 + 1/r}\right)^{1/2}, \tag{37.1}$$

where r is the photocurrent to dark current ratio ($I_{\text{p}}/I_{\text{d}}$). From Equation 37.1, it is obvious that the S/N ratio is large only when I_{p} and r are large. In this chapter, we focus on the modeling of I_{p} and I_{d} in QWIP design and optimization.

37.2 Theoretical Model of QWIPs

A QWIP uses a multiple quantum well (MQW) structure to detect light (Levine et al., 1987). This structure is realized by growing alternate material layers of different bandgaps on a suitable substrate. The MQW creates a series of energy subbands, $E_1, E_2, \ldots$, in each QW unit. The QWs are doped n-type with a doping density N_{D} so that there are free electrons in the ground subband E_1 with a Fermi energy E_{F}. The band structure of a typical QWIP made of $\text{Al}_x\text{Ga}_{1-x}\text{As}/\text{In}_y\text{Ga}_{1-y}\text{As}$ is shown in Figure 37.1. In Figure 37.1, x is the aluminum mole fraction, y is the indium mole fraction, W is the well width, and B is the barrier thickness. An inter-subband absorption brings an electron from E_1 to E_n where the conductivity across the layers is higher. A photocurrent is thus created.

37.2.1 Dark Current

The dark current I_{d} of a QWIP can be qualitatively separated into three main components. They are direct tunneling (DI) between adjacent ground states, thermally assisted tunneling (TAT) near the top of the barrier, and TE over the barrier. These processes are indicated in Figure 37.2.

The TAT current originates from the finite thermal spreading of in-plane electron energy in E_1. Upon a scattering process, the in-plane energy combines with the out-of-plane energy, with which the electron tunnels more efficiently through the QW barrier. Therefore, the barrier transmission depends on the total energy rather than its energy components. Although each current component differs in details, such as the densities of initial and final states, the major factor in determining their magnitudes is the impedance of the QW barrier. By adopting an energy-dependent tunneling probability $\gamma(E, V)$ for the barrier, the total

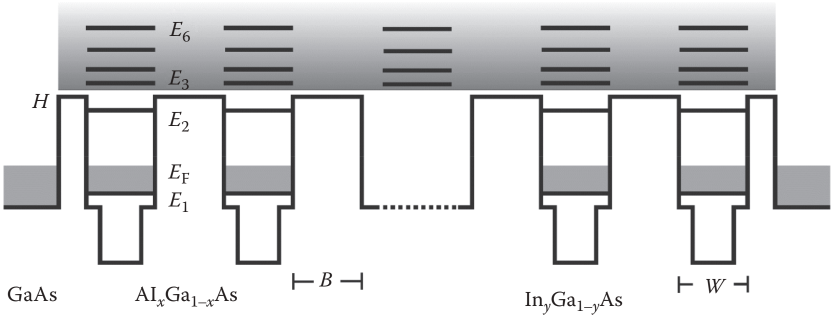

E_6
E_3
H
E_2
E_{F}
E_1
GaAs
$\text{Al}_x\text{Ga}_{1-x}\text{As}$
B
$\text{In}_y\text{Ga}_{1-y}\text{As}$
W

FIGURE 37.1 A typical energy diagram of an n-type QWIP along the z axis. E_1 to E_6 are the resonant QW states. The shaded region is the global conduction band.

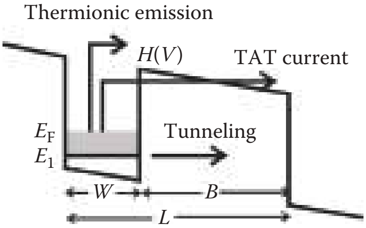


FIGURE 37.2 Different dark current transport processes in a QWIP. (TAT, thermally assisted tunneling.)

dark current can be obtained by summing over all electron energies (Levine et al., 1990):

$$\begin{aligned} J_d(V,T) &= \int_{E_1}^{\infty} e\rho_d(E,T)\, v_d(V)\mathrm{d}E, \\ &= \int_{E_1}^{\infty} e\left[\frac{g_{2D}}{L} f(E)\gamma(E,V)\right] v_d(V)\mathrm{d}E, \\ &= \int_{E_1}^{\infty} e\frac{m^*}{\pi\hbar^2 L}\frac{\gamma(E,V)\, v_d(V)}{1+\exp\left(\frac{E-E_F-E_1}{kT}\right)}\mathrm{d}E. \end{aligned} \tag{37.2}$$

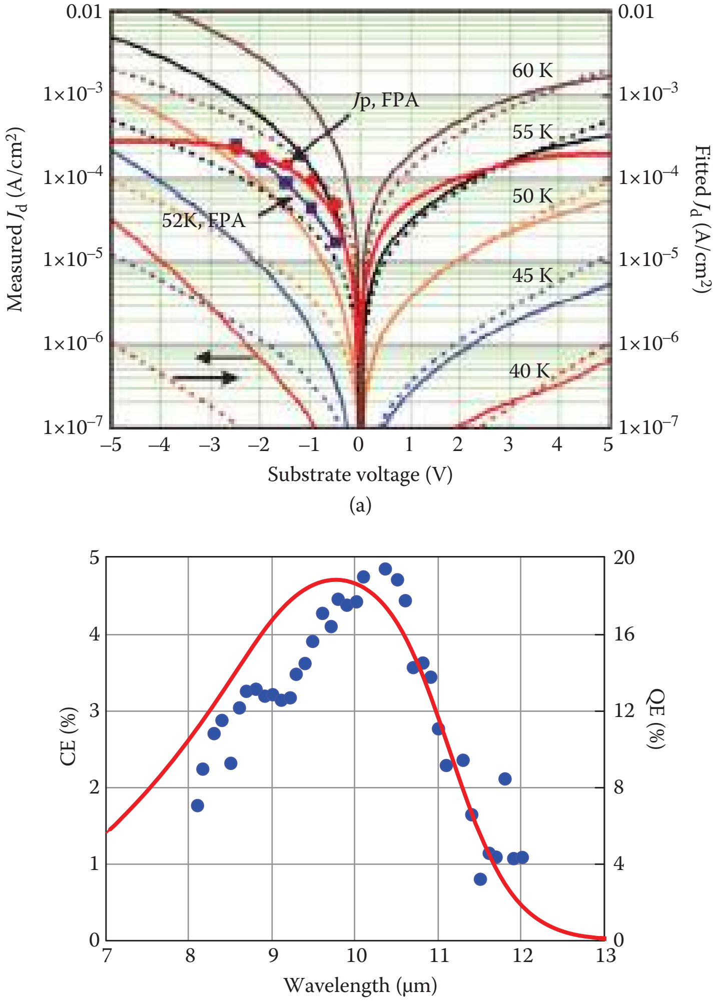
In Equation 37.2, $J_d(V,T)$ is the total dark current density under a bias of V per period at an operating temperature T, E the electron total energy, $\rho_d(E)$ the three-dimensional (3D) electron density per electron-volt, $v_d(V)$ the drift velocity, g_{2D} the 2D density of states, L the QW period length, $f(E)$ the Fermi–Dirac distribution, $\gamma(E,V)$ the tunneling probability of an electron with E under V, and m^* the electron effective mass in the well. The value of γ can be obtained from the usual Wentzel–Kramers–Brillouin (WKB) approximation (Choi, 1997) of a trapezoidal barrier with a barrier height H, a barrier thickness B, and an electron effective mass $m_b{}^*$ in the barrier. Specifically,

$$\begin{aligned} \gamma(E,V) &= \exp\left\{\frac{-4B}{3e\hbar V}(2m_b^*)\left[(H(V)-E)^{3/2}-(H(V)-eV-E)^{3/2}\right]\right\} \quad \text{for } E < H(V)-\text{eV}, \\ &= \exp\left\{\frac{-4B}{3e\hbar V}(2m_b^*)\,(H(V)-E)^{3/2}\right\} \quad \text{for } H(V) > E > H(V)-\text{eV}, \\ &= 1 \text{ for } E > H(V), \end{aligned} \tag{37.3}$$

where H is generally a function of V because there is a potential drop across the well as depicted in Figure 37.2, which depends on the W/L ratio. The drift velocity was experimentally found to be linearly proportional to the electric field (V/L) with a low-field mobility μ initially and then becomes saturated at a constant velocity v_{sat} when the field is high. Therefore, $v_d(V)$ can be represented as follows (Levine et al., 1990):

$$v_d(V) = \frac{\mu V/L}{\sqrt{1+\left(\mu V/Lv_{sat}\right)^2}}. \tag{37.4}$$

To fit the measured J_d, one can adjust H, μ, and v_{sat} in accounting for uncertainties in the material composition and quality.



(a)

(b)

FIGURE 37.3 (a) The measured J_d (solid curves), fitted J_d (dashed curves), and measured J_p (thick solid curves) of a test detector are shown. The symbols are measured from a focal plane array (FPA). (b) The predicted and measured quantum efficiency (QE) and conversion efficiency (CE) of the FPA at −3 V based on angled sidewall optical coupling are shown.

Figure 37.3 shows the typical current–voltage (I–V) characteristics of a QWIP. The substrate voltage refers to the bias applied on the substrate while the top contact is grounded. They are measured from a QWIP with a cutoff wavelength $\lambda_c = 11.2$ μm (Choi et al., 2011). The QWIP is made up of 60 periods of 700 Å $Al_{0.19}Ga_{0.81}As$ and 50 Å GaAs. The wells are doped with N_D of 1.7×10^{18} cm^{-3}. By adopting a constant $H = 191$ meV, $\mu = 400$ cm^2/V · s, and $v_{sat} = 1 \times 10^7$ cm/s, the dark current I–V characteristics under positive bias can be explained in a large range of temperatures. In this example, H is independent of V because the W/L ratio ($= 50/750 = 0.067$) is very small. The adopted H is consistent with the expected band offset from the barrier composition, and the values of μ and v_{sat} are nearly the same for all experimental LWIR detectors. This consistency shows that Equation 37.2 is a satisfactory dark current model for QWIPs. According to the symmetrical band structure shown in Figure 37.1, the I–V characteristics of a QWIP should be symmetric with respect to bias polarity. The higher measured dark current shown in Figure 37.3

under the negative bias is attributed to dopant migration during the material growth, which lowers the barrier height in the barrier region immediately after the doping.

37.2.2 Photoconductive Gain

The photocurrent I_p of a QWIP can be expressed as $I_\mathrm{p} = A\rho_\mathrm{p}ev_\mathrm{d}$, where A is the detector area, ρ_p the photoelectron density, and v_d the electron drift velocity. The parameter ρ_p can be further expressed as $G\tau/(A\ell)$, where G is the total photoelectron generation rate in the entire detector volume $A\ell$, ℓ the detector total thickness, and τ the average recombination lifetime of a photoelectron. After photoexcitation, a photoelectron only retains its energy in a period τ before it recombines into one of the QWs. Therefore, $G\tau/(A\ell)$ represents the time averaged photoelectron density. If we define the quantum efficiency (QE) η as the fraction of incident photon flux Φ being absorbed by the detector and converted into photoelectrons, the generation rate G will be $\eta\Phi$. Hence, I_p is given by

$$\begin{aligned} I_\mathrm{p} &= A\rho_\mathrm{p}ev_\mathrm{d}, \\ &= A\left(\frac{G\tau}{A\ell}\right)e\left(\frac{\ell}{\tau_\mathrm{tr}}\right) = \left(\frac{\eta\Phi\tau}{\ell}\right)e\left(\frac{\ell}{\tau_\mathrm{tr}}\right) = \eta\Phi\left(\frac{\tau}{\tau_\mathrm{tr}}\right)e, \\ &= \eta\Phi g e, \end{aligned} \tag{37.5}$$

where τ_tr is the time taken by a photoelectron to travel across the entire detector thickness ℓ, and $g = \tau/\tau_\mathrm{tr} = \tau v_\mathrm{d}/\ell$ is defined as the photoconductive gain. Equation 37.5 indicates that the measured number of photoelectrons (I_p/e) is g times the number of photons absorbed ($\eta\Phi$), resulted in a gain. Since the lifetime τ of a photoelectron is usually shorter than the transit time τ_tr in a QWIP, g is usually less than unity. Its photocurrent will then be lower than that of a photodiode with a gain of unity for the same QE. The value of g is usually deduced from noise measurements and theoretically fitted to the same functional form of v_d with τ as a fitting parameter. Experimentally, we found $\tau = 25$ ps for most LWIR QWIPs.

37.2.3 Photocurrent

Expanding the last line of Equation 37.5, the photocurrent is given by

$$\begin{aligned} I_\mathrm{p}(V, T_B) &= eg(V)\int_{\lambda_1}^{\lambda_2} \eta(V,\lambda)\Phi(T_B,\lambda)\mathrm{d}\lambda, \\ &= eg(V)\int_{\lambda_1}^{\lambda_2} \eta(V,\lambda)\Omega A L(T_B,\lambda)\mathrm{d}\lambda, \\ &= \frac{\pi}{4F^2+1}Aeg(V)\int_{\lambda_1}^{\lambda_2} \eta(V,\lambda)L(T_B,\lambda)\mathrm{d}\lambda. \end{aligned} \tag{37.6}$$

In Equation 37.6, Ω is the solid angle of the light cone sustained at the detector toward the lens. $L(T_\mathrm{B}, \lambda)$ is the photon spectral radiance of the scene, which has a unit of number of photons arrived per unit time per unit detector area per unit solid angle per unit wavelength. Expressing Ω in terms of the usual specification for a lens system, the f-number F, which is the ratio between the focal length f and the lens diameter D, $\Omega = \pi \sin^2(\theta) = \pi(D/2)^2/[f^2 + (D/2)^2] = \pi/(4F^2 + 1)$, where θ is the half-cone angle.

The parameter $g(V)$ is the photoconductive gain at V, λ_1 and λ_2 are the lower and upper wavelengths encompassing the detector absorption spectrum, and $\eta(V,\lambda)$ is the detection QE spectrum. $\eta(V,\lambda)$ can be decomposed into $\eta_{\text{abs}}(\lambda)\gamma(V)$, where η_{abs} is the fraction of incident photons being absorbed and $\gamma(V)$ is the tunneling probability of the photoelectron out of the barrier. For the typical QWIP designs, the upper subbands are above the barrier and thus $\gamma(V)$ is close to unity. Therefore, we usually assume $\eta(V,\lambda) \approx \eta_{\text{abs}}(\lambda)$, and label $\eta_{\text{abs}}(\lambda)$ simply as $\eta(\lambda)$ unless the effect of γ is apparent.

37.2.4 Absorption Coefficient

The magnitude of the photocurrent I_{p} in Equation 37.6 depends on the QE spectrum $\eta(\lambda)$. In QWIPs, the optical transitions are generated from the transition dipole moment between two envelope eigenfunctions in the z-direction. The optical electric field E of the radiation must be oriented in this direction to drive the transitions. When the optical field is polarized in the x–y plane, the dipole moment is zero due to conservation of momentum, and therefore, there is no absorption. Therefore, the optical absorption of the material depends on the direction of light propagation, and the absorption coefficient α is anisotropic. The parameter α is defined as $(-1/\Phi)(\mathrm{d}\Phi/\mathrm{d}x)$, where Φ is the photon flux and x is the traveling distance. In general, when the light is traveling at an angle θ relative to the normal of the layer surface, only the vertical electric component $E_z = E \cdot \sin(\theta)$ in the transverse magnetic (TM) mode is able to initiate the transitions. The transverse electric mode, on the other hand, is totally uncoupled. Therefore, for an unpolarized light, only half of its intensity is being absorbed, and α of the remaining half varies as $\sin^2\theta$. When θ is zero under the usual detection orientation, i.e., under normal incidence, α will be zero unless there is an optical coupling structure to convert the incident electric polarization from horizontal to vertical. Overall, $\eta(\lambda)$ of a QWIP depends on both the material intrinsic absorption properties and the detector physical structure.

For a parallel propagating light with vertical polarization, the absorption coefficient $\alpha(\hbar\omega)$ depends on the oscillator strength between the ground state and one of the upper states and the number of free carriers in the ground state. Quantitatively, $\alpha(\hbar\omega)$ is given by (Choi, 1997)

$$\alpha(\hbar\omega) = \frac{N_{\text{s}}}{L}\frac{\pi e^2 \hbar}{2m^* \sqrt{\varepsilon_{\text{r}}}\,\varepsilon_0 c}\sum_{n=2}^{\infty} f_n \rho_n(\hbar\omega + E_1), \tag{37.7}$$

where $N_{\text{s}} = N_{\text{D}}W$ is the 2D doping density per QW period, L the length of a QW period, m^* the effective electron mass of individual layers, ε_{r} the real part of the dielectric constant, f_n the oscillator strength between E_1 and E_n, and ρ_n the normalized density of states due to the presence of line broadening mechanisms. The oscillator strength f_n is given by

$$f_n = \frac{2\hbar^2}{m^*\left(E_n - E_1\right)}\left|\langle\psi_{\text{n}}\left|\frac{\partial}{\partial z}\right|\psi_1\rangle\right|^2, \tag{37.8}$$

where ψ_{n} is the eigenfunction of E_n. The overall absorption lineshape depends mostly on the energy distribution of E_n and the respective f_n with a weaker dependence on the individual ρ_n. We assume a Gaussian shape for ρ_n, i.e.,

$$\rho_n(E) = \frac{1}{\sqrt{2\pi}\sigma_{\text{n}}}\exp\left(-\frac{\left(E - E_n\right)^2}{2\sigma_{\text{n}}^2}\right), \tag{37.9}$$

where σ_{n} is the line broadening parameter. In the next section, we show modeling procedures for $\alpha(\hbar\omega)$.

37.3 Material Absorption Modeling

The eigenfunction ψ_n and eigenvalue E_n of an MQW can be obtained from the transfer matrix method (TMM) (Choi et al., 2002). In this method, ψ_j in the jth material layer is written as two counter-propagating plane wave states:

$$\psi_j = A_j\, e^{i\,k_j\,z} + B_j\, e^{-\,i\,k_j\,z}, \tag{37.10}$$

where A_j and B_j are two constants to be determined, and k_j is the wavevector given by

$$k_j(E) = \frac{\sqrt{2\,m_j^*(E)}}{\hbar}\sqrt{E - \Delta E_j}, \tag{37.11}$$

where $m_j^*(E)$ is the energy-dependent effective mass, E and ΔE_j are the electron energy and the conduction band offset measured from the GaAs conduction band edge, respectively. The constants A_j and B_j relate to A_{j+1} and B_{j+1} of the adjacent $(j+1)$th layer through the boundary conditions that both ψ and $(1/m^*)\mathrm{d}\psi/\mathrm{d}z$ are continuous across the interface:

$$\begin{aligned}\begin{bmatrix}A_j\\ B_j\end{bmatrix} &= \frac{1}{2}\begin{bmatrix}(1+\gamma_{j,j+1})\,e^{i\,(k_{j+1}-k_j)\,d_{j,j+1}} & (1-\gamma_{j,j+1})\,e^{-\,i\,(k_{j+1}+k_j)\,d_{j,j+1}}\\ (1-\gamma_{j,j+1})\,e^{i\,(k_{j+1}+k_j)\,d_{j,j+1}} & (1+\gamma_{j,j+1})\,e^{-\,i\,(k_{j+1}-k_j)\,d_{j,j+1}}\end{bmatrix}\begin{bmatrix}A_{j+1}\\ B_{j+1}\end{bmatrix}\\ &= \frac{1}{2}\,M_{j\,,j+1}\begin{bmatrix}A_{j+1}\\ B_{j+1}\end{bmatrix},\end{aligned} \tag{37.12}$$

where $\gamma_{j,j+1} = (m_j{}^*k_{j+1})/(m_{j+1}{}^*k_j)$ and $d_{j,j+1}$ is the spatial coordinate of the interface. The energy level structure of a multilayer material system can be obtained from the maxima of the global transmission coefficient T_G of a plane wave through the material. When the incident energy of the wave coincides with one of E_n's, strong resonant transmission occurs and T_G attains a local maximum. Assuming the plane wave incident from the left, the coefficients A_1 and B_1 for the first material layer on the left and the coefficients A_p and B_p of the last pth layer are connected by

$$\begin{bmatrix}A_1\\ B_1\end{bmatrix} = \frac{1}{2^{p-1}}\,M_{1,p}\begin{bmatrix}A_p\\ B_p\end{bmatrix}, \tag{37.13}$$

where $M_{1,p} = M_{1,2}M_{2,3}\ldots M_{p-1,p}$, and $M_{j,j+1}$ is defined in Equation 37.12. By assuming that there is no wave traveling to the left in the last period, one can set $A_p = 1$ and $B_p = 0$, with which all the coefficients in the layers are determined. The value of T_G at a given energy E is then equal to

$$T_\mathrm{G}(E) = \frac{1}{\left|A_1(E)\right|^2}\frac{v_p(E)}{v_1(E)} = \frac{2^{2p-2}}{\left|a_{11}(E)\right|^2}\frac{m_1^*(E)}{m_p^*(E)}\frac{k_p(E)}{k_1(E)}, \tag{37.14}$$

where v is the electron group velocity and a_{11} the first diagonal element of $M_{1,p}$. The local maxima in $T_\mathrm{G}(E)$ determine the locations of E_n. In general, these local maxima are less than unity unless the material structure is symmetrical. Calculating $A_\mathrm{n}(E)$ and $B_\mathrm{n}(E)$ at $E = E_n$ will yield ψ_n.

37.3.1 Material Parameters

The TMM calculation can be handled efficiently by a personal computer. The material parameters required for the InGaAs/AlGaAs material system had been collected and summarized (Shi and Goldys, 1999).

To begin with, the unstrained bandgap of the *i*th layer in the unit of meV is

$$E^u_{g,j} = 1424 + 1186\,x_j + 370\,x_j^2 - 1499\,y_j + 429\,y_j^2, \tag{37.15}$$

where x_j and y_j are the Al and In molar ratios in the *j*th layer, respectively. (In the present calculation, x and y both cannot be nonzero in a given layer, i.e., AlInGaAs quaternary compound is not considered.) Since the material layers are under strain, the values of bandgaps are modified. Assuming the strain is accommodated exclusively in the InGaAs well layers, which are usually much thinner than the GaAs/AlGaAs layers, the modified bandgap for $y_j > 0$ is

$$E^s_{g,j} = E^u_{g,j} + 2\,a\left(1 - \frac{c_{12}}{c_{11}}\right)\varepsilon_{11}, \tag{37.16}$$

where $a = -8700 + 2700y_j$ meV is the deformation potential, $c_{12}/c_{11} = (5.32 - 0.79y_j)/(11.81 - 3.48y_j)$ is the ratio of the elastic constants, and $\varepsilon_{11} = -0.405y_j/(5.6533 - 0.405y_j)$ is the hydrostatic component of the strain induced by the adjacent GaAs/AlGaAs material layers. Therefore, in the present assumption, the bandgap of InGaAs is widened while that of GaAs and AlGaAs are unchanged. The band offset ΔE_j is taken as

$$\begin{aligned} \Delta E_j &= \beta_x\,(E^u_{g,j} - 1424) \quad \text{if} \quad x_j > 0 \\ &= \beta_y\,(E^s_{g,j} - 1424) \quad \text{if} \quad y_j > 0, \end{aligned} \tag{37.17}$$

where β_x is the band offset parameter for the GaAs/AlGaAs interface and β_y is for the InGaAs/GaAs interface. From fitting detectors with different wavelengths, we notice that β_x increases from 0.60 to 0.77 when x decreases from 0.4 to 0.16, and β_y is 0.63 when $y < 0.34$. The effective mass at the conduction band edge is given by

$$m_j^* = 0.0665\,(1 + 1.256\,x_j - 0.579\,y_j). \tag{37.18}$$

Taking the band nonparabolicity into account, the effective mass $m_j{}^*(E)$ can be expressed in the form of

$$m_j^*(E) = m_j^*\left(1 + \frac{E - \Delta E_j}{E_{g,j}}\right), \tag{37.19}$$

where $E_{g,j}$ is the strain modified bandgap. With all these material parameters, the input parameters in Equation 37.14 are known. Since m^* varies from layer to layer, strictly speaking, Equations 37.7 and 37.8 need to be evaluated individually in each layer. For simplicity, we assume m^* to be the effective mass at the GaAs conduction edge since optical transitions mostly occur in the GaAs layers.

37.3.2 Numerical Examples

In this section, we present some examples and compare them to experiments. To measure the α lineshape, infrared (IR) light is incident on a 45° polished facet at the edge of a substrate (referred to as the edge coupling method), so that the optical electric field pointing at the z-direction is a factor of $1/\sqrt{2}$ of that of the TM component. The measured quantity is the responsivity (R) or the conversion efficiency ($CE \equiv \eta g$). The values of R and CE are related by

$$R(\lambda) = \frac{I_p}{P(\lambda)} = \frac{e\eta g\Phi(\lambda)}{\frac{hc}{\lambda}\Phi(\lambda)} = \frac{\lambda}{\frac{hc}{e}}\eta g = \frac{\lambda}{1.24\times 10^{-6}}CE(\lambda)\ \left[\frac{\mathrm{A}}{\mathrm{W}}\right], \tag{37.20}$$

where $P(\lambda)$ is the incident optical power. Since in the weak absorption limit, η is directly proportional to α (see Section 37.4.4 for details on this coupling scheme), we can compare either the experimental $CE(\lambda)$ lineshape to the theoretical $\alpha(\lambda)$ or $R(\lambda)$ lineshape to $\lambda\alpha(\lambda)$. In Figure 37.4, we show the calculated $f_n(\lambda)$ of a typical QWIP, labeled as Det. A, which is made of 21 periods of 500 Å $Al_{0.27}Ga_{0.73}As$/46 Å GaAs QWs. The TMM calculation was performed on a single QW unit. The result will be the same if the calculation is performed on the entire MQW when E_1 of individual wells are not coupled (Choi et al., 2002), and in this case, the total f is independent of B.

Substituting the f spectrum into Equation 37.7, $\alpha(\lambda)$ can be obtained. Figure 37.5 shows the calculated $\alpha(\lambda)$ for Det. A along with three other GaAs/AlGaAs detectors having different cutoff wavelength λ_c and doping density N_D. Figure 37.5 also shows the measured CE lineshape of each detector in arbitrary units. To fit the measurements, the values of β_x needs to vary from 0.665 to 0.715 for the nominal x from 0.27

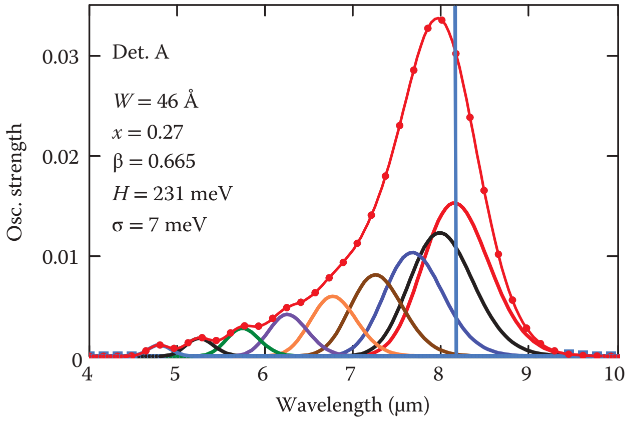


FIGURE 37.4 The calculated oscillator strengths f_n of a QWIP with $W = 46$ Å, $B = 500$ Å, $x = 0.27$, and $\sigma = 7$ meV. The curve with circles is the total oscillator strength. The vertical line shows λ at which the final state energy $= H - E_1$.

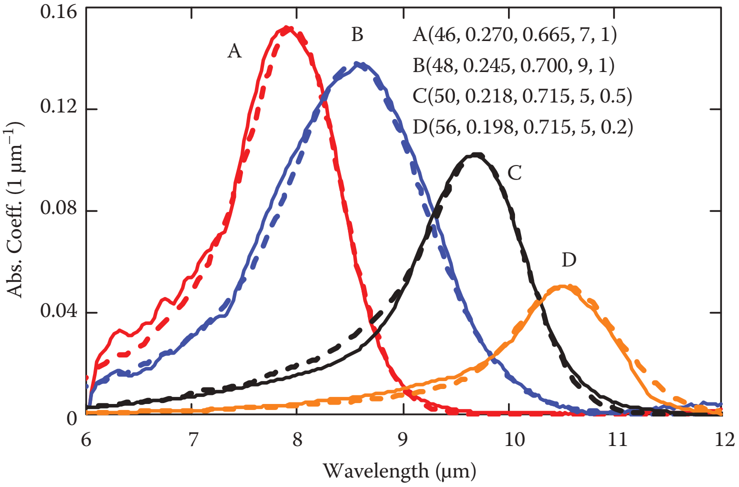


FIGURE 37.5 The measured CE lineshapes (solid curves) from four materials. The dashed curves are the calculated α assuming $(W, x, \beta_x, \sigma, N_D)$ shown in the figure, where N_D is in 10^{18} cm^{-3}. The barrier thickness is 500 Å for all detectors except 600 Å for Det. D.

to 0.198. The broadening parameter σ ranges from 5 to 9 meV for the first upper state, and it increases by about 0.5 meV for each successive upper state. With the two fitting parameters, β_x and σ, the intrinsic absorption lineshape can be satisfactorily explained.

The bandwidths shown in Figure 37.5 are rather limited. To broaden the bandwidth, one approach is to use superlattices (SLs) to create miniband-to-miniband (M–M) transitions. When the barrier thickness B is small, E_n states among different wells are coupled to form minibands M_n. Since the degenerate E_1 states develop into a range of occupied states in M_1, the absorption can cover a wider spectral range. This bandwidth can further be expanded if the QWs in the SL have different well widths (Choi et al., 2002).

To cover the entire 8- to 14-μm LWIR window, we designed a structure (Li et al., 2006) that contains seven SLs, and they are separated by 600-Å $Al_{0.17}Ga_{0.83}As$ barriers. Each SL contains seven 30-Å $Al_{0.23}Ga_{0.77}As$ barriers and four 70-Å and four 75-Å wells alternately placed together. The doping is 4×10^{17} cm^{-3} in the wells while the barriers are undoped. Since the eight miniband ground states within M_1 have different energies, their electron populations are also different. In this example, the first state in M_1 is at 39.4 meV and the last state is at 55.7 meV, while the common Fermi energy for the present doping is at 57.8 meV. Therefore, the electron population in the first state is 8.8 times larger than the last state, resulting in a significant modification of the lineshape from that of MQWs. Calculating Equation 37.7 with a different effective 3D doping density (N_s/L) for each miniband ground state according to the different Fermi levels, the spectral lineshape of the SL can be obtained. Figure 37.6 shows the calculated lineshape for the present SL. It agrees well with the measurement using a constant σ of 7 meV. The spectral responsivity is seen to cover the entire LWIR window. Following this material design approach, a QWIP with any cut-on and cutoff wavelengths can be designed.

By using InGaAs/AlGaAs QWs to increase the combined barrier height, a mid-wavelength IR QWIP can be developed (Choi et al., 2002). The illustrated structure contains eight $In_{0.34}Ga_{0.66}As/Al_{0.4}Ga_{0.6}As$ SL units separated by 500-Å $Al_{0.4}Ga_{0.6}As$ blocking barriers. Each unit consists of five wells and four barriers. The wells are made of 5-Å GaAs/25 Å InGaAs/5Å GaAs layers, and the barriers are made of 30-Å AlGaAs layers. There are top and bottom GaAs contact layers. All SL layers are doped to 2×10^{18} cm^{-3}. The modeled responsivity curve matches the experiment if $\sigma = 27.5$ meV as shown in Figure 37.7. This much larger σ is usually observed in the MWIR regime.

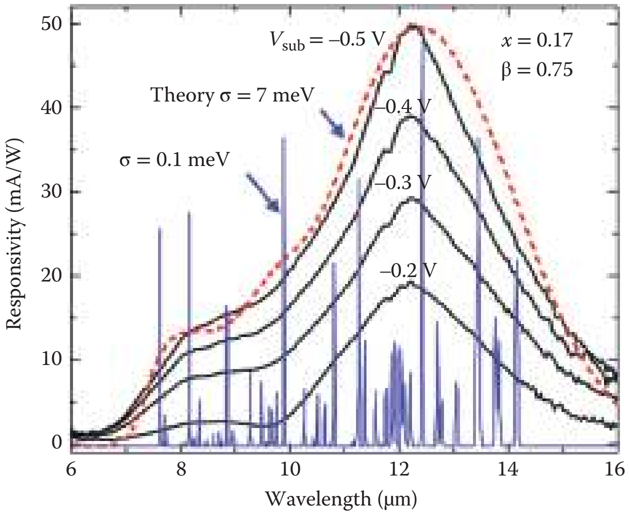


FIGURE 37.6 The measured responsivity from an SL structure with alternate well widths (solid curves). The fitted spectral lineshape of λα(λ) with σ = 7 meV is shown as the dashed curve. The curve with σ = 0.1 meV is to reveal the spectral distribution of individual M-M transitions.

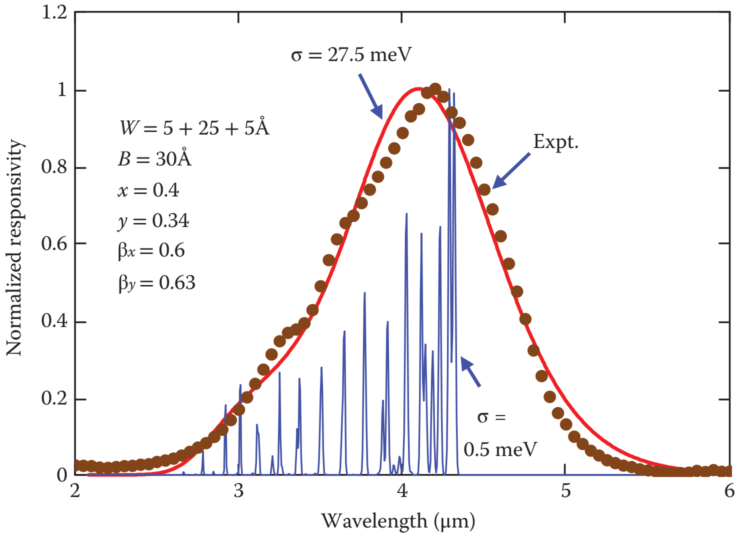


FIGURE 37.7 The measured normalized responsivity (circles) and the calculated $\lambda\alpha(\lambda)$ with $\sigma = 27.5$ meV (solid curve) for the combined transitions in the SL with blocking barriers. The curve with $\sigma = 0.5$ meV is to reveal the spectral distribution of individual M-M transitions.

37.4 Detector QE Modeling

Although $\alpha(\lambda)$ for parallel propagating light can be modeled accurately, the detector QE under normal incidence is still far from definite because it also depends on the optical coupling scheme. The search for an optimum coupling structure for a given application occupies the bulk of QWIP research to date. To yield the theoretical QE, one has to know the EM field distribution inside the detector volume in the presence of a polarization-dependent absorption coefficient. This task can only be solved by employing 3D numerical EM modeling. Since this capability was not readily available in the past, the common QWIP structures are not optimized.

With the advent of computer hardware and software technologies, one can now perform realistic and sophisticated EM modeling using commercial software on a personal computer. These advances have greatly facilitated the QWIP development. There are many numerical methods in solving EM problems. The more common ones are finite element method (FEM), finite-difference time-domain method, finite boundary method, and modal transmission-line (MTL) method, etc. In this chapter, we describe using FEM to model the theoretical QE of QWIPs.

FEM solves Maxwell's equation,

$$\nabla \times \left(\mu_r^{-1} \nabla \times \boldsymbol{E}\right) - k_0^2 \varepsilon_{rc} \boldsymbol{E} = 0, \tag{37.21}$$

via trial basis functions under a user-defined boundary condition. In Equation 37.21, μ_r and $\varepsilon_{rc} = \varepsilon_r - j\varepsilon_i$ are the relative permeability and relative permittivity, respectively, and $k_0 = 2\pi/\lambda$ is the free-space wavevector. Through variational principles, FEM transforms Equation 37.21 into a large set of linear equations, which can be solved by matrix multiplications (Sadiku, 2001). To yield a valid solution, it is important to present a correct boundary condition in the model.

The present modeling is performed in the RF Module of a commercial EM solver, COMSOL Multiphysics version 3.5 (COMSOL AB, Nov 2008). The modeling procedures involve selecting the EM analysis mode, building the 2D or 3D detector geometry, defining constants, variables, and functions, inputting subdomain properties, selecting appropriate boundary conditions, building mesh structures, setting solver parameters, performing computation, and using postprocessing to yield the required information.

To model the QE of a QWIP, we note that η can be expressed as (Choi et al., 2012, 2013a)

$$\begin{aligned}\eta(\lambda) &= \frac{1}{P_0}\int_V \mathrm{d}I(\mathbf{r}) = \frac{1}{P_0}\int_V \alpha(\lambda)I(\mathbf{r},\lambda)d^3r \\ &= \frac{\alpha(\lambda)}{A\frac{c\varepsilon_0}{2}E_0^2}\int_V \frac{n(\lambda)c\varepsilon_0}{2}\left|E_z(\mathbf{r},\lambda)\right|^2 d^3r = \frac{n(\lambda)\alpha(\lambda)}{AE_0^2}\int_V \left|E_z(\mathbf{r})\right|^2 d^3r, \qquad (37.22)\end{aligned}$$

where P_0 is the incident optical power, V the detector active volume, I the optical intensity associated with E_z, α the absorption coefficient, A the detector area, and n the material refractive index. Equation 37.22 states that the QE of a QWIP can be calculated from the volume integral of $|E_z|^2$ in the presence of a finite α. If E_z in Equation 37.22 is replaced by the total E, it will be applicable to detectors with isotropic absorption as well. Since E_0 and E_z are linearly proportional to each other, E_0 can be set arbitrarily. Besides E_0, the only free input parameter in Equation 37.22 is the wavelength-dependent $\alpha(\lambda)$ (if $n(\lambda)$ of the material is known accurately). For a known $\alpha(\lambda)$ from Equation 37.7, there will be no free parameters, and the value of $\eta(\lambda)$ for a given combination of material and geometry can be uniquely and unambiguously determined.

37.4.1 Modeling Validation and Phase Coherence

Since numerical EM modeling involves many steps, assumptions, and idealizations of materials and structures, it is critical to validate the model through theoretical and experimental means. One important assumption in the model is the complete phase coherence of radiation. While it is true for radio waves, it is generally not the case in the IR wavelengths and the rationale for EM modeling needs further explanation. If the optical field distribution in a detector indeed depends on the degree of coherence of light, the present modeling clearly will not be relevant. It follows that a QWIP will not have a definite QE, and its value depends on the light source characteristics. For a blackbody source, it is well known that the radiation is incoherent and its coherence length ℓ_c is only about $0.8\lambda_{\max}$, where $\lambda_{\max}$ is the free-space wavelength at the intensity maximum (Donges, 1998). At a blackbody temperature $T_B = 300$ K, $\lambda_{\max}$ is 10 μm, and ℓ_c is 8 μm in free space and 2.7 μm in GaAs. Such a short ℓ_c would suggest phase coherence exists only in a very small detector volume and coherent EM modeling would not be meaningful.

The abovementioned picture is certainly true when considering all of the radiations from the blackbody. Consistent with this picture, it also follows that the modeled coherent QE at $\lambda_{\max}$ will not be valid for all the blackbody radiations, which is also certainly the case. Nonetheless, when all the incoherent wave packets from the blackbody are Fourier transformed into the plane wave components, which happen to assume the Planck distribution, each individual wavelength component is in fact phase coherent. (By the definition of a plane wave with a definite λ, its coherence length is infinite.) Since the detector absorbs one photon with a definite λ at a time, not a wave packet with a broad spectrum at a time, the detector is interacting with each Fourier component individually, and because of this specific absorption feature, the spectrally dispersed radiation appears to be completely coherent to the detector at the individual wavelength component basis. Following this reasoning, there will only be one QE value for each λ, and it is the value evaluated under phase coherent condition. The modeled QE is thus an intrinsic property of the detector independent of the light source. Consequently, the detector maintains the same optical properties when detecting objects with different coherence lengths caused by, for example, different scene temperatures.

For experimental verification, when one of the Fourier components is separated out from a blackbody by an instrument, either using a grating or a filter, its coherence length $\ell_c = \lambda^2/\Delta\lambda$ becomes finite, where $\Delta\lambda$ is bandwidth of the instrument (Fowles, 1975). It is more convenient to use a light source with a long ℓ_c, i.e., a small $\Delta\lambda$, so that no QE convolution is need. However, even if $\Delta\lambda$ is substantial in comparison with the variation of modeled $\eta(\lambda)$, one can still make comparison by characterizing the spectral power density $P(\lambda)$ of the source around the nominal incident wavelength λ_0. The weighted theoretical $\eta(\lambda_0) = \int \eta(\lambda - \lambda_0)P(\lambda - \lambda_0)d\lambda / \int P(\lambda - \lambda_0)d\lambda$ can then be compared to experiment.

It is interesting to mention that when $\Delta\lambda$ of the instrument is broader than the detector absorption bandwidth Γ, the detector can only sense part of each wave packet emitted by the instrument, with which the coherence length appears to be longer at λ^2/Γ instead of $\lambda^2/\Delta\lambda$ for the original wave packet. Therefore, a detector with a given Γ will not be able to perceive the true phase coherence of a source if the source ℓ_c is shorter than λ^2/Γ. In the following, we present examples of modeling validation.

37.4.2 Planar Solar Cell

A planar GaAs solar cell provides a simple test for the modeling since it lacks any x–y spatial variations. Its material structure includes an active GaAs absorbing layer, a front antireflection coating, and a back metal reflector as shown in Figure 37.8 (Grandidier et al., 2012). With the known complex refractive indices of GaAs, and assuming 100% transmission for the antireflection (AR)-coating and 100% reflection for the reflector, a simple classical model for QE using ray-tracing technique is represented by the dash curve in Figure 37.9. On the other hand, using the complex refractive indices from all the constituent materials, the 2D modeling based on Equation 37.22 with isotropic absorption yields the solid curve. It coincides

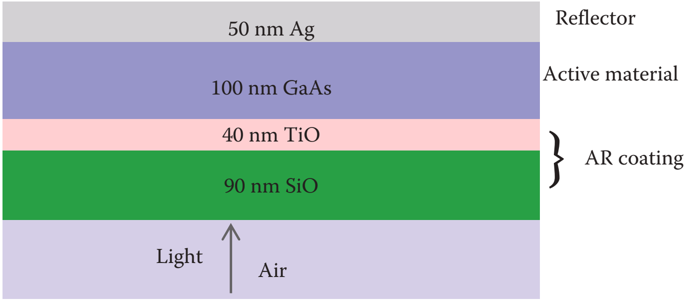


FIGURE 37.8 The modeled layer structure of a planar solar cell.

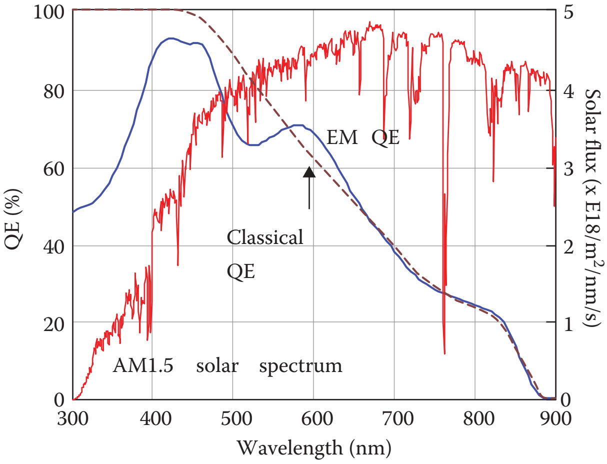


FIGURE 37.9 The quantum efficiency (QE) obtained by raytracing (dashed curve) and FEM (solid curve). The figure also shows the solar flux under 1-sun. (EM, electromagnetic.)

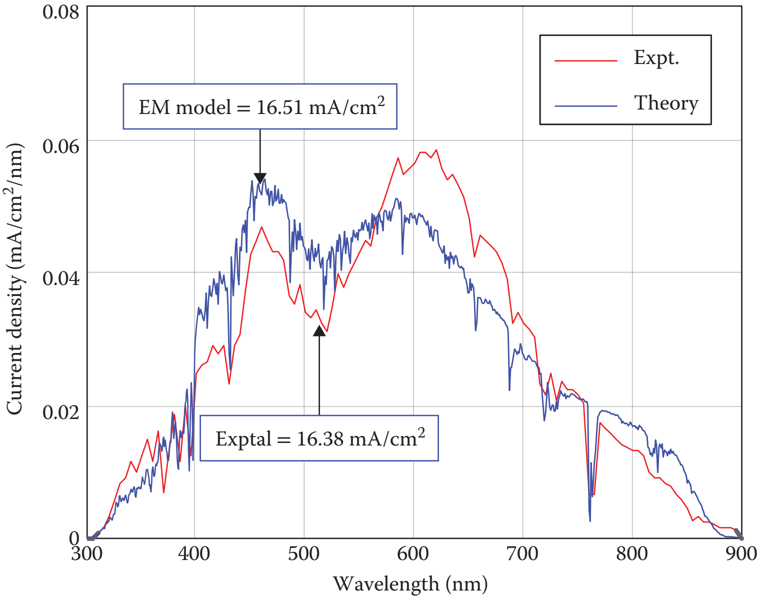


FIGURE 37.10 The measured and modeled photocurrent density under 1-sun without fitting parameters. (EM, electromagnetic.)

with the classical solution when the coatings in the modeling are effective. At the same time, the predicted photocurrent spectrum in Figure 37.10 is very similar to the measured spectrum (Grandidier et al., 2012) without any fitting parameters. In particular, the model accurately predicts two photocurrent peaks at 430 and 600 nm, which are absent in the simplified classical model. Furthermore, the model predicts the integrated photocurrent to be 16.51 mA/cm^2 under 1-sun condition, which is in excellent agreement with the measured 16.38 mA/cm^2. Therefore, without knowing the actual material properties in the experimental study, the agreement between theory and experiment is satisfactory for this device structure.

37.4.3 InAs/GaInSb IR Photodiode

When the detector material is fabricated into detector pixels, it acquires spatial variations. In this example, we show an IR photodiode in Figure 37.11 with a linear dimension of 22 μm in a 25-μm pitch array. It consists of a 2-μm thick p-InAs/Ga$_{0.75}$In$_{0.25}$Sb strained SL (p-SLS2) layer and a 0.5-μm thick wider bandgap n-SLS1 layer to form a photodiode. The photodiode is sandwiched between a 0.5-μm thick wider bandgap p-contact layer (p-SLS3) and a 2-μm thick n-InAs$_{0.9}$Sb$_{0.1}$ contact layer. On top of the p-contact, there is a 12-μm wide gold layer for ohmic contact. Light is incident from the n-contact layer. The detector is assumed to be surrounded by a semi-infinite epoxy material with $n = 1.5$ and with negligible IR absorption.

Since there is no AR-coating in this detector, Fabry–Pérot (FP) oscillations arise from optical interference between the top and bottom material interfaces. Because of the inhomogeneous phase change along the top surface, the FP oscillations are irregular. In Figure 37.12, we show the theoretical α spectrum of the p-SLS2 layer and the deduced α spectrum of n-InAs$_{0.9}$Sb$_{0.1}$ contact layer from open literature. The latter spectrum is obtained by shifting the experimental n-InAs spectrum by $\lambda = 0.7$ μm in accounting for the bandgap reduction (Miles et al., 1990). We further assume α of the n-SLS1 and p-SLS3 to be the same as that of p-SLS2 but at the shorter wavelengths (Dixon et al, 1961). The $n(\lambda)$ value in Equation 37.22 is taken to be a constant of 3.5, which is the average between 3.4 for InAs and 3.7 for GaSb at 10 μm. All IR absorption is assumed to be isotropic. Since the absorption of all the layers except the p-SLS2 layer will be dissipated as heat and no photocurrents will be generated in these layers, their absorption reduces the QE of the p-SLS2 layer in IR detection.

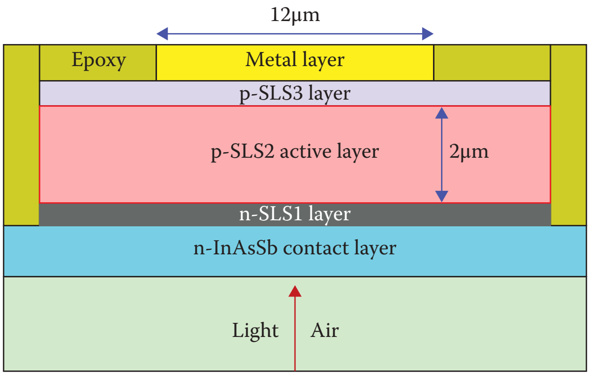


FIGURE 37.11 The material layer structure of a SLS photodiode. The pixel pitch is 25 μm.

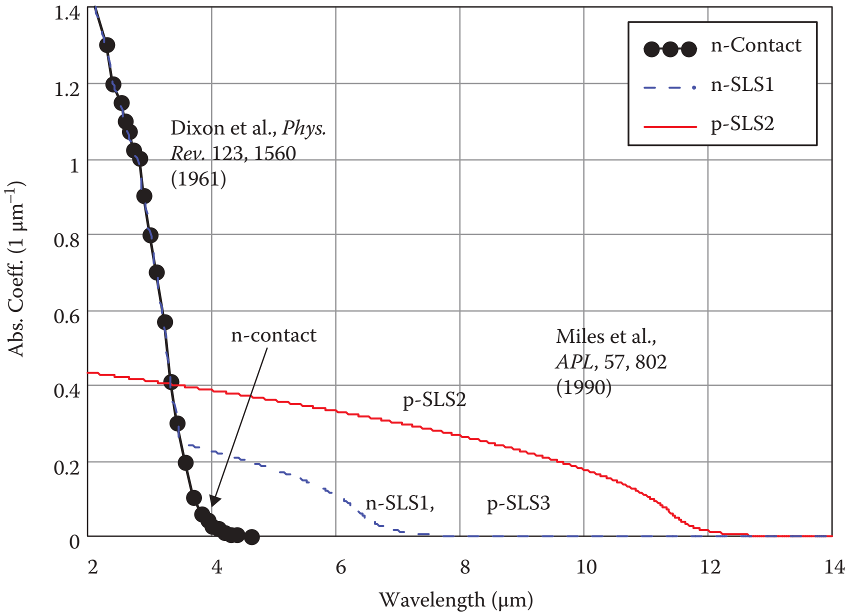


FIGURE 37.12 The theoretical and experimental α of individual layers obtained from open literature.

With these material and detector structures, the detector QE spectrum is modeled and the result is shown in Figure 37.13. Without adjusting any input parameters, the theoretical prediction matches satisfactorily with the experiment both in magnitude and in spectral lineshape in the entire spectral range. This example shows the validity of the modeling in the presence of spatial variations.

37.4.4 Edge Coupling for QWIPs

In Section 37.3.2, we assumed optical coupling efficiency using a 45° facet to be λ-independent in order to employ this coupling scheme for material characterization. It is worth checking its validity (Choi et al., 2012). Figure 37.14 shows the detector geometry and the calculated E fields with $E_0 = 377$ V/m. The detector has a 3-μm thick active QWIP layer, a 1-μm top contact layer, and a 1-μm gold layer. The detector is 100-μm wide. The value of α is assumed to be 0.15 μm^{-1}, and it is λ independent.

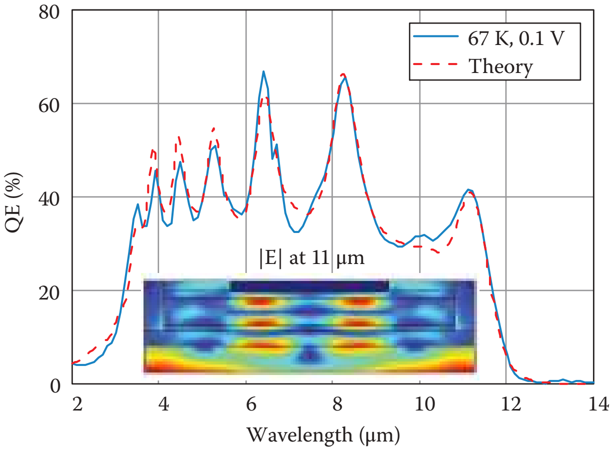


FIGURE 37.13 The calculated and the measured quantum efficiency (QE) of the p-SLS2 layer. The experimental data are provided by M. Jhabvala at NASA Goddard Space Flight Center. The insert shows the color plot of the |*E*| distribution at λ = 11 μm.

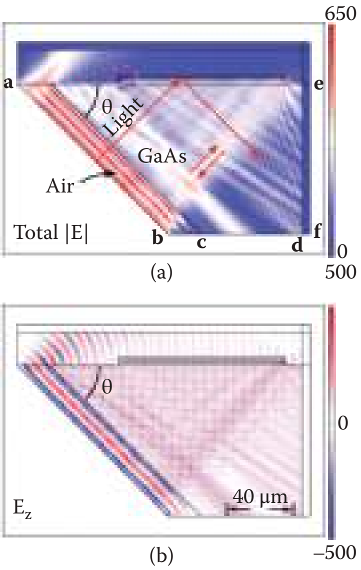


FIGURE 37.14 (a) The detector geometry at wedge angle θ of 45°. The arrows show the typical light paths. (b) The corresponding E_z is shown.

The 2D modeling with anisotropic absorption shows that QE at θ = 45° as well as other angles are nearly independent of λ as shown in Figure 37.15, justifying the experimental approach. The small oscillations are due to FP oscillations established in one of the detector corners shown in Figure 37.14a, but they will not be noticeable in the presence of edge roughness. In Figure 37.15, we also show the classical QE calculated from

$$\eta_c(\theta) = \frac{4n}{(1+n)^2}\frac{\cos\theta}{2}\left[1-\exp\left(-\alpha\sin^2\theta\cdot\frac{2\ell}{\cos\theta}\right)\right], \tag{37.23}$$

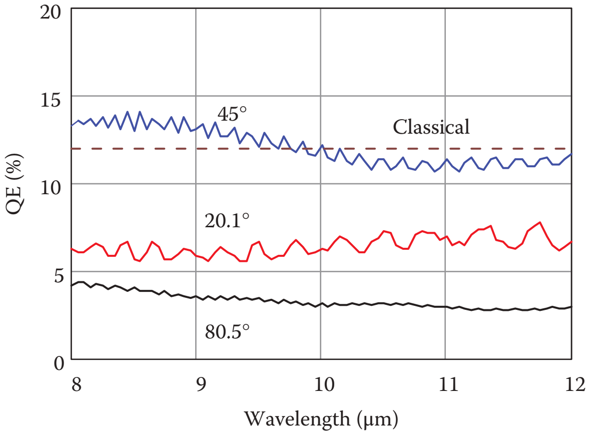


FIGURE 37.15 The calculated quantum efficiencies (QEs) at three different wedge angles. The dashed lines show the classical value at θ = 45°.

where $\ell = 3$ μm is the active layer thickness and $\theta = 45°$ is the angle of the polished wedge, which is also the angle of incidence measured from the detector normal. In Equation 37.23, $4n/(1+n)^2$ is the transmission coefficient of the GaAs substrate. The term $\cos(\theta)/2$ accounts for the smaller projected detector area in the direction of light and the fact that only the TM mode is coupled. The factor $\alpha \cdot \sin^2(\theta)$ is the absorption coefficient under oblique incidence, and $2\ell/\cos(\theta)$ is the total optical path length inside the detector accounting for the internal reflection at the top of the detector. In this classical model, if α is constant respect to the wavelength, η_c will also be wavelength independent. In Figure 37.15, η_c is shown to agree with the mean value from EM modeling. Therefore, EM modeling indeed supports the edge coupling approach in material characterization.

37.4.5 Quantum Grid Infrared Photodetector

Equipped with a quantitative model, we can revisit some of the QWIP coupling structures used in the past. Figure 37.16a shows the cross section of a structure known as the quantum grid infrared photodetector (QGIP) (Choi et al., 2004). It uses a metal layer on top of long and linear grid lines to scatter light incident from the bottom. The width w of the grid line acts as a half-wave dipole antenna and its dimension determines the detection wavelength. This detector geometry can also be modeled by 2D MTL method (Tamir and Zhang, 1996). Figure 37.16b compares the MTL solution and the present FEM solution for the same QGIP structure, assuming the same constant ε_i of $9.722 + i$. The close agreement in the QE spectrum between these two methods and the agreement between the methods and the experiment (Choi et al., 2004; not shown here) confirm the validity of both EM models. The E_z field distribution under the individual metal grid lines shows the expected scattering pattern from a dipole antenna.

37.4.6 Plasmonic-Enhanced QWIP

The structures described earlier were modeled in 2D. In this example, we perform 3D modeling as the structure cannot be reduced to 2D. The present structure, shown in Figure 37.17, uses surface plasmon to couple incident light into a QWIP (Wu et al., 2010). Surface plasmon is the resonant oscillation of conduction electrons at the interface between a metal and a semiconductor under optical excitation. The detector in this case contains a 400-Å thick gold film deposited on a 0.528-μm-thick InGaAs/InP active material.

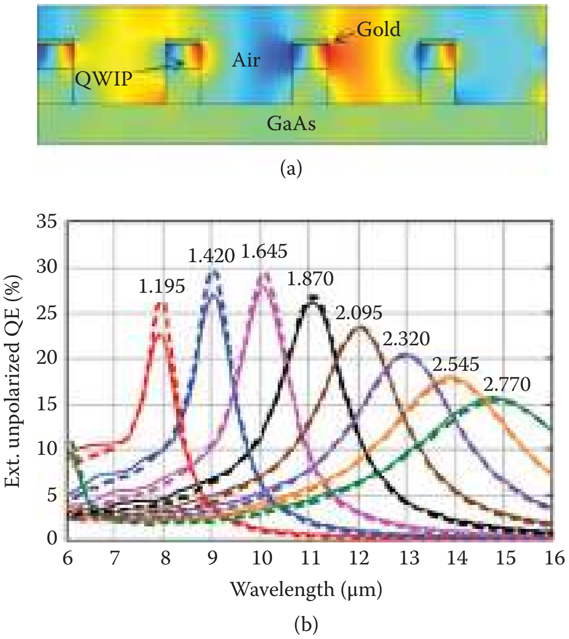


FIGURE 37.16 (a) The QGIP structure and the E_z distribution for $w = 1.87$ μm and $\lambda = 11.1$ μm. (b) The quantum efficiency (QE) spectra calculated based on FEM (solid curves) and MTL (dashed curves) having the same detector parameters. The numbers are w in microns. (QWIP, quantum well infrared photodetectors.)

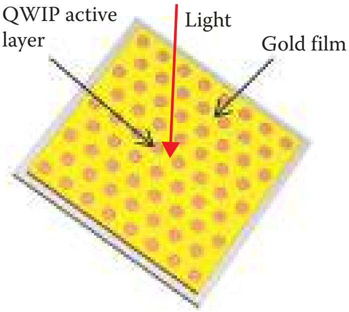


FIGURE 37.17 The 3D perspective of a plasmonic-enhanced quantum well infrared photodetectors (QWIP). Light is shining through the perforated gold layer from the top.

The gold layer is perforated with circular holes in a hexagonal lattice with spacing of 2.9 μm, and the hole diameter is 1.4 μm. The light incident on the gold layer excites surface plasmons in the opposite side with a resonant wavelength determined by the hole pattern. The resonant surface plasmons then create an intense vertical field at the gold/semiconductor interface. The active material has a low doping, with which the peak α is calculated to be 0.05 μm^{-1}. The modeled QE assuming this constant α for E_z is shown in Figure 37.18. It has a narrow peak at $\lambda = 8.3$ μm. Meanwhile, from the measured responsivity spectrum (Wu et al., 2010) and the estimated gain of 8.5 from a similar detector structure (Eker et al., 2010), the experimental peak QE is deduced to be 12.6%, which agrees with the modeling to within 10%, and the two spectra have similar lineshapes.

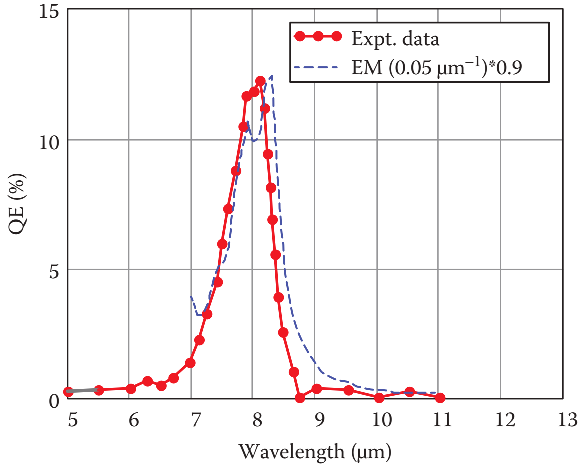


FIGURE 37.18 The measured (solid curve) and modeled (dashed curve) quantum efficiency (QE) of the plasmonic-enhanced QWIP. (EM, electromagnetic.)

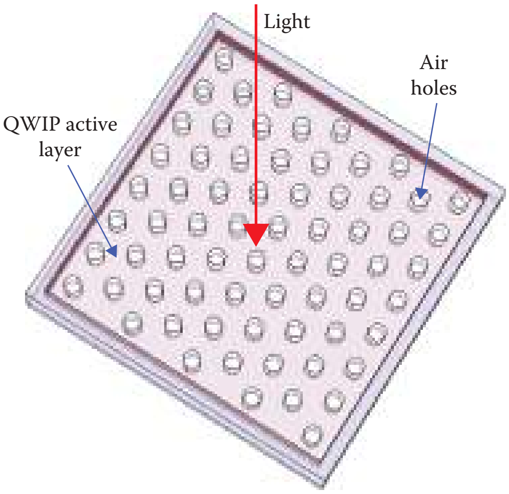


FIGURE 37.19 The 3D perspective of the PCS-QWIP. (QWIP, quantum well infrared photodetectors.)

37.4.7 Photonic Crystal Slab QWIP

To show another 3D optical structure, Figure 37.19 shows a photonic crystal slab QWIP (PCS-QWIP) studied (Kalchmair et al., 2011). The PCS is made of QWIP material, which is filled with periodic holes. These holes create a periodic modulation of refractive index in the x- and y-directions. Together with refractive index difference between GaAs and air in the z-direction, the PCS creates a series of characteristic photonic bands and bandgaps for the parallel propagating light. The normal incident light excites the stationary photonic eigenstates at the Γ-point of these bands in strong and sharp resonances. The present detector consists of an array of holes with spacing of 3.1 μm and diameter of 1.24 μm etched through the active and bottom contact layers. There is no metal layer in the optical path. The active material thickness is 1.5 μm, and the bottom contact thickness is 0.5 μm. The PCS is suspended in the air at a height t_{air} of 2.0 μm above the GaAs substrate.

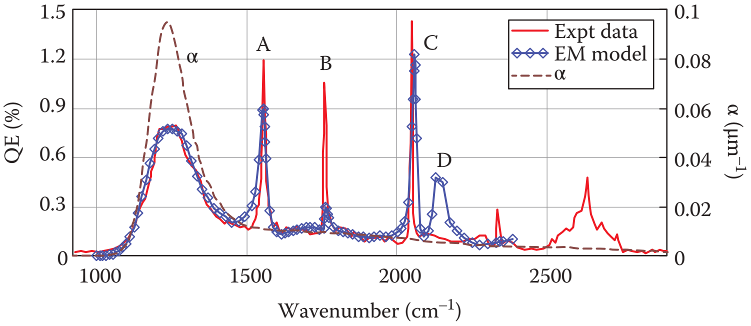


FIGURE 37.20 The measured photocurrent spectrum in arbitrary unit (solid curve) and the calculated QE spectrum based on the displayed α spectrum. (EM, electromagnetic.)

The measured photocurrent spectrum is shown in Figure 37.20. Based on the material α spectrum shown in Figure 37.20, which is deduced from the edge-coupled detector, the QE spectrum is modeled and it indeed contains the characteristic sharp peaks. However, the calculated peaks do not align exactly with the measurement. To obtain a better alignment, the theoretical hole spacing is reduced slightly from 3.1 to 2.9 μm as shown in Figure 37.20. This discrepancy could be due to the assumption of an average GaAs refractive index n to be 3.239 over a wide range of wavelengths. In reality, the value of n varies from 3.34 at 4 μm to 3.04 at 11 μm. Furthermore, the magnitudes of these peaks depend weakly on t_{air}. To obtain a larger peak at position B in Figure 37.20, t_{air} is adjusted from 2.0 to 0.9 μm. A different t_{air} could be due to the sagging of the PCS in the air at the operating temperature. From the modeling, the sharpness of these peaks is caused by two factors. One is the nearly symmetrical detector structure, both in vertical and horizontal directions, which induces strong resonances. The second is the weak material absorption at the peak wavelengths, which introduces only small damping effects on the resonances. The close match of the main peak and some of the side peaks lends support to the present modeling approach.

The large responsivity peaks A to C in the absorption tail illustrate that a large relative photoresponse can be obtained by using resonance in spite of a small intrinsic material absorption. Besides the examples presented earlier, we have also modeled other coupling structures, and the agreements between theory and experiment are similar to those presented (Choi et al., 2013a). They include air resonant cavities, linear and cross gratings, random-gratings, corrugated-QWIPs (C-QWIPs), etc. After studying all these examples, we found that the present FEM is highly accurate when comparing with other theoretical models. It also agrees well with experiments in both magnitudes and spectral lineshapes when both modeling and experimental uncertainties are taken into account.

37.5 EM Design of Resonator-QWIPs

The present FEM solves Maxwell equations numerically and thus should be able to optimize any detectors mentioned earlier as well as exploring new ideas. Among the many possible mechanisms and geometries, one structure that can yield a large QE is a photon trap. The function of a photon trap is to capture and hold incident light inside the detector until it is absorbed by the detector.

37.5.1 Resonator-QWIP Design

One realization of a photon trap is the resonator-QWIP (R-QWIP) as shown in Figure 37.21 (Choi et al., 2013a). It consists of an active layer, a thin GaAs substrate layer, and a number of diffractive elements (DEs) made of the same GaAs material on top. The topmost surface is then covered with a metal layer, and the detector is encapsulated with a low refractive index material.

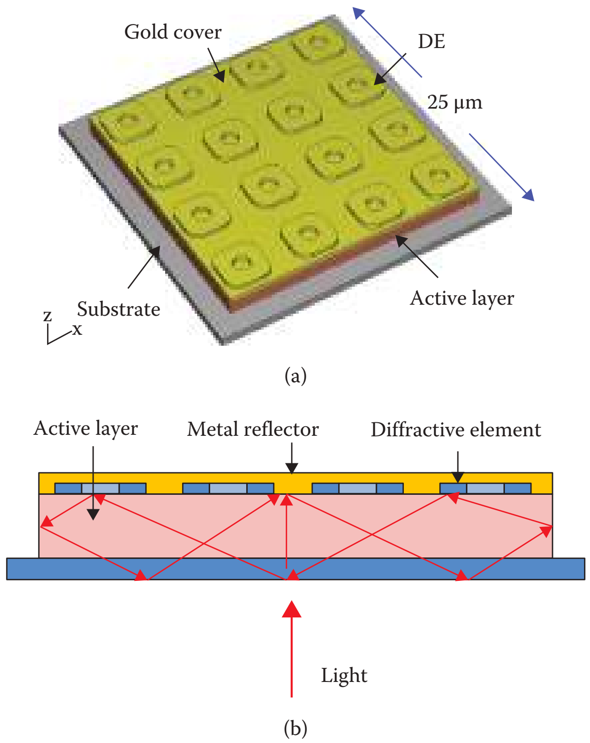


FIGURE 37.21 (a) A 3D perspective of an R-QWIP. Light is incident from the bottom. (b) The cross-section and one possible optical path.

The function of the DEs is to diffract incident light such that the subsequent angles of incidence at all detector boundaries are larger than the critical angle for total internal reflection (~18° for GaAs/air interface). If this is the case, the light will be totally confined in the detector. To account for interference effects, the size and shape of the detector volume are adjusted such that the scattered optical paths form a constructive interference pattern inside the detector. Under this condition, the newly incident light will be able to reinforce the light already under circulation, and the optical energy can be accumulated and stored in the detector as in a resonator. Therefore, by designing the detector into a resonator with a diffractive surface, i.e., an R-QWIP structure, an effective photon trap can be obtained.

To demonstrate the light storage capability, we first set α to zero so that there is no optical absorption in the detector. We then adjust the detector dimensions such that the intensity enhancement ratio (r_{ie}) given by

$$r_{ie} = \frac{1}{\frac{c\varepsilon_0}{2}E_0^2}\frac{nc\varepsilon_0}{2}\frac{1}{V}\int_V |E|^2 d^3r \tag{37.24}$$

reaches a maximum. The detector dimensions include the shape, size, and spacing of the DEs, the DE distribution pattern, the material layer thicknesses, the detector areal dimensions, etc. The quantity r_{ie} compares the average intensity built up inside the active volume relative to the incident intensity in the air. Figure 37.22 shows that r_{ie} can attain a value of 20 at 9.6 μm with the optimized structure shown in Figure 37.21a. For QWIPs, only the vertical intensity initiates absorption. The corresponding r_{ie} in Figure 37.22 for the vertical polarization component is 11. Such a large vertical r_{ie} produces a QE of 71% when α becomes finite and equals to 0.2 μm^{-1}, despite the very thin active thickness of 1.2 μm is adopted.

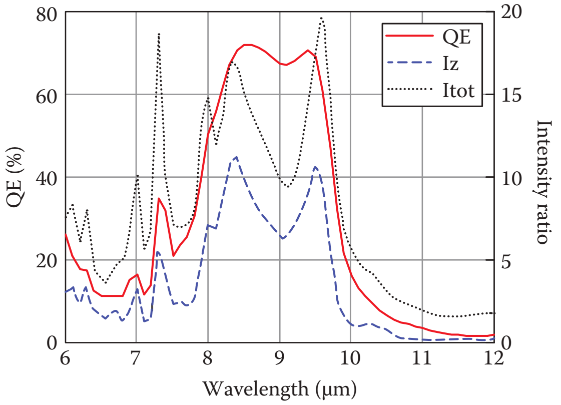


FIGURE 37.22 The calculated r_{ie} for the total intensity (dotted curve) and the vertical intensity (dashed curve) assuming α = 0. The figure also shows the calculated QE (solid curve) if α = 0.2 μm^{-1}. The detector dimension is 22 μm. (DE, diffractive element.)

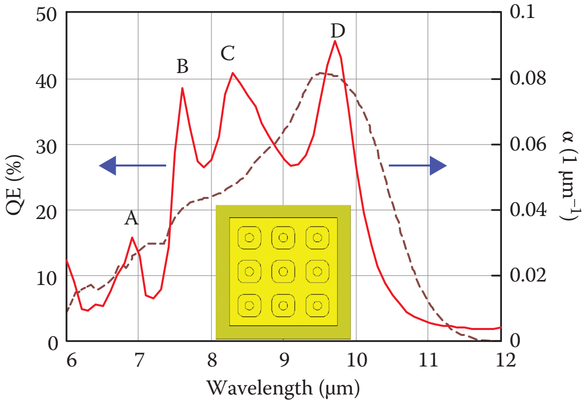


FIGURE 37.23 The measured α spectrum and the modeled quantum efficiency (QE) of an R-QWIP if α = 0.08 μm^{-1}. The pixel size is 20 μm, and the pixel pitch is 25 μm. The insert shows the DE pattern used for optical coupling.

37.5.2 R-QWIP Realization

We have conducted experimental research into R-QWIPs, and found QE as high as 70% in materials with doping larger than 1×10^{18} cm^{-3} (Choi et al., 2013b, 2014, 2015). In this section, we present a more recent R-QWIP to show its advantages. The material consists of 19 QW periods. Each period is made of 50-Å GaAs/60 Å $Al_{0.22}Ga_{0.78}As$/50-Å GaAs/500-Å $Al_{0.22}Ga_{0.78}As$ double-barrier (DW) structure and each well is doped to $N_D = 0.5\times10^{18}$ cm^{-3}. The adoption of the DW structure is to widen the absorption width, and the theoretical peak α is 0.08 μm^{-1} for this N_D. The measured α lineshape based on edge coupling is shown in Figure 37.23. Its bandwidth of 3 μm is about twice that of the single well design. This experimental lineshape and a theoretical peak value of 0.08 μm^{-1} is used for QE modeling. A matching resonator structure for this spectral range is shown in the insert of Figure 37.23. Assuming a λ-independent α of 0.08 μm^{-1} and n = 3.05, the resonator shows four resonant peaks (A, B, C, and D) in Figure 37.23, and the highest peak value is 45%.

The measured *I*–*V* characteristics of the R-QWIP averaged over 1600 pixels are shown in Figure 37.24a, and the QE spectral lineshape measured at −1.1 V averaged over 1000 pixels is shown in Figure 37.24b. Figure 37.24b also shows the calculated spectral lineshape using the material α(λ). It is seen that the relative

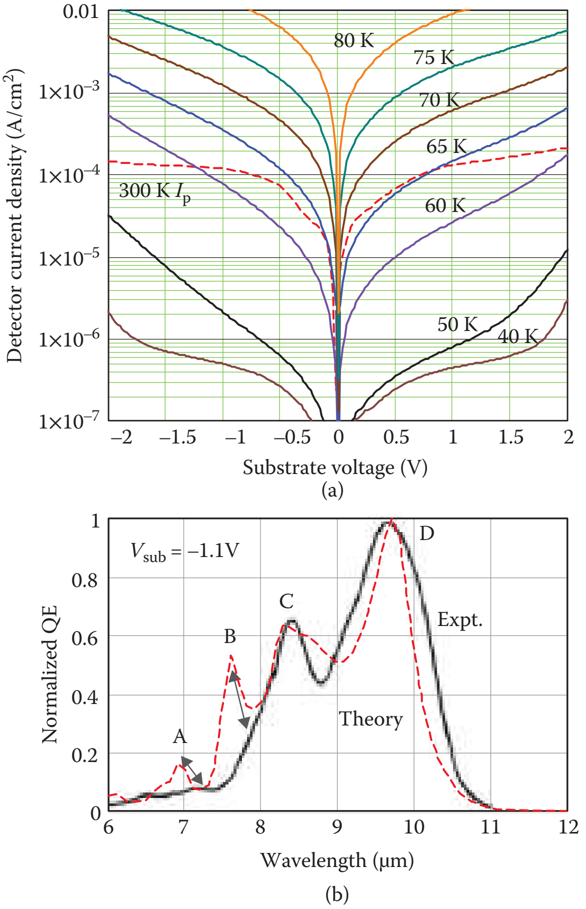
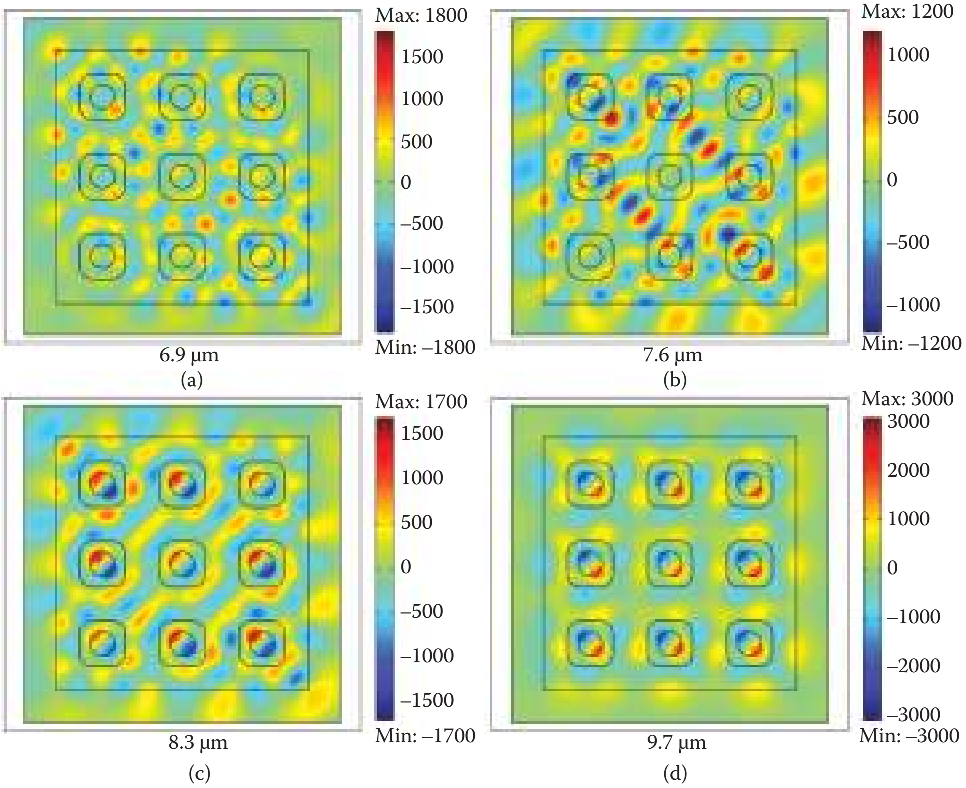


FIGURE 37.24 (a) The average *I*–*V* characteristics of 1600 R-QWIP pixels. (b) The measured average quantum efficiency (QE) lineshape of 1000 pixels in an FPA at −1.1 V and the modeled lineshape.

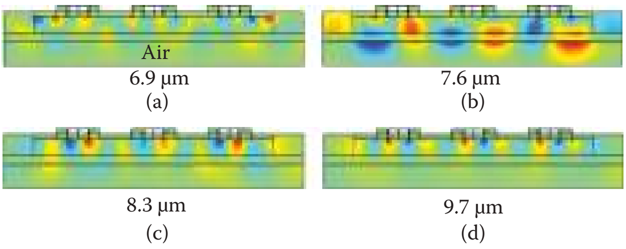
magnitudes of the experimental C and D peaks are as predicted while those of A and B peaks are much lower. The lower peaks also shift to longer wavelengths. This wavelength shift is expected when a constant n of 3.05 suitable only for 10 μm is used in modeling, while the actual n is known to increase toward shorter wavelengths. In addition to the less accurate n, the E_z distributions shown in Figure 37.25 at λ for A and B peaks indicate that they are higher order resonant modes that require higher precisions in the global structural dimensions. With the presence of processing nonuniformity, these resonances will be reduced.

For the QE magnitude, the main peak at 9.7 μm is measured to be 20% at −2 V and 37% at +2 V. The QE at positive bias is thus consistent with the modeling in this case. The polarity asymmetry is usually observed in R-QWIPs and is related to the highly localized optical intensity created in these detector structures as shown in Figure 37.26. For the two main peaks C and D, E_z is much stronger at the top of the active layer, where the QW layers are more active under positive bias. Therefore, this polarity produces a larger detection QE (Choi et al., 2014). In our simple dark current model described in Section 2.1, we assumed a linear potential drop in the MQW structure. In reality, the potential drop is slightly curved with potential drop near the cathode larger than the anode (Choi, 1997). This nonlinearity makes $\gamma(V)$ in $\eta_{\mathrm{abs}}(\lambda)\gamma(V)$ larger near the top contact region than the substrate under positive bias. While the change of $\gamma(V)$ along z does not affect the average I_{d} with uniformly excited thermal electrons, it does promote larger I_{p} under positive bias when the optical intensity is localized. Therefore, R-QWIPs usually prefer positive bias operation. For future modeling efforts, the inclusion of $n(\lambda)$ and $\gamma(z)$ are clearly needed.



6.9 µm
(a)
7.6 µm
(b)
8.3 µm
(c)
9.7 µm
(d)

FIGURE 37.25 The calculated E_z at the four peak wavelengths in the middle plane of the active layer. E_0 is 377 V/m. Note the change of scale in the plots.



6.9 µm
(a)
7.6 µm
(b)
8.3 µm
(c)
9.7 µm
(d)

FIGURE 37.26 E_z at the center cross-section of the R-QWIP with distribution in (a), (b), (c), and (d) corresponding to peak A, B, C, and D, respectively. The value of E_z decreases exponentially from the top toward the substrate in (a), (c), and (d).

The advantages of the R-QWIPs can be seen by comparing the corrugated-QWIP performance (C-QWIP) in Figure 37.3a and the R-QWIP performance in Figure 37.24a. The C-QWIP uses angled sidewalls to couple normal incident light (Choi et al., 2011, 2012). The cutoff wavelength λ_c of these two materials are similar, which is 11.2 µm for the C-QWIP and 10.6 µm for the R-QWIP. The C-QWIP shows $J_d = 9 \times 10^{-4}$ A/cm^2 at 3 V and 60 K and $J_p = 1.5 \times 10^{-4}$ A/cm^2 at 3 V. For the R-QWIP, $J_d = 2.5 \times 10^{-5}$ A/cm^2 at 1 V and 60 K and $J_p = 1.5 \times 10^{-4}$ A/cm^2 at 1 V. While the difference in the dark currents of the two detectors is a factor of 36, the photocurrents are the same. Accounting for the longer λ_c for the C-QWIP, the R-QWIP improves the I_p/I_d ratio by approximately a factor of 10. Figure 37.27 shows an IR image taken by the present R-QWIP FPA, which shows its excellent sensitivity and uniformity. The FPA operating condition is as follows: temperature = 61 K, applied bias = −1.1 V, $F/\# = 2.5$, signal integration

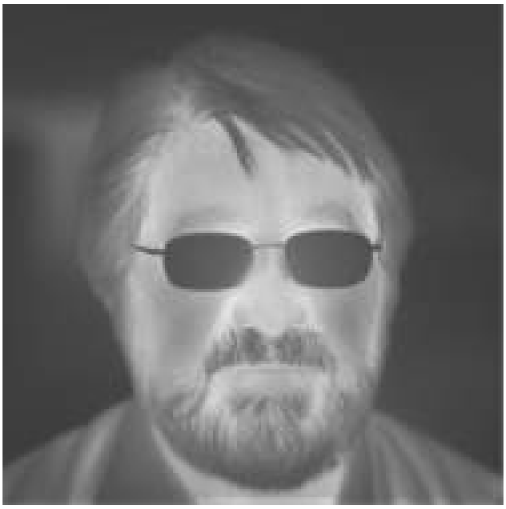


FIGURE 37.27 This image was taken by the R-QWIP FPA with 1024 × 1024 pixels.

time = 3.06 ms, and signal processing = 1-point background subtraction. Under this operating condition, the pixel operability is 99.5% and thermal sensitivity is 45 mK. Even higher sensitivity is expected at a lower temperature or at positive bias.

37.6 Conclusion

In this chapter, we have discussed the theoretical methods in modeling the optoelectronic properties of QWIPs. We found that analytical methods can be used to explain the detector dark current and photoconductive gain. The intrinsic material absorption properties can be calculated from the TMM, and the detector QE can be solved by performing finite-element EM modeling. With these methods, all of the relevant optoelectronic properties of a QWIP can be predicted with very few empirical parameters. This ability greatly facilitates QWIP FPA production, since it allows more consistent and capable products be designed and manufactured.

Due to limited space, some of the advantages of EM modeling have not been discussed. For example, EM modeling can be used to design structures that use thinner materials without antireflection coatings or optical filters to reduce manufacturing cost and improve uniformity; structures that are more immune to fabrication nonuniformity and have less pixel spatial and spectral cross-talk; structures that have special functionality such as extremely narrowband detection, two-color detection, broadband detection, and polarization detection; and structures that are very small in size but are still sensitive so that ultra-high-density FPAs can be produced. For the last benefit, the model predicts that a QE of 56% can be obtained with an absorption coefficient of 0.2 μm^{-1} and pixel size of $5 \times 5\ \mu m^2$. Its realization will enable the production of very high-resolution FPAs with much smaller system size, weight, power, and cost. Besides their usefulness to QWIPs, EM modeling and the photon trap concept are also applicable to other sensors and energy convertors in the IR and other nearby wavelengths. Therefore, one can foresee the increasing vital role played by EM modeling in advancing different optoelectronic technologies in the future.

References

Choi KK (1997) *The Physics of Quantum Well Infrared Photodetectors*. River Edge, NJ: World Scientific.

Choi KK (2012) Electromagnetic modeling of edge coupled quantum well infrared photodetectors. *J. Appl. Phys*. 111:124507.

Choi KK, Bandara SV, Gunapala SD, Liu WK, and Fastenau JM (2002) Detection wavelength of InGaAs/AlGaAs quantum wells and superlattices. *J. Appl. Phys.* 91:551–564.

Choi KK, Chen CR, Goldberg AC, Chang WH, and Tsui DC (1998) Performance of corrugated quantum well infrared photodetectors. *Proc. SPIE* 3379:441–452.

Choi KK, Dang G, Little JW, Leung KM, and Tamir T (2004) Quantum grid infrared spectrometer. *Appl. Phys. Lett.* 84:4439–4441.

Choi KK, Jhabvala, MD, Forrai D, Sun J, and Endres D (2011) Corrugated quantum well infrared photodetectors for far infrared detection. *Opt. Eng.* 50:061005-1–6.

Choi KK, Jhabvala MD, Forrai DP, Waczynski A, Sun J, and Jones R (2012) Electromagnetic modeling of quantum well infrared photodetectors. *IEEE J. Quant. Electron.* 48:384–393.

Choi KK, Jhabvala MD, Forrai DP, Waczynski A, Sun J, and Jones R (2013a) Electromagnetic modeling and design of quantum well infrared photodetectors. *IEEE J. Sel. Top. Quant. Electron.* 19:1–10.

Choi KK, Jhabvala MD, Sun J, Jhabvala CA, Waczynski A, and Olver K (2013b) Resonator-quantum well infrared photodetectors. *Appl. Phys. Lett.* 103:201113.

Choi KK, Jhabvala MD, Sun J, Jhabvala CA, Waczynski A, and Olver K (2014) Resonator-QWIPs and FPAs. *Proc. SPIE* 9070:907037.

Choi KK, Sun J, and Olver K (2015) Resonator-QWIP FPA development. *Proc. SPIE* 9451:94512K.

Dixon JR and Ellis JM (1961) Optical Properties of n-type indium arsenide in the fundamental absorption edge region. *Phys. Rev.* 123:1560–1566.

Donges A (1998) The coherence length of black-body radiation. *Eur. J. Phys.* 19:245–249.

Eker SU, Arslan Y, Onuk AE, and Besikci C (2010) High conversion efficiency InP/InGaAs strained quantum well infrared photodetector focal plane array with 9.7 μm cut-off for high-speed thermal imaging. *IEEE J. Quantum Electron.* 46:164–168.

Fowles GR (1975) *Introduction to Modern Optics*, 2nd Edition. New York, NY: Holt, Rinehart, and Winston.

Grandidier J, Callahan DM, Munday JN, and Atwater HA (2012) Gallium arsenide solar cell absorption enhancement using whispering gallery modes of dielectric nanospheres. *IEEE J. Photovolt.* 2:123–128.

Hinnrichs M and Guptab N (2008) Comparison of QWIP to HgCdTe detectors for gas imaging. *Proc. SPIE* 6940:69401Q.

Jhabvala MD et al. (2011) Performance of the QWIP focal plane arrays for NASA's Landsat Data Continuity Mission. *Proc. SPIE* 8012:80120Q.

Kalchmair S, Detz, H, Cole GD, Andrews AM, Klang P, Nobile M, Gansch R, Ostermaier C, Schrenk W, and Strasser G (2011) Photonic crystal slab quantum well infrared photodetector. *Appl. Phys. Lett.* 98:011105.

Levine BF, Bethea CG, Hasnain G, Shen VO, Pelve E, Abbott RR, and Hsieh SJ (1990) High sensitivity low dark current 10 μm GaAs quantum well infrared photodetectors. *Appl. Phys. Lett.* 56:851–863.

Levine BF, Choi KK, Bethea CG, Walker J, and Malik RJ (1987) A new 10 micron infrared detector using inter-subband absorption in resonant tunneling GaAs/GaAlAs superlattices. *Appl. Phys. Lett.* 50:1092–1094.

Li J, Choi KK, Klem JF, Reno JL, and Tsui DC (2006) High Gain, broadband InGaAs/InGaAsP quantum well infrared photodetectors. *Appl. Phys. Lett.* 89:081128.

Miles RH, Chow DH, Schulman JN, and McGill TC (1990) Infrared optical characterization of InAs/Ga$_{1-x}$InxSb superlattices. *Appl. Phys. Lett.* 57:801–803.

Sadiku MNO (2001) Numerical techniques in electromagnetics, 2nd Edition. Boca Raton, FL: CRS Press.

Shi JJ and Goldys EM (1999) Intersubband optical absorption in strained double barrier quantum well infrared photodetectors. *IEEE Trans. Elect. Dev.* 46:83–88.

Tamir T and Zhang S (1996) Modal transmission-line theory of multilayered grating structures. *J. Lightwave Technol.* 14:914–927.

Wu W, Bonakdar A, and Mosheni H (2010) Plasmonic enhanced quantum well infrared photodetector with high detectivity. *Appl. Phys. Lett.* 96:161107.

38

Optical Modulators

38.1 Introduction.. 363
38.2 QCSE-Based Modulators 364
Figures of Merit • Structure • Franz–Keldysh Effect • Quantum-Confined Stark Effect • Modeling Absorption Spectrum • Poisson-Drift-Diffusion Model • Schrödinger Equation for QW States and Excitons • Permittivity Spectrum • Numerical Results
38.3 Electro-Optic Effect Modulators 371
The Mach–Zehnder Modulator • Traveling-Wave Design
38.4 Plasma-Effect Modulators 374
38.5 Si Electroabsorption versus Electrorefractive Effect 374
38.6 Resonant Modulators 377

Dominic F. G. Gallagher

and

Dmitry Labukhin

38.1 Introduction

The modulation of an optical signal is a key requirement of modern optical communications. Even when the transmitter is a laser diode, which can itself be directly modulated, there are advantages of using an unmodulated laser and an external modulator on the output.

Optical modulators can be constructed using many different physical effects. The main ones of interest are as follows:

1. Linear electro-optic, or Pockel's effect—an electrostatic field is used to modulate the refractive index of certain crystals.
2. The plasma effect—the alteration of the free carrier density in a semiconductor can change both the absorption and the refractive index.
3. The Franz–Keldysh (FK) effect and the related quantum-confined Stark effect (QCSE)—where an electrostatic field changes the band structure of a semiconductor, altering both the absorption spectrum and the refractive index.

In choosing a modulator, the following factors should be taken into account:

1. Do you want a phase modulator or an absorption modulator?
2. The modulation bandwidth needed.
3. Its energy efficiency, typically measured in pJ/bit.
4. The optical bandwidth required—some designs are based on resonant structures and can modulate only a narrow range of wavelengths, others can work over very wide optical bandwidths.
5. Size.

There is no such thing as a pure phase modulator or pure absorption modulator. The Kramers–Kronig relationship states that the dispersion spectrum (real part of the refractive index) can be known given the absorption spectrum over all frequencies, and vice versa:

$$\varepsilon'(\omega) = 1 + \frac{2}{\pi} P \int_0^{\infty} \frac{\omega' \varepsilon''(\omega')}{\omega'^2 - \omega^2} \mathrm{d}\omega' \tag{38.1}$$

$$\varepsilon''(\omega) = -\frac{2\omega}{\pi} P \int_0^{\infty} \frac{\varepsilon'(\omega') - 1}{\omega'^2 - \omega^2} \mathrm{d}\omega' \tag{38.2}$$

where the dielectric constant $\varepsilon = \varepsilon' + i \cdot \varepsilon''$, and P is the *Cauchy principal value*. The Kramers–Kronig relationships tell us that the largest change in real index occurs near a change in absorption. Thus the most efficient phase modulators are likely to have absorption at nearby wavelengths.

38.2 QCSE-Based Modulators

38.2.1 Figures of Merit

Design of an electroabsorption modulator (EAM) requires accurate computation of its key characteristics. Among most important figures of merit are (a) on/off ratio defined as the output light power ratio between on and off states of the EAM; (b) the change in absorption coefficient per unit applied voltage; and (c) maximum data transmission speed. Calculating (a) and (b) is based on accurate modeling of absorption spectrum as a function of applied voltage, which will be the main focus of this chapter. Computing (c) involves modeling carrier dynamics. A detailed account of the latter can be found in [1–3].

38.2.2 Structure

Let us consider a typical traveling-wave EAM shown in Figure 38.1. This type of structure is used in this chapter as an example. Although other configurations have been reported as well (e.g., VCSEL-type EAM [4]), the traveling-wave type is the most common and the general principles of its operation can be extended to others with some modifications.

In a traveling-wave configuration, the light propagation dimension (z in Figure 38.1) is much greater than light- and carrier-confining dimensions (x and y in Figure 38.1) and, therefore, z can be separated from the x and y for computational efficiency. This means that we can split the device along axis z into a finite

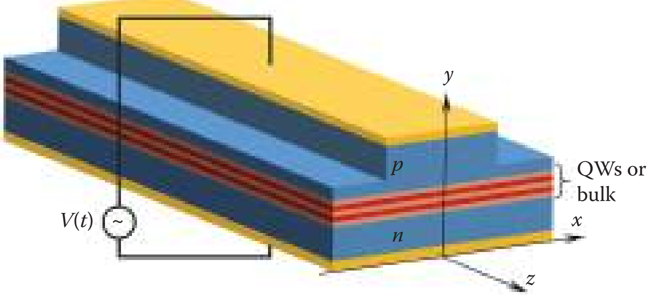


FIGURE 38.1 Traveling-wave EAM structure. The light signal propagates in the z-direction. Light confinement in the x–y plane can be realized via various configurations, the rigde waveguide shown in the figure being just one example. Epitaxial layers are grown in the y-direction. Top and bottom layers are metallic contacts. QWs and barriers comprise the active region. Next to the metallic contact layers are the substrate (n-doped) and cladding (p-doped). (QW=quantum well.)

number of short segments and assume optical properties within each segment to be z-independent, thus reducing the problem of modeling the optical properties to two transversal dimensions, x and y. Furthermore, in many cases, when the variation of the optical field and carrier density in one transversal direction (e.g., y in Figure 38.1) is much greater than in the other, the problem becomes one-dimensional. Then the calculated optical properties, namely, *material absorption* and *refractive index*, are used as input to an optical mode solver to compute the *effective index* and *modal absorption* of a particular waveguide mode that carries the optical signal. Finally, the result is passed as input to a one-dimensional z-propagation algorithm to obtain the device's characteristics. The first stage of this sequence, modeling material absorption and refractive index, is the focus of this chapter.

The optical absorption in EAMs is caused by several processes, among which are carrier transitions between conduction and valence band, free-carrier absorption (FCA), intraband transitions, photon–phonon interaction, etc. By far, the strongest contribution within the relevant spectral bandwidth is made by the first of these. It can be controlled by electrical bias applied across the epitaxial structure via one of the two main physical mechanisms: the FK effect in bulk layer structures or the QCSE in quantum wells (QWs). It is the latter one that is used in the majority of EAMs due to its higher efficiency. However, the FK effect is briefly examined here for the benefit of general understanding of these two phenomena.

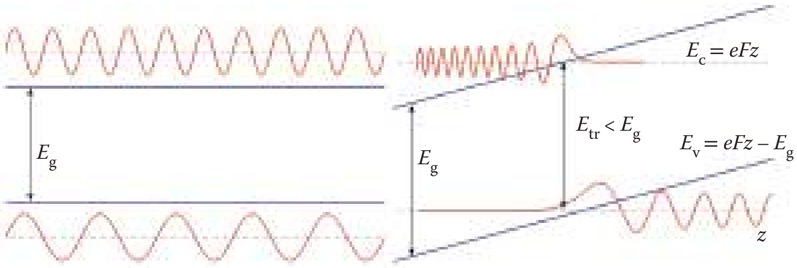
38.2.3 Franz–Keldysh Effect

The Franz–Keldysh (FK) effect is the change in photon absorption spectrum induced by reverse bias applied to a bulk semiconductor structure (Figure 38.1). In an unbiased bulk layer, the electron and hole states are represented by space-harmonic functions. The probability of a photon-absorbing transition between two states is proportional to the overlap integral between their respective wavefunctions. When electric field F is applied across spatial direction y, the electron and hole wavefunctions change their shape and become Airy functions as shown in Figure 38.2b.

As the overlap integral between them changes, the probability of the photon absorption changes accordingly. Furthermore, the minimum absorption frequency falls below the bandgap, as can be seen from Figure 38.2b. This enables the control of the transmitted power of the light traveling through the EAM with applied voltage at a given wavelength (Figure 38.3). However, absorption change per applied voltage is relatively low and QW structures are preferred due to the higher efficiency of the QCSE.

38.2.4 Quantum-Confined Stark Effect

The QCSE is a counterpart of the FK effect in QW structures. It has, however, an additional feature: the presence of thin peaks in the absorption spectrum caused by the excitonic transitions. Indeed, unlike the bulk structures, QWs facilitate strong electron and hole confinement (see Figure 38.4), which causes



(a) (b)

FIGURE 38.2 Electron and hole wavefunction in (a) an unbiased bulk semiconductor with bandgap E_g and when (b) a constant electric field F is applied across spatial direction y.

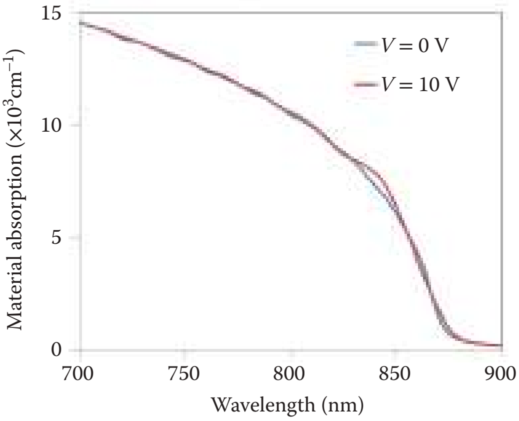


FIGURE 38.3 Change in absorption spectrum in 70-nm bulk AlGaAs under reverse bias due to the FK effect.

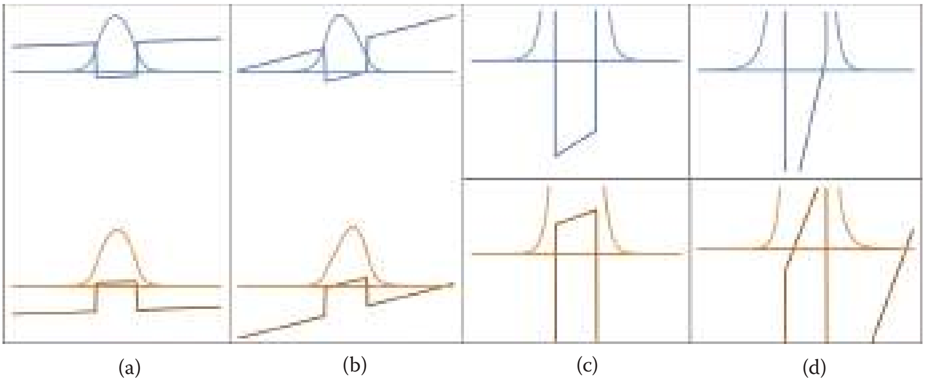


FIGURE 38.4 QW potential, electron and hole wavefunctions (a) without and (b) with a reverse bias. Plots (c) and (d) are (a) and (b), respectively, zoomed in to demonstrate the redshift in the energy levels.

the two carrier types to form bound states, *excitons*, similar to those between a proton and an electron in a hydrogen atom. These states introduce new discrete energy levels in addition to those of the QWs: but whereas the contribution to the absorption spectrum from the QW transitions has a staircase-like shape, the exciton contribution appears as peaks in the absorption spectrum.

Like in bulk layers, applying a reverse bias to the QW potential creates a double effect (see Figure 38.4). First, the electron and hole wavefunctions start "escaping" through the lower potential barrier, which decreases their overlap and, consequently, the probability of photon absorption. Second, their energy levels move closer to one another, which results in the redshift of the respective absorption peaks.

It can be seen from Figure 38.5 that if the absorption is measured at wavelength λ_0 near the exciton peak, the net effect of the two phenomena is the *increase* in absorption. This enables the control of the amount of light passing through the EAM by varying electric bias applied to the device. The material absorption change is much higher than in the case of the bulk structure. However, the material absorption is factored by the optical confinement factor, which is much smaller in QWs. To overcome this problem, multiple quantum wells (MQWs) are normally used to increase the efficiency of the device.

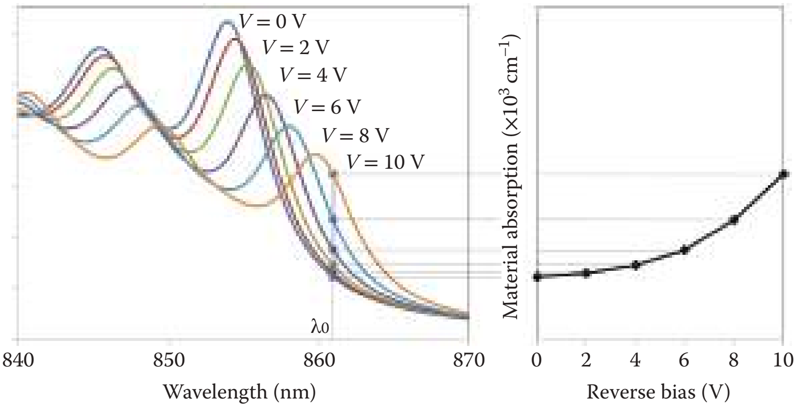


FIGURE 38.5 Change in material absorption spectrum in a 10-nm single QW AlGaAs EAM under reverse bias due to the QCSE effect. The peaks represent exciton's positions.

38.2.5 Modeling Absorption Spectrum

Modeling of the absorption spectrum as a function of voltage bias consists of several stages. Two most important effects contributing to the absorption is the carrier transitions between (a) energy levels created by QW confinement and (b) exciton binding energy levels. Photon absorption rate is proportional to the probability of the transition between an occupied state n in the valence band (hole) and an unoccupied state m in the conduction band (electron) multiplied by the Fermi population level $\alpha_{mn} \propto P_{mn}(f_m - f_n)$. The total absorption is given by the double sum over all relevant transitions mn, while each probability P_{mn} is determined from the overlap between the respective electron and hole wavefunctions. Therefore, the central part of the modeling is solving the Schrödinger equation in the QW potential and with an excitonic term added to its Hamiltonian. The QW potential, in its turn, is created by the juxtaposition of epitaxial layers with different bandgaps and applied electric field. It can be calculated using the Poisson-drift-diffusion model. In the case of EAMs, the coupling between the Schrödinger and Poisson equations can be neglected, which significantly reduces computational load.

38.2.6 Poisson-Drift-Diffusion Model

The Poisson-drift-diffusion model is based on three first-order differential equations: the carrier continuity equations for electrons and holes and the Gauss's law (see Chapter 3.1 in [5]),

$$\begin{aligned}\frac{\partial n}{\partial t} &= G_n - R_n - \nabla \cdot (\mu_n n \nabla \phi - D_n \nabla n)\\ \frac{\partial p}{\partial t} &= G_p - R_p - \nabla \cdot (\mu_p p \nabla \phi - D_p \nabla p)\end{aligned} \tag{38.3}$$

$$0 = \nabla \cdot \varepsilon \nabla \phi + q_e(p - n + N_D + N_A) \tag{38.4}$$

where $n = n(\boldsymbol{\rho}_{xy})$, $p = p(\boldsymbol{\rho}_{xy})$ are the electron and hole densities; $\boldsymbol{\rho}_{xy} = \{x, y\}$ is the spatial coordinate; $G_{n,p}(\boldsymbol{\rho}_{xy})$ and $R_{n,p}(\boldsymbol{\rho}_{xy})$ are generation and recombination rates, respectively; $\mu_{n,p}(\boldsymbol{\rho}_{xy})$ is electron and hole mobilities; $D_{n,p}(\boldsymbol{\rho}_{xy})$ is the electron and hole diffusion coefficients; $\varphi(\boldsymbol{\rho}_{xy})$ is the static field potential; $\varepsilon(\boldsymbol{\rho}_{xy})$ is the electric permittivity; $N_{D,A}(\boldsymbol{\rho}_{xy})$ is the donor and acceptor doping levels; and q_e is the electron charge.

Since in most EAMs the carrier density is very low, the generation and recombination terms can be discarded. With the reverse bias, the currents are also assumed negligible and, therefore, for the steady-state solution, the time derivatives are set to zero:

$$\begin{aligned} \nabla\cdot(\mu_n n\nabla\phi - D_n\nabla n) &= 0\\ \nabla\cdot(\mu_p p\nabla\phi - D_p\nabla p) &= 0\\ \nabla\cdot\varepsilon\nabla\phi + q_\mathrm{e}(p - n + N_\mathrm{D} + N_\mathrm{A}) &= 0 \end{aligned} \tag{38.5}$$

Equation 38.5 is solved with respect to $\{n, p, \varphi\}$, whereas $\mu_{n,p}$, $D_{n,p}$, ε, and $N_{\mathrm{D,A}}$ are usually assumed to be independent of $\{n, p, \varphi\}$ and are precomputed from the epitaxial layer composition. The potential φ is then used to find the spatial profile of the band edges:

$$E_{\mathrm{c,v}}(\boldsymbol{\rho}_{xy}) = E_{\mathrm{c0,v0}}(\boldsymbol{\rho}_{xy}) - \phi(\boldsymbol{\rho}_{xy}) \tag{38.6}$$

In addition to static field potential, the Poisson-drift-diffusion model calculates the distribution of carriers and, consequently, the distribution of charge in the PIN-junction as a function of applied bias. This is essential for calculating the device's characteristics in an electric circuit, such as capacitance, resistance, and, eventually, the modulation response.

38.2.7 Schrödinger Equation for QW States and Excitons

The next step after finding the band edge potentials $E_{\mathrm{c,v}}(\boldsymbol{\rho}_{xy})$ is to solve the Schrödinger equation in order to find energy levels and spatial distributions of electron and hole wavefunctions. There exist several methods to do so, with different tradeoffs between accuracy and computational efficiency. The general form of the Schrödinger equation is

$$\hat{H}\Psi = E\Psi \tag{38.7}$$

The Hamiltonian operator $\hat{H}$ in the presence of excitonic effect is split into electron, hole, and exciton parts [6]:

$$\begin{aligned} \hat{H} &= \hat{H}^\mathrm{e} + \hat{H}^\mathrm{h} + \hat{H}^\mathrm{exc}\\ \hat{H}^\mathrm{e} &= \hat{H}^\mathrm{e}_\mathrm{kin} + E_\mathrm{c}(\boldsymbol{r}_\mathrm{e})\\ \hat{H}^\mathrm{h} &= \hat{H}^\mathrm{h}_\mathrm{LK} + E_\mathrm{v}(\boldsymbol{r}_\mathrm{h}) \end{aligned} \tag{38.8}$$

where $\hat{H}^\mathrm{e}_\mathrm{kin} \propto \nabla^2_\mathrm{e}$ is the kinetic part of the electron Hamiltonian, $E_{\mathrm{c,v}}(\mathbf{r}_{\mathrm{e,h}})$ is the QW potentials with applied bias for conduction and valence bands, respectively, as calculated by solving Equations 38.5 and 38.6. Note that the spatial coordinates for electrons and holes are different. $\hat{H}^\mathrm{h}_\mathrm{LK}$ is the Luttinger–Kohn Hamiltonian for holes. The latter can be a 4×4 matrix operator, in case if only heavy and light holes are taken into account [6]. When highly strained structures are considered, the influence of the spin-orbit split-off band becomes significant and it needs to be included in $\hat{H}^\mathrm{h}_\mathrm{LK}$, increasing its size to a 6×6 matrix operator [29,30]. The exciton part, $\hat{H}^\mathrm{exc}$, is due to the Coulomb interaction between electron and holes. The additional term in the Hamiltonian of the system is, therefore,

$$\hat{H}^\mathrm{exc} = \frac{q_\mathrm{e}}{4\pi\varepsilon_0\varepsilon}\frac{1}{|r_\mathrm{e} - r_\mathrm{h}|} \tag{38.9}$$

As the first step, the energies and the wavefunctions of the conduction and valence bands without the exciton contribution are obtained by solving the Schrödinger equation with respective Hamiltonians:

$$\begin{aligned}\hat{H}^{\mathrm{e}}\psi_m^{\mathrm{e}}(\mathbf{k},\mathbf{r}) &= E_m^{\mathrm{e}}\psi_m^{\mathrm{e}}(\mathbf{k},\mathbf{r})\\ \hat{H}_{\mathrm{LK}}^{\mathrm{h}}\psi_n^{\mathrm{h}}(\mathbf{k},\mathbf{r}) &= E_n^{\mathrm{h}}\psi_n^{\mathrm{h}}(\mathbf{k},\mathbf{r})\end{aligned} \tag{38.10}$$

It is assumed that the scale of the spatial variation of parameters in the x- and z-directions is insignificant in comparison to that in y-direction. The solution of these equations is, therefore, obtained by separating their spatial dependence: the wavefunction is assumed to be harmonic in the x- and z-directions and (Equation 38.10) is reduced to a one-dimensional eigenvalue problem with a y-dependent Hamiltonian only (note that in some references quoted here the perpendicular axis is denoted z, not y).

The one-dimensional Schrödinger equation can be solved using a finite-difference method or a so-called "shooting" method. In the case of the effective mass approximation (parabolic dispersion relation $E(\mathbf{k}_{xz})$), it only needs to be solved for $\mathbf{k}_{xz} = 0$, where $\mathbf{k}_{xz}$ is an in-plane wave vector (see, e.g., [7]).

If nonparabolicity of the dispersion curve needs to be taken into account, Equation 38.10 must be solved for a range of $\mathbf{k}_{xz}$ in the vicinity of zero. This can present a considerable computational challenge. The axial approximation removes the dependency of $\hat{H}_{\mathrm{LK}}^{\mathrm{h}}$ on azimuthal angle θ and the two-dimensional vector $\mathbf{k}_{xz}$is replaced with its magnitude k, which significantly reduces computational time (see, e.g., [6]).

After the energies and wavefunctions in Equation 38.10 are found, the second step is to compute the exciton transition energy levels and wavefunctions with the full Hamiltonian:

$$\hat{H}\,\Psi^X(\mathbf{r}_{\mathrm{e}},\mathbf{r}_{\mathrm{h}}) = E^X\Psi^X(\mathbf{r}_{\mathrm{e}},\mathbf{r}_{\mathrm{h}}) \tag{38.11}$$

Here X denotes different exciton states, such as 1s, 2s, 2p, etc., called so by analogy with the atomic states. Two main methods of solving this problem are the momentum-space method [6] and the variation method [7].

In the momentum-space method, the exciton function is expanded in terms of conduction and valence wavefunctions $\psi_n^{\mathrm{e}}(\mathbf{k}_{xz}, y_{\mathrm{e}})$ and $\psi_m^{\mathrm{h}}(\mathbf{k}_{xz}, y_{\mathrm{h}})$ (found by solving Equation 38.10) in the $\mathbf{k}_{xz}$ space. As a result, Equation 38.11 is reduced to an eigenvalue problem with the integral operator on the right-hand side (see Equations 38.18 and 38.19 in [6]). Similarly to Equation 38.10, the axial approximation can be applied to full Hamiltonian $\hat{H}$: vector $\mathbf{k}_{xz}$ is replaced with its absolute value k, and the dependence of the wavefunctions on the azimuthal angle is confined to the phase term only: $\varphi(\mathbf{k}_{xz}) \to \varphi(k)e^{j\theta l}$. As mentioned earlier, the exciton states with $l = 0, \pm1, \pm2$, and ±3, are s, p, d, and f.

The exact mathematical details can be found in [6], but it is important to mention here that the final equation will contain a quadruple integral: over the spatial domain of the electron wavefunctions y_{e}, over that of the hole wavefunctions y_{h} spatial domain, and over the momentum angle of the in-plane vector $\mathbf{k}$; and over its magnitude k. The method was demonstrated to be in excellent agreement with the experimental results [6]. However, computing the quadruple integral for large structures (on the order of tens of nanometers in y-direction) can become prohibitively slow.

The variational method [7] is used to increase computational efficiency, albeit at the cost of accuracy. The exciton function in Equation 38.11 is expanded in terms of conduction and valence wavefunctions in real space ρ, not $\mathbf{k}$. Only the 1s exciton state is taken into account (so that the θ dependence is dropped, $\boldsymbol{\rho} \to \rho$):

$$\Psi^{1s}(\rho, y_{\mathrm{e}}, y_{\mathrm{h}}) = \phi_{mn}^{1s}(\rho)\psi_m^{\mathrm{e}}(y_{\mathrm{e}})\psi_n^{\mathrm{h}}(y_{\mathrm{h}}) \tag{38.12}$$

where $\phi_{mn}^{1s}(\rho)$ is a 1s trial function with variational parameter λ:

$$\phi_{mn}^{1s}(\rho) = \sqrt{\frac{2}{\pi}}\frac{e^{-\rho/\lambda}}{\lambda} \tag{38.13}$$

38.2.8 Permittivity Spectrum

Equations 38.7 through 38.13 serve to calculate the spectral positions of the exciton and QW transitions. The next step is to calculate the strength and the shape of the absorption $\alpha(\omega)$ as a function of light frequency ω due to these transitions and the corresponding change in the refractive index $n_{\text{re}}(\omega)$, both making part of the complex permittivity $\varepsilon(\omega) = [n_{\text{re}}(\omega) + i\alpha(\omega)c_0/(2\omega)]^2$, where ω is the angular frequency of light and c_0 is the speed of light in vacuum. The expression for total complex permittivity can be written as

$$\varepsilon(\omega) = \varepsilon^{\text{opt}} + \sum_{m(\text{e})} \sum_{n(\text{h})} (\varepsilon_{mn}^{\text{QW}}(\omega) + \varepsilon_{mn}^{\text{exc}}(\omega)) \tag{38.14}$$

where ε_{opt} is the permittivity for low optical frequencies, $\varepsilon_{mn}^{\text{QW}}(\omega)$ and $\varepsilon_{mn}^{\text{exc}}(\omega)$ is the permittivity change due to the QW and exciton transition between hole state n and electron state m. Whereas the QW contribution is a staircase-like function with steps located at energies $(E_m - E_n)$, the exciton transition is a delta function found at energies $(E_m - E_n - E_X)$. The exact formulae for the momentum-space and variational methods can be found in [6,7], respectively.

Two things must be pointed out. First, as mentioned earlier, the contribution both to the QW "steps" and exciton peaks is proportional to the overlap integral between the respective hole and electron functions:

$$\varepsilon_{mn}^{\text{QW,exc}}(\omega) \sim \int \psi_m^{\text{e}}(y)\psi_n^{h*}(y)dy \tag{38.15}$$

Second, due to the finite lifetime of the exciton states, these peaks have a spectral width $\simeq 1/\tau_{\text{exc}}$. Similarly, the steps of the staircase-like function of the QW contribution are smoothed out by a QW linewidth function, whose width is inversely proportional to the lifetime of the QW states. Mathematically, it is expressed as a convolution integral between $\varepsilon_{mn}^{\text{QW,exc}}(\omega)$ and respective linewidth functions $L_{mn}^{\text{QW,exc}}(\omega)$. The latter are complex-valued functions satisfying Kramers–Kronig relations (Equations 38.1 and 38.2) and they contribute to both the real and imaginary parts of the complex permittivity. Strictly speaking, each electron–hole pair should have its own $L_{mn}^{\text{QW,exc}}(\omega)$, but often in practice the same function is used for all pairs mn. There have been a few publications dedicated to calculating the shape of $L^{\text{QW,exc}}(\omega)$ (see, e.g., [9]). Still in many cases, a simple Lorenzian shape with an experimentally measured spectral width is used in calculations with satisfying accuracy [6].

38.2.9 Numerical Results

Below, experimental and modeling results are shown for several structures. Figure 38.6 shows the comparison between experimental results obtained for an SiGe modulator [10] and those simulated using the variational method. The variational method demonstrates excellent agreement in positions and shapes of the primary excitonic peaks and their dependence on the applied electric bias. It is not as accurate, however, in modeling higher energy spectra.

Furthermore, experimental results for an AlGaAs EAM, reported in [6], are compared with curves computed by the variational method and the momentum-space method (Figure 38.7).

It can be seen that the latter captures the general shape of the curve much more accurately, especially at higher energies. It can be argued, however, that the working frequency of the EAM is below the bandgap (to the left of the excitonic peak at zero bias) and it is, therefore, the most important spectral area to model. Moreover, the behavior in this area is largely dependent on the correct modeling of the exciton shape, not covered by either of the methods (see [9]). Therefore, in many cases, the variational method is preferred as a computationally fast alternative of the momentum-space method, especially for large MQW structures.

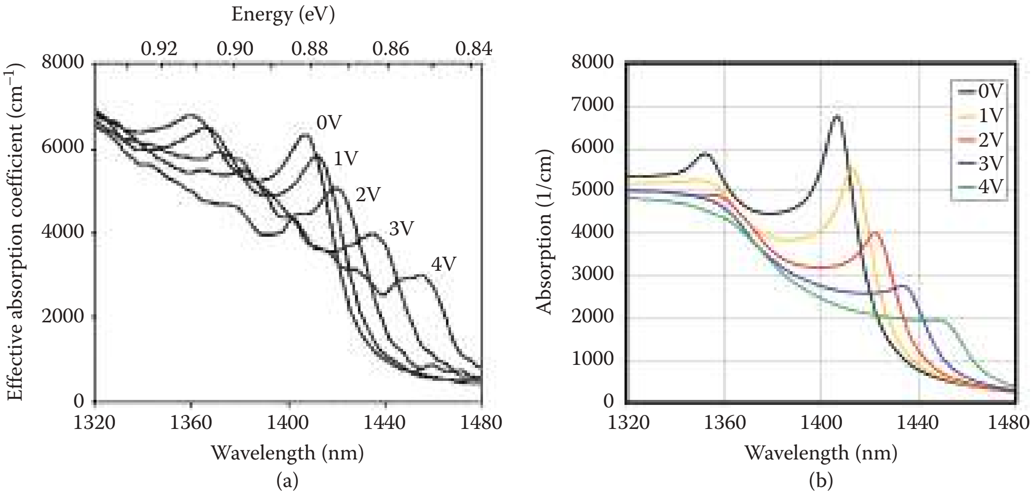


FIGURE 38.6 SiGe modulator. Transverse electric (TE) material absorption spectra for various values of the bias voltage: (left) experimental data from [10] (see detailed description of the structure in Section II of [10]) and (right) variational method by Photon Design Harold EAM.

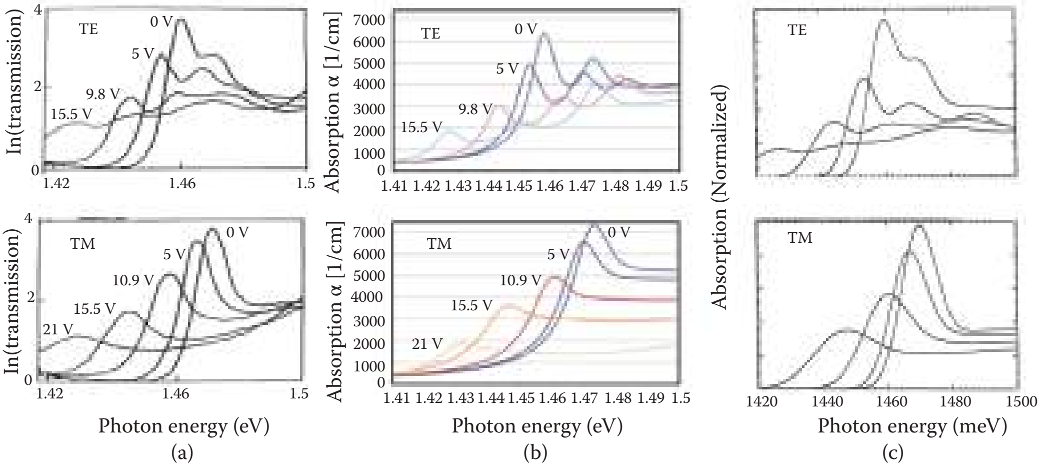


FIGURE 38.7 AlGaAs modulator [6] ([9.4 nm AlGaAs]/$Al_{0.3}Ga_{0.7}As$ QW, see detailed description of the structure in [31]). Absorption spectra for various values of the bias voltage: (a) experimental data from; (b) simulated curves using variational method for (top) TE polarization and (bottom) TM polarization; and (c) simulated curves using momentum-space method.

38.3 Electro-Optic Effect Modulators

The linear electro-optic effect, otherwise known as the Pockel's effect is present in crystals that do not have an inversion symmetry. This means they need an anisotropic dielectric tensor. Popular materials exploiting the Pockel's effect include $LiNbO_3$, $LiTaO_3$,and β-barium borate (BBO). The Pockel's effect is effectively instantaneous, once the electrostatic field has been established, making it suitable for exceedingly high-speed modulators.

$LiNbO_3$ has been a very popular material for electrorefractive modulators for many years because of its relatively large electro-optic coefficient [11,12], i.e., variation of refractive index with an applied electrostatic field. The rest of this section focuses on $LiNbO_3$ not because it is the only material available, but because of its dominant industrial position in photonics.

$LiNbO_3$ is strongly anisotropic implying a nontrivial dielectric tensor ε_{ij}. Since the tensor is symmetric, the double indexing conventionally simplified as

$$\begin{pmatrix} D_x \\ D_y \\ D_z \end{pmatrix} = \begin{pmatrix} \varepsilon_1 & \varepsilon_6 & \varepsilon_5 \\ \varepsilon_6 & \varepsilon_2 & \varepsilon_4 \\ \varepsilon_5 & \varepsilon_4 & \varepsilon_3 \end{pmatrix} \begin{pmatrix} E_x \\ E_y \\ E_z \end{pmatrix} \tag{38.16}$$

In addition, the variation of each $\varepsilon_{\mathrm{ij}}$ with applied electrostatic field depends on the direction of the electrostatic field E, leading to a second-rank tensor:

$$\Delta\left(\frac{1}{\varepsilon_i}\right) = \sigma_{ik}E_k \tag{38.17}$$

This implies that different optical polarizations will experience different modulations. In $LiNbO_3$, only a few of the electro-optic tensor coefficients are significant:

$$\sigma_{13} = 8.6\ \mathrm{pm/V}, \sigma_{22} = 3.4\ \mathrm{pm/V}, \sigma_{33} = 30.8\ \mathrm{pm/V}, \sigma_{51} = 28\ \mathrm{pm/V} \tag{38.18}$$

Unlike for example, silicon modulators, there is no macroscopic movement of charge except that associated with the capacitance of the dielectric: $C = \epsilon_r\epsilon_0 A/d$; d is thickness, A is area. This leads to a very high-speed modulator capable of working to 30 GHz or more [13]. In fact, the change in refractive index is near instantaneous once the electric field has been established and the speed of an $LiNbO_3$ modulator is limited almost entirely by it capacitance and the speed of the electrical circuit driving it.

As the electro-optic tensor coefficients given earlier imply, the effect is anisotropic in both the direction of the applied field and in the modulation of the optical dielectric tensor. The crystal must be carefully oriented to obtain the desired effect. The two favored orientations are denoted x-cut and z-cut. An x-cut crystal will create a strong modulation of a TE-like polarized optical mode in response to an electrostatic field in the same transverse axis as the optical one, exploiting the σ_{33} coefficient. An x-cut crystal requires a vertical electrostatic field to modulate the same TE-like polarized optical mode and exploits the σ_{51} coefficient.

38.3.1 The Mach–Zehnder Modulator

The change in refractive index can be converted to a change in signal amplitude using a Mach–Zehnder interferometer design (Figure 38.8). With sufficient voltage, a 180°phase shift between the two optical paths is created by the time the waveguides combine, causing full cancellation of the signal.

A modulator can be made using a field between an electrode placed on top of the $LiNb0_3$ chip and a ground plane underneath. However, since the chip is rather thick, this leads to a low field strength, so typically both electrodes are placed on top of the chip either side of the waveguide, as shown in Figure 38.9.

The design in Figure 38.9 has one big flaw. The electrodes are very close to the optical mode; this causes substantial optical losses especially in the TM-like polarization. To cure this, a cap layer can be deposited on the $LiNbO_3$ before the electrodes are deposited. The buffer has another advantage in that it alters the velocity of the electrical signal along the modulator, allowing it to better match the velocity of the optical signal. This substantially improves the modulation bandwidth.

The disadvantage of the buffer is a reduction in electric field strength in the vicinity of the optical mode, requiring a longer modulator to achieve the same phase change. In Figure 38.10, you can see the voltage contours of a design with a 100-nm SiO_2 buffer. Notice that a lot of potential is dropped across the buffer layer. This typically leads to a ~30% drop in modulator efficiency.

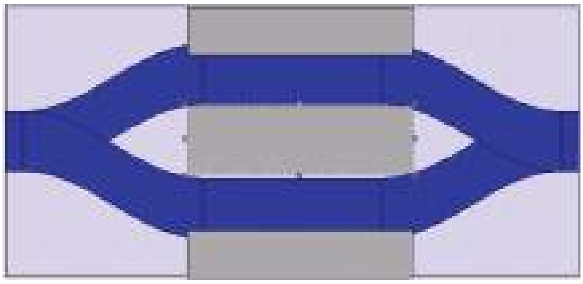


FIGURE 38.8 An $LiNbO_3$ modulator based on a Mach–Zehnder interferometer. A 3-dB splitter divides the input signal between two arms. One arm is given a positive electric field and the other arm a negative field, creating an increase in mode effective index in one branch and a similar decrease in effective index in the other.

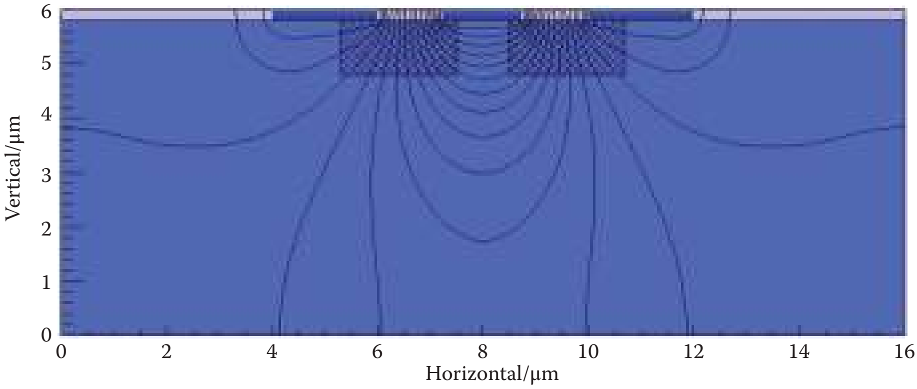


FIGURE 38.9 Cross section of a typical x-cut $LiNbO_3$ Mach–Zehnder modulator showing the Ti-diffused waveguides (hashed), three electrodes on top of the structure and electric potential contour lines. Typically, the outer electrodes are grounded and the central one carries the modulating signal. The opposite electrostatic field direction across each waveguide creates a natural push-pull effect for the Mach–Zehnder modulator. Note how the strongest field is generated in the middle of the waveguides.

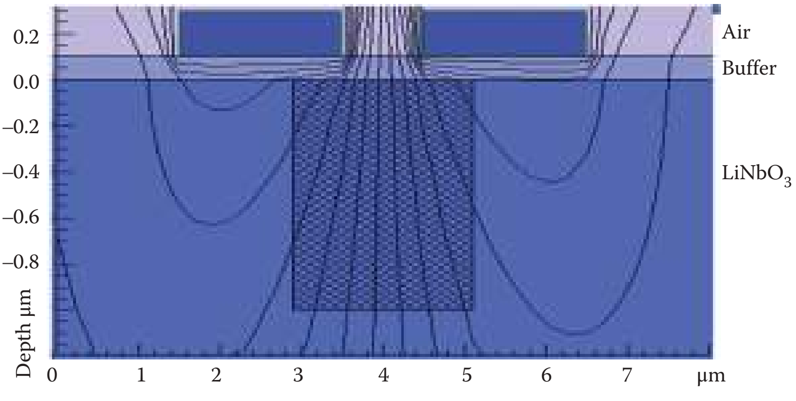


FIGURE 38.10 An x-cut $LiNbO_3$ modulator cross section with a cap layer creating an offstand between the electrodes and the optical mode. But note the strong field drop across the cap layer. This will reduce the electric field in the vicinity of the optical mode.

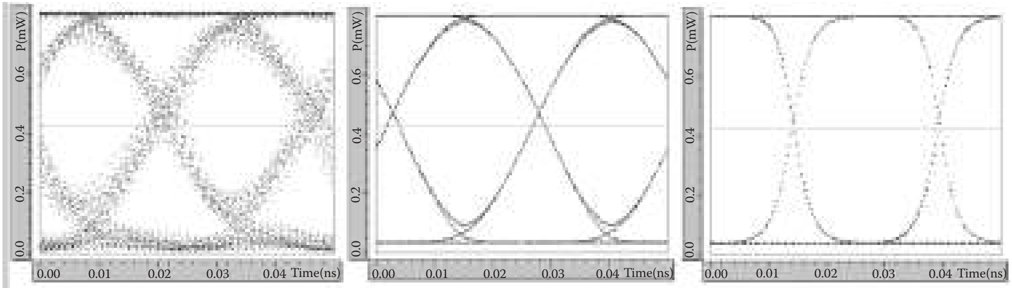


FIGURE 38.11 Eye diagrams of a traveling-wave $LiNbO_3$ MZI modulator at 40 Gb/s, computed with an optical circuit simulator PicWave [14] that includes traveling-wave time domain models for both the optical field and the electrodes: (left) with an unmatched electrode contacted a few millimeters from the end of the 20-mm-long modulator; (middle) contacted at the end; and (right) with a velocity-matched electrode.

38.3.2 Traveling-Wave Design

Because the change in effective index with voltage is typically very small, ~0.0001 per volt, the modulator has to be very long. This means that both the optical signal and the electrical signal will take a significant time to travel along the device. In order to maximize the modulation bandwidth, it is advantageous to match the velocity of the optical and electrical signals by appropriate design of the electrodes. Figure 38.11 shows some simulations of a traveling-wave $LiNbO_3$ Mach–Zehnder modulator under 40 Gb/s NRZ data stream. The figures show “eye diagrams” of the modulation response. To create an eye diagram the circuit is driven by a random bit sequence, and response of each clock period is superimposed. A clear open “eye” indicates a good signal.

38.4 Plasma-Effect Modulators

The plasma effect is the name given to the interaction of free electrons with an optical field. A plasma naturally occurs in a metal, which is why all metals are opaque. In order to make a modulator, one needs a way to control the carrier density, which is not realistic in a metal. In a semiconductor, the free-carrier population can be readily controlled by either injecting carriers into a diode (forward bias) or by lightly doping a region of the semiconductor and using a gate electrode or a reverse-biased junction to create an electrostatic field to deplete the carriers from the lightly doped region.

This section focuses on silicon-based plasma-effect modulators. Silicon is becoming a popular platform for manufacturing photonic components, largely because of the benefits of access to the huge electronic silicon manufacturing industry, its mature foundries and tools. Although silicon is not capable of generating light efficiently, i.e., making a laser, it is capable of absorbing light or modulating its phase by controlling the plasma effect in the optical path. Silicon optical modulators are now capable of reaching tens of gigahertz modulation speed, nearly rivalling the best $LiNbO_3$ modulators and in a much smaller volume, with much lower power requirements.

38.5 Si Electroabsorption versus Electrorefractive Effect

Silicon can act as an optical modulator either by a variation of the optical absorption in the silicon [15]—an “electroabsorptive modulator” or more popularly by modulation of the refractive index [16–19]—a so-called “electrorefractive modulator”. In fact this distinction is somewhat arbitrary, since all changes in the real part of the refractive index will cause a change in the imaginary part, and vice versa. As discussed in Introduction, the real and imaginary changes do not generally happen at the same wavelength and the relationship is described by the so Kramers–Kronig relationship. At a wavelength of 1.55 μm and for low doping densities, the real and imaginary changes are given approximately by [17]

$$\Delta n = -8.8x10^{-22}\Delta N - 8.6x10^{-18}\,(\Delta P)^{0.8}$$

$$\Delta\alpha = +8.5 \times 10^{-18}\Delta N + 6 \times 10^{-22}\Delta P$$

A big disadvantage of an electrorefractive modulator is that usually we want to modulate the amplitude of an optical signal and to create an amplitude modulation out of such a modulator generally requires interference effects in for example, a Mach–Zehnder interferometer [18–20] or a ring resonator [16,17, 21,22]. Both of these have a strong wavelength-dependent response, especially the ring resonator. Despite this, silicon electrorefractive modulators have arguably had the greater impact due to their lower power and higher modulation bandwidths.

Some of the earliest silicon modulator designs are based on carrier injection in a PIN diode [20–22]. Here a forward current is injected into the diode and the free charge carriers in the intrinsic region cause a change in refractive index. Figure 38.12 shows the cross section of a typical carrier-injection PIN diode modulator. The effect is only weakly dependent on wavelength, though once included in a Mach–Zehnder or resonant cavity structure, this advantage is lost. The effect is however only weakly dependent on temperature unlike QCSE or other bandgap-dependent effects.

An alternative Si modulator structure reverse biases a pn (or PIN) junction [18,19,23,24]. The reverse bias causes carrier depletion in the junction, and an associated alteration of the local refractive index. Figure 38.13 shows a possible geometry, one with a vertical junction [8]. In general, it is preferable to move the p-n junction to one side so most modulation is done by p-type carriers. This is because the p-carriers create less absorption at wavelengths of interest for a given modulation depth, since the absorption peak is moved further out.

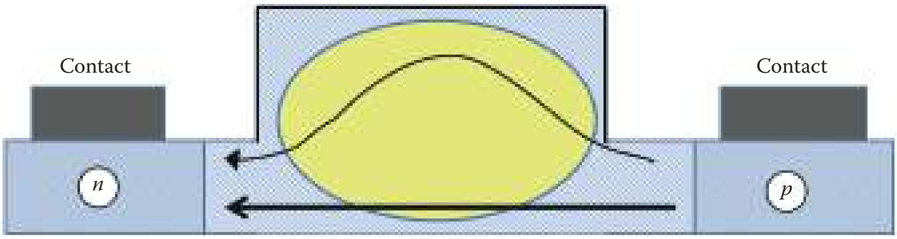


FIGURE 38.12 A PIN diode type silicon optical modulator. The silicon is etched to form a rib waveguide in the intrinsic region. Forward current interacts with the optical mode in the rib, changing its velocity or effective index.

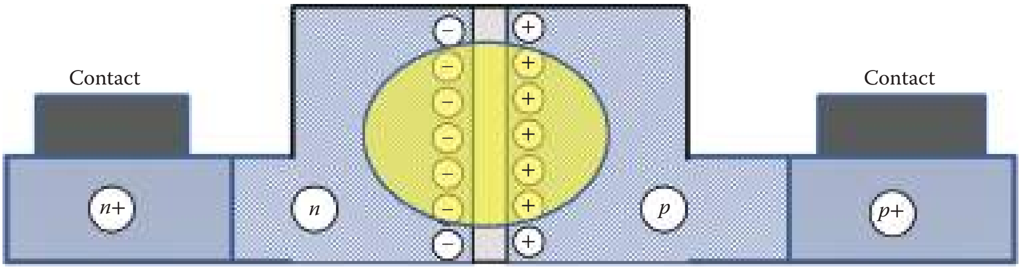
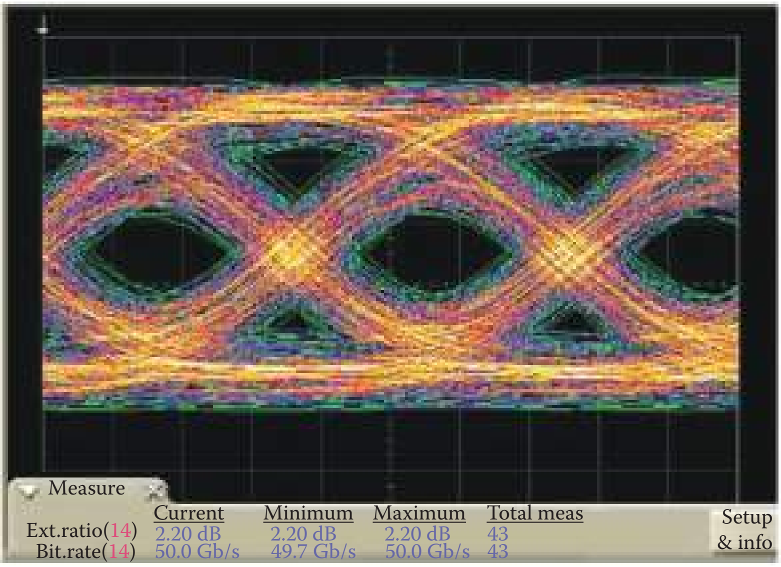


FIGURE 38.13 A reverse-bias type Si optical modulator. In these designs, the reverse bias caused electrostatic charge to build up at the p-n junction. These charges change the local refractive index.

Geometries similar to the former with asymmetric p-n junction have achieved a Vπ of 10 V in a 1-mm waveguide and bandwidths in excess of 50 Gb/s for a 1-mm long asymmetric Mach–Zehnder interferometer [19]. An experimentally measured eye diagram from this modulator is shown in Figure 38.14.

Figure 38.15 shows a recent design with a horizontal junction corresponding to a capacitor device [23]. The device is forward biased to cause carrier accumulation across the capacitor. This device achieved a modulation capability of 10 Gb/s at a length of 625 μm. Because the p and n carriers are close to each other, this design does not avoid the higher absorption of the n-type carriers.

Although most Si optical modulators are primarily electrorefractive, some electroabsorptive modulators have been demonstrated. In this case the free carriers are increased to a sufficient density that substantial FCA occurs. FCA is caused by carriers close to the band edge absorbing a photon and being excited to a state further up the conduction band (electrons) or down into the valence band (holes). The excited carriers then rapidly lose their energy by nonradiative transitions. An interesting design was recently proposed based on a Schottky diode [25]. The structure is predicted to reach a 3-dB bandwidth at 7 GHz for a 500-μm-long traveling-wave design. This is equivalent to a data rate of over 15 Gb/s using NRZ coding. Other groups have demonstrated EAMs using SiGe QWs—the QCSE effect (see Section 38.2) or by depositing graphene on the silicon [15,26,27].



	Current	Minimum	Maximum	Total meas
Ext.ratio(14)	2.20 dB	2.20 dB	2.20 dB	43
Bit.rate(14)	50.0 Gb/s	49.7 Gb/s	50.0 Gb/s	43

FIGURE 38.14 Experimental measurement of 50 Gb/s data stream in a depletion-type Silicon modulator in an asymmetric Mach–Zehnder interferometer configuration.

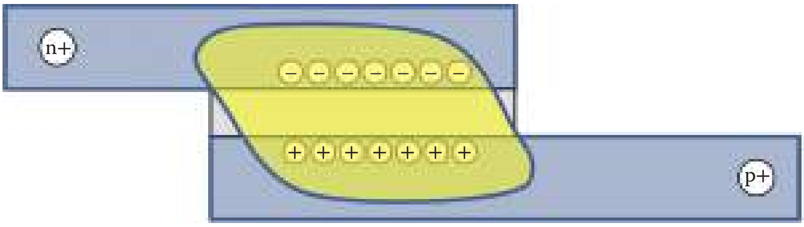


FIGURE 38.15 A capacitor-like structure also works by charge build up at the interface.

38.6 Resonant Modulators

In previous paragraphs, we outlined the use of traveling-wave designs such as the Mach–Zehnder interferometer to implement a modulator. Because the change in refractive index tends to be small for most modulator effects, this leads to long devices—hundreds of microns to tens of millimeter. This is not practical for many applications and so many have looked into resonant structures to reduce the modulator size. The principle is simple—in a resonant cavity, the light will remain trapped, bouncing or circulating around for a time much greater than L/c, where L is some characteristic dimension of the device. In this section, we focus on a popular structure—the ring resonator [16,17,21,22]—but many other designs are possible, for example, a photonic crystal cavity [28].

A resonator is popularly characterized by its "Q" factor, which can be defined as the photon lifetime divided by the optical period. A ring resonator has a roundtrip time of

$$t_{\text{round}} = \frac{2\pi R}{c}\mu \tag{38.19}$$

where μ is the characteristic refractive index, R is the ring radius. The photon lifetime of a cavity for optical frequency f is Q/f. This means that a ring resonator can increase the time spent in the ring by a factor of

$$\frac{Q\lambda_0}{2\pi R\mu} \tag{38.20}$$

This gives the modulator more time to change the phase of the signal, and allows the modulator to be shrunk be a corresponding factor. The disadvantage is that the ring resonator optical bandwidth is reduced to f/Q, making it sensitive to small changes in input wavelength.

A typical resonator configuration is shown in Figure 38.16 (left). This design has only one tangential waveguide and will act as a phase modulator. If a second tangential waveguide is added (Figure 38.16, right) then the modulator will direct the light into this when in resonance.

The high Q lowers the modulation speed of the device as well as its optical bandwidth. Figure 38.17 shows a PicWave [14] simulation of an eye diagram for a 20-μm-diameter silicon ring resonator modulator driven in carrier depletion mode with a 20 Gb/s NRZ bit pattern. Notice the overshoot caused by the high Q of the cavity, in this case over 6000. In practice, this ringing would be filtered out by a low-pass filter on the detector.

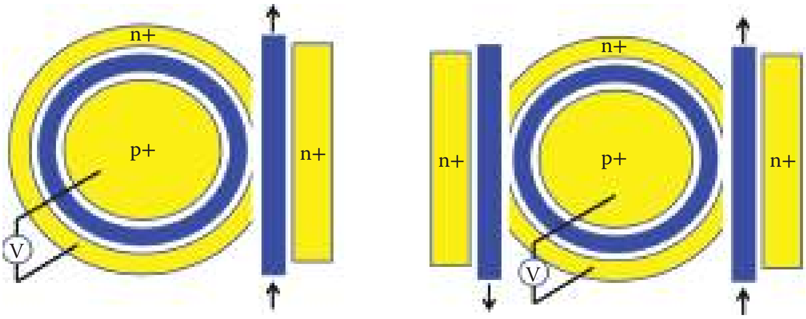


FIGURE 38.16 (Left) A silicon ring resonator modulator. The waveguides in blue (dark) are surrounded by electrodes imposing a reverse bias on the p-n junction (depletion-mode modulator). This design is a phase modulator. (Right) An amplitude modulator.

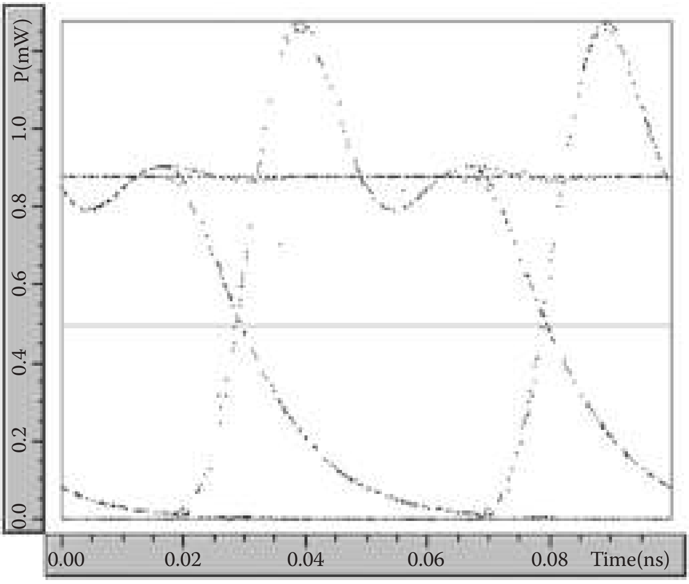


FIGURE 38.17 Eye diagram of a 20-μm-diameter silicon ring resonator modulator. The overshoot is caused by the high Q of the cavity.

References

1. Lefevre KR, Anwar AFM (1997) Electron escape time from single quantum wells. *IEEE J Quantum Electron*. 33: 187–191.
2. Hojfeldt S (2002) Modeling of carrier dynamics in electroabsorption modulators. PhD Thesis. Technical University of Denmark
3. Nikolaev V, Avrutin E (2003) Photocarrier escape time in quantum-well light-absorbing devices: Effects of electric field and well parameters. *IEEE J Quantum Electron*. 16: 24–26.
4. van Eisden J et al. (2007) Modulation properties of VCSEL with intracavity modulator. *Proc SPIE*. 6484: 64840A.
5. Piprek J (2003) *Semiconductor Optoelectronic Devices: Introduction to Physics and Simulations*. Academic Press London
6. Chao CYP, Chuang SL (1993) Momentum-space solution of exciton excited states and heavy-hole-light-hole mixing in quantum wells. *Phys Rev B*. 48: 8210–8221.
7. Mares PJ, Chuang SL (1993) Modeling of self-electro-optic-effect devices. *J Appl Phys*. 74: 1388–1397.
8. Alloatti L et al. (2011) 42.7 Gbit/s electro-optic modulator in silicon technology. *Opt Express*. 19: 11841–11851.
9. Cho HS, Prucnal PR (1989) Effect of parameter variation in the performance of GaAs/AlGaAs multiple-quantum-well electroabsorption modulator. *IEEE J Quantum Electron*. 25: 1682–1690.
10. Kuo YH et al. (2006) Quantum-confined stark effect in Ge–SiGe quantum wells on Si for optical modulators. *IEEE J Sel Topics Quantum Electron*. 12: 1503–1513.
11. Alferness RC, Joyner CH, Buhl LL, Korotky SK (1983) High-speed travelling-wave directional coupler switch/modulator for l = 1.32 μm. *IEEE J Quantum Electron*. 19: 1339–1341.
12. Kubota K, Noda J, Mikami O (1980) Traveling wave optical modulator using a directional coupler $LiNbO_3$ waveguide. *IEEE J Quantum Electron*. 16: 754–760.

13. Kondo J et al. (2002) 40-Gb/s X-cut $LiNbO_3$ optical modulator with two-step back-slot structure. *J Lightwave Technol.* 20: 2110–2114.
14. PicWave - optoelectronic simulator, ver. 5. www.photond.com/products/picwave.htm.
15. Hu YT et al. (2014) Broadband 10 Gb/s graphene electro-absorption modulator on silicon for chip-level optical interconnects. In Proceedings of 2014 IEEE Electron Devices Meeting (IEDM), San Francisco, CA.
16. Dong P et al. (2009) Low Vpp, ultralow-energy, compact, high-speed silicon electro-optic modulator. *Opt Express*. 17: 22484–22490.
17. Lipson M (2006) Compact electro-optic modulators on a silicon chip. *IEEE J Sel Topics Quantum Electron*. 12: 1520–1526.
18. Liu A et al. (2007) High-speed optical modulation based on carrier depletion in a silicon waveguide. *Opt Express*. 15: 660–668.
19. Thompson D et al. (2012) 50-Gb/s silicon optical modulator. *IEEE Phot Tech Lett*. 24: 234–236.
20. Zhou GR et al. (2008) Effect of carrier lifetime on forward-biased silicon Mach-Zehnder modulators. *Opt Express*. 16: 5218–5226.
21. Manipatruni S, Xu Q, Schmidt B, Shakya I, Lipson M (2007) High speed carrier injection 18 Gb/s silicon micro-ring electro-optic modulator. In Proceedings of 2007 Lasers and Electro-Optics Society, Lake Buena Vista, FL (LEOS 2007), 537–538.
22. Xu Q, Manipatruni S, Schmidt B, Shakya J, Lipson M (2007) 12.5 Gbit/s carrier-injection-based silicon microring silicon modulators. *Opt Express*. 15: 430–436.
23. Abraham A, Olivier S, Marris-Morini D, Vivien L (2014) Evaluation of the performances of a silicon optical modulator based on a silicon-oxide-silicon capacitor. In Proceedings of 2014 IEEE 11th International Conference on Group IV Photonics, 3–4.
24. Gardes FY, Tsakmakidis KL, Thomson D, Reed GT, Mashanovich GZ, Hess O (2007) Micrometer size polarisation independent depletion-type photonic modulator in Silicon on insulator. *Opt Express*. 15: 5879–5884.
25. Jeong U, Han D, Lee D, Lee K, Kim J, Park J (2014) A broadband silicon electro-absorption modulator (EAM) using a Schottky diode. *Proc SPIE*. 8988: 89881M.
26. Liu M, Yin X, Zhang X (2012) Double-layer graphene optical modulator. *Nano Lett*. 12(3): 1482–1485.
27. Phare CT, Lee YD, Cardenas J, Lipson M (2015) Graphene electro-optic modulator with 30 GHz bandwidth. *Nat Photonics*. 9: 511–514.
28. Tanabe T, Nishiguchi K, Kuramochi E, Notomi M (2009) Low power and fast electro-optic silicon modulator with lateral p-i-n embedded photonic crystal nanocavity. *Opt Express*. 17: 22505–22513.
29. Chao CYP, Chuang SL (1992) Spin-orbit-coupling effects on the valence-band structure of strained semiconductor quantum wells. *Phys Rev B*. 46: 4110–4122.
30. Lever L et al. (2010) Design of Ge-SiGe quantum-confined stark effect electroabsorption heterostructures for CMOS compatible photonics. *J Lightwave Technol*. 28: 3273–3281.
31. Weiner JS et al. (1985) Strong polarization-sensitive electroabsorption in GaAs/AlGaAs quantum well waveguides. *Appl Phys Lett*. 43: 1148–1150.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

VIII

Solar Cells

39 Solar Cell Fundamentals *Matthias Müller* 383
Introduction • The PV System • The Solar Spectrum • Real Operating Conditions of Nonconcentrated Solar Cells and Modules • Solar Cell Optics • Conversion into Electrical Energy • Modeling of Ideal *I*–*V* Characteristics • Nonideal *I*–*V* Characteristics • Numerical Simulation and Models for Crystalline Silicon Solar Cells • Real Working Behavior of Solar Cells • Conclusion and Outlook

40 Multijunction Solar Cells *Matthew Wilkins and Karin Hinzer* 415
Introduction • Device Model and Simulation Workflow • Mesh Discretization • Optics • Drift-Diffusion Calculation • Current–Voltage Calculation • QE Calculation • Photon Recycling and Luminescent Coupling • Summary

41 Nanostructure Solar Cells *Urs Aeberhard* 441
Introduction: Nanostructures in Photovoltaics • Physical Mechanisms of NSSC Device Operation • Modeling of NSSC Device Characteristics • Case Study: Single Quantum Well p-i-n SC • Summary

42 Nanowire Solar Cells: Electro-Optical Performance *Bernd Witzigmann* 475
Introduction • Efficiency of Nanostructured Solar Cells • Computational Models • Light Capture and Emission • Electronic Properties • Summary

43 Thin-Film Solar Cells *Matthias Auf der Maur, Tim Albes, and Alessio Gagliardi* 497
Introduction • Technology Overview • Optical Modeling • Transport Modeling • Specific Modeling Issues

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

39

Solar Cell Fundamentals

39.1 Introduction 383
39.2 The PV System 386
39.3 The Solar Spectrum 387
39.4 Real Operating Conditions of Non-concentrated Solar Cells and Modules 388
Irradiation • Temperature
39.5 Solar Cell Optics 392
Photogenerated Current
39.6 Conversion into Electrical Energy 395
39.7 Modeling of Ideal *I*–*V* Characteristics 396
I–*V* Parameters and Quantum Efficiency • Maximum Efficiency of Ideal Single Junction Solar Cells
39.8 Nonideal *I*–*V* Characteristics 399
Series and Parallel Resistance • Recombination
39.9 Numerical Simulation and Models for Crystalline Silicon Solar Cells 402
Numerical Simulation • Models for Crystalline Silicon Solar Cells
39.10 Real Working Behavior of Solar Cells 406
Light Intensity Behavior • Temperature Behavior • Energy Yield
39.11 Conclusion and Outlook 410

Matthias Müller

39.1 Introduction

Solar cells convert the energy of the sun light to electricity. It is an optical to electrical energy conversion. In principle, they make use of the photovoltaic (PV) effect, i.e., the generation of a potential difference between two electrodes under illumination. It was firstly observed by A. E. Becquerel (Becquerel, 1839) and practically implemented in a solar cell at Bell Laboratories in the United States (Chapin et al., 1954). Today it is believed that PV energy generation will be a part of the world's base of energy supply established in the twenty-first century. Conservative estimates by the International Energy Agency (IEA) show a share of the world electrical energy generation of about 1% nowadays to about 16% in 2050 (IEA, 2014). This would correspond to a yearly energy generation of about 6300 TWh compared to slightly more than 200 TWh today.

Looking back on the past decades, we can see that the development in solar research and industry was very fast and economically successful, which enabled such a bright perspective for solar electricity. The price for solar modules reduced with increasing manufactured PV module capacity following a power-law relation: the so called PV learning curve could be established which is shown in Figure 39.1.

For example, the price of solar modules decreased by about 80% between the years 2008 and 2015. About 1000 GW module shipments may be reached in 2024 for 10% market growth, or even in 2021 for 30% market growth, which corresponded to the market development in the past 15 years. Those cost reductions

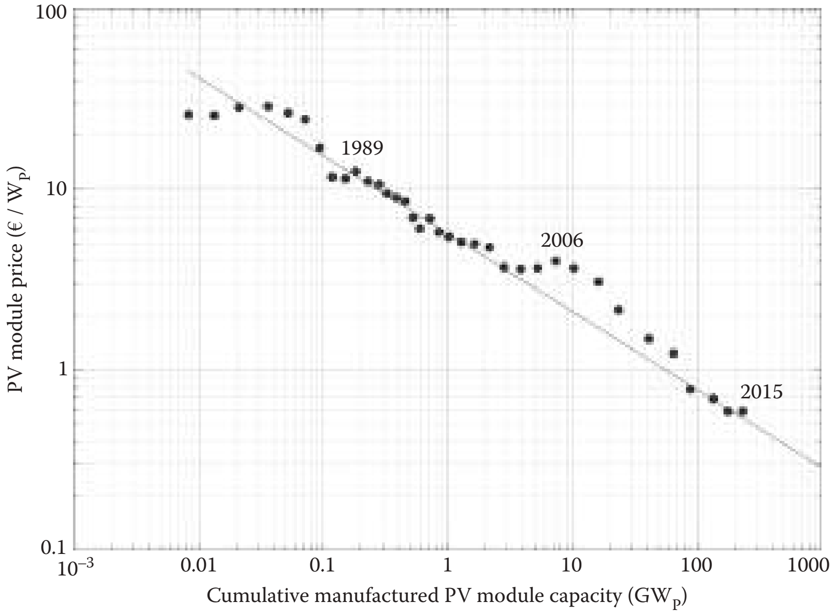


FIGURE 39.1 PV learning curve (Fraunhofer ISE, 2013; IEA, 2014; pvXchange, 2016) and a fit of a power law model to the years 1990–2000 and 2012–2015. In between the years 2000–2010, there was a shortage on raw silicon which slowed down the price reduction.

result largely from economies of scale and technological improvements. In 2015, the limit of 200 GW-shipped solar modules was exceeded, and the same volume is expected to be the future world production capacity to maintain the global solar energy supply in the twenty-first century starting from 2025 (IEA, 2014).

The current solar cell production is dominated by wafer-based crystalline silicon (c-Si) solar cells, which make about 90% of the market (ITRPV, 2015) and only 10% are thin-film technologies based mainly on cadmium telluride (CdTe) and copper indium gallium diselenide (CIGS). That is why the author focuses on the c-Si technology within this chapter.

Looking closely on c-Si, the shares of different wafer technologies that are applied nowadays are shown in the upper graph of Figure 39.2 (ITRPV, 2015), where mainly boron-doped p-type material is used (95%). The cast silicon materials multicrystalline (mc) and high-performance multicrystalline (HPmc-Si) currently make up to over 65%, whereas monocrystalline, mainly Czochralski (Cz), make 35%.

The share of different solar cell technologies is shown in the lower graph of Figure 39.2, where the main product currently is the full-area back-surface-field (BSF) solar cell (Mandelkorn and Lamneck, 1972), whereas the passivated emitter and rear cell (PERC) technology (Blakers et al., 1989) shows increasing market shares in the near future (ITRPV, 2015). The passivated emitter and rear totally diffused solar cell (PERT) (Zhao et al., 1999), the Si-heterojunction (SHJ) solar cell (Tanaka et al., 1992) and the back contacted solar cells—mainly interdigitated back contact solar cells (IBC) (Lammert and Schwartz, 1977)—are expected to play a minor role in the coming years but may be candidates for future high-efficiency solar cells with higher market shares (ITRPV, 2015). The structure of these different solar cells is shown in Figure 39.3.

This chapter gives a general overview about the working principle and basic models of solar cells. The view will be extended to the energy yield delivered by solar cells under real world conditions, i.e., the view will be extended to solar cells working in solar modules and the corresponding cell demands. Examples will

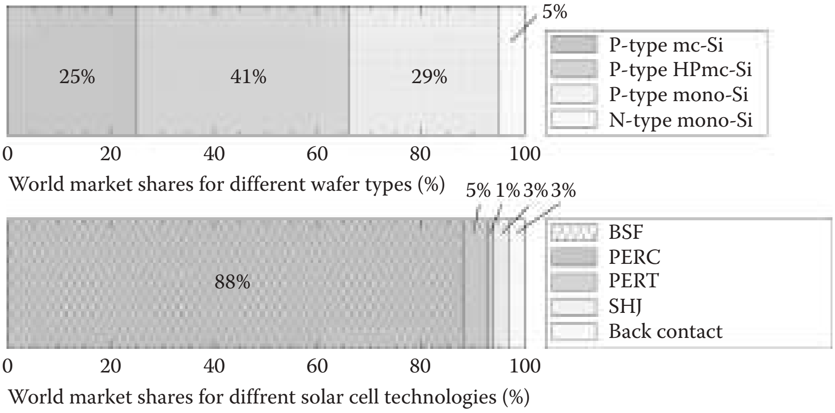


FIGURE 39.2 Share of different silicon wafer technologies for solar cell application and share of different solar cell technologies. [BSF, back-surface-field; PERT, passivated emitter and rear totally diffused; PERC, passivated emitter and rear cell; SHJ, Si-heterojunction. (ITRPV, 2015)]

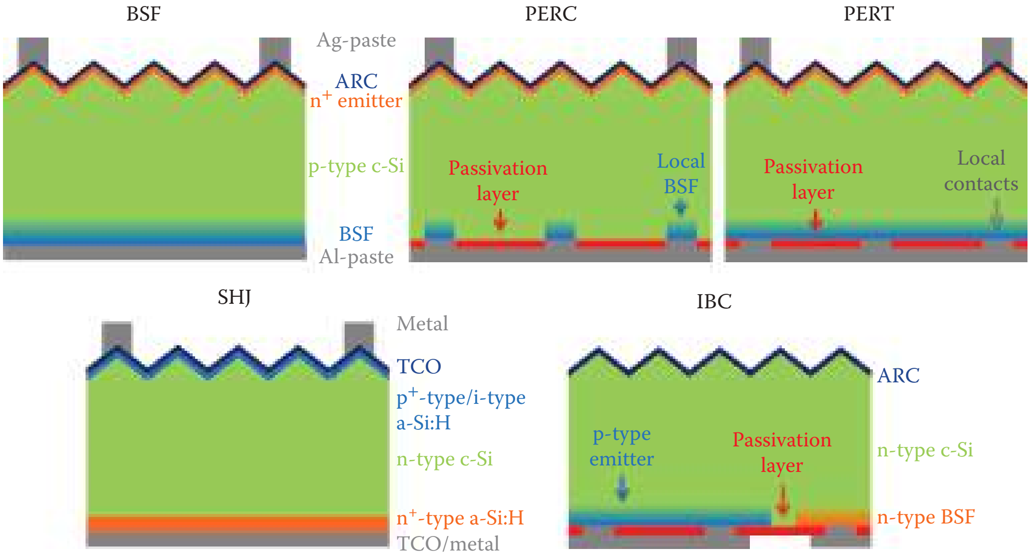

BSF
PERC
PERT
Ag-paste
ARC
n^+ emitter
p-type c-Si
BSF
Al-paste
Passivation layer
Local BSF
Passivation layer
Local contacts
SHJ
Metal
TCO
p^+-type/i-type a-Si:H
n-type c-Si
n^+-type a-Si:H
TCO/metal
IBC
ARC
p-type emitter
Passivation layer
n-type c-Si
n-type BSF

FIGURE 39.3 Schematic of typical solar cell structures in their two-dimensional cross section representation where antireflection coating (ARC) and transparent conductive oxide (TCO) are used. (BSF, back-surface-field; PERC, passivated emitter and rear cell; PERT, passivated emitter and rear totally diffused; SHJ, Si-heterojunction; IBC; interdigitated back contact solar cell.)

be presented considering only nonconcentrated crystalline silicon-based solar cells for 1-sun application. For additional information and explanations which are not given in this chapter, the author would like to refer to the notable literature by Wuerfel (2005), Green (1982), and Goetzberger et al. (1998). Additionally, the author would like to refer to the free online resources for PV engineers and scientists PVEducation (PVEDUCATION.ORG, 2016) and PV Lighthouse (PV Lighthouse, 2016).

39.2 The PV System

The solar cell is only a part, even so the functional device, of a full PV system. Such systems may be grid-connected as shown in Figure 39.4, or stand-alone systems, which means that all produced electricity is consumed locally. Those stand-alone PV systems are good options for rural areas of developing countries and special applications (e.g., parking meter). However, in comparison, grid-connected systems produce nowadays and will produce in the future a higher share of electricity. A grid-connected system consists of solar modules, which contain the solar cells. These modules are typically connected in series (called module string) to an inverter. Those inverters first do Maximum Power Point Tracking (MPPT) of the module strings and secondly convert the direct current produced by solar cells in alternating current. The inverter is connected to the load and to the electricity grid via an AC isolation switch and via an import/export electricity meter (compare Figure 39.4).

A current trend for grid-connected systems is the combination with storage systems. It increases energy self-consumption, as it is expected that during noon an excess of solar energy is produced. This excess energy can then be stored and used during times without irradiation. The trend will be enhanced, as the price for electricity is decreasing more and more at noon time when more solar power plants are connected to the grid. So the storage system will help for grid stabilization as PV electricity is a fluctuating energy challenging the electricity grid in the future. Those battery PV systems are controlled by an energy management system (EMS) which regulates production, load, storage, and grid connection (compare Figure 39.4).

Looking in more detail at the structure of a solar module as shown in Figure 39.5, one can see that solar cells are basically embedded between glass and plastics.

A typical solar module contains 60 solar cells on an area of approximately 1.7 m^2. They are connected in series by tabbing ribbons to strings of 20 cells while bus ribbons interconnect the three cell strings. Those cell strings are connected in parallel by a bypass diode which accounts for lower power losses during shading conditions. Finally, the bus ribbons are connected to the junction box, where cables with connectors go out.

Solar modules are made to produce electricity for at least 20–30 years; hence all used materials have to withstand changing environmental conditions. Solar glass, that in most cases is low-iron rolled glass or float glass, is situated on the side facing the sky and has to withstand rain, hail and snow. Besides, the front glass guarantees mechanical stability against wind and snow load. It may be manufactured with an antireflection coating (ARC) and/or textured, i.e., with a structured front surface for improved light trapping. The rear side is typically covered with polyethylene terephthalate (PET) polymer backsheet which features moisture resistance and electrical isolation. If replaced by another glass to produce a glass-glass module, the solar

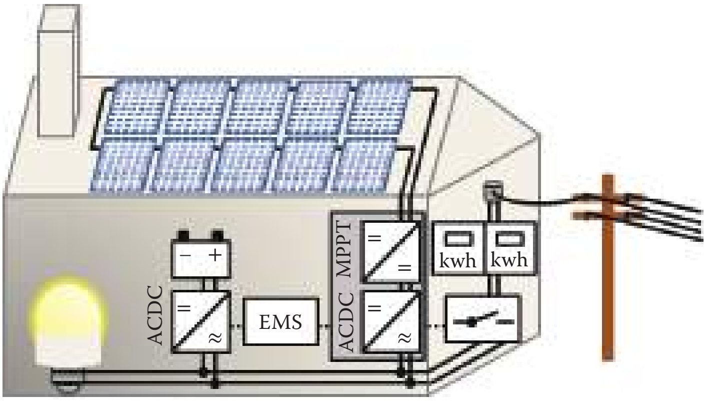


FIGURE 39.4 Schematic of a PV system with solar modules, cables, inverter, battery energy managements system (EMS), battery, load, AC isolation switch, and import/export electricity meter.

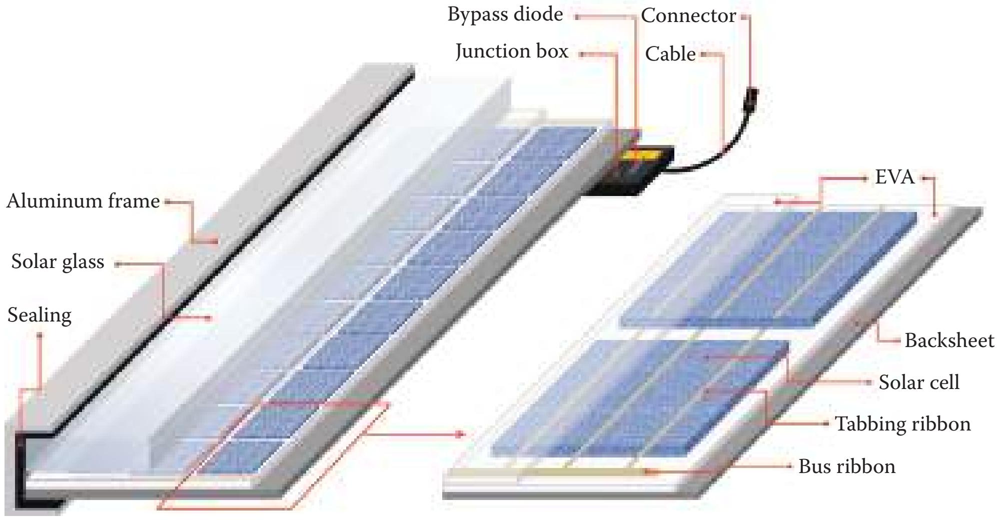


FIGURE 39.5 Structure of a solar module. EVA, ethylenvinylacetate.

module lifetime increases by about 5–10 years. Finally, the solar cells are embedded between the solar glass and the backsheet in Ethylenvinylacetate (EVA) as an encapsulant. It is a polymer which cross-links at about 150°C during manufacturing. Solar modules are sealed at the edges with silicone and an aluminum frame.

39.3 The Solar Spectrum

Our sun is basically a fusion power plant which emits electromagnetic radiation from its surface, called photosphere, into the solar system. It has a spectrum similar to a blackbody with a temperature of 5777 K, which is shown in Figure 39.6.

The light intensity reaching the planet earth is reduced by a factor of 2.17×10^{-5} due to geometrical considerations. The factor is the square of the suns radius divided by the square of the distance sun–earth. The resulting spectrum above the earth's atmosphere is shown in Figure 39.6 denoted as AM0, where AM stands for air mass index and the "zero" denotes that the light doesn't travel through the earth's atmosphere. This extraterrestrial solar spectrum AM0 has an irradiance of 1366.1 $\mathrm{W/m^2}$ as defined within (ASTM E490-00a, 2014), which is called the solar constant. The latest measured value is 1360.8 $\mathrm{W/m^2}$ (Kopp and Lean, 2011). This spectrum is relevant for solar cell power measurements for space applications.

Light traveling through the atmosphere faces absorption, scattering, and reflection effects which change the spectrum and the irradiance as shown schematically in Figure 39.7.

Absorption in the earth's atmosphere takes place in its molecules: ozone, oxygen, water vapor, carbon dioxide, nitrogen dioxide. Scattering in the atmosphere takes place at particles that are smaller than the wavelength of the light (Rayleigh scattering, e.g., N_2, O_2 molecules) and which are larger than the wavelength of the light (Mie scattering, e.g., aerosol particles, cloud droplets). The third mechanism is reflection of the light in the atmosphere, mainly at the clouds, and finally at the ground. This introduces a direct, a diffuse, and a ground-reflected component of irradiance on the module plane. The diffuse component describes all light coming from any region of the sky, mainly Rayleigh-scattered light which we see as blue sky on sunny days.

The standard spectrum AM1.5g is the result of light passing the earth's atmosphere approximately on a very sunny day and it is shown in Figure 39.6. The air mass index AM defines the direct optical path length through the atmosphere of the earth. Thus, AM1.5 is one and a half times the path length; "g" stands for

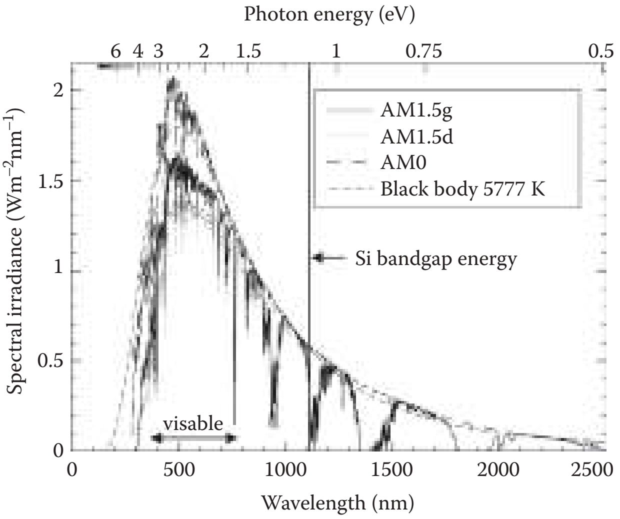


FIGURE 39.6 Comparison of different spectra: blackbody spectrum with a temperature of 5777 K, extraterrestrial spectrum AM0 (ASTM E490-00a, 2014), reference spectra AM1.5g and AM1.5d (ASTM G173-03, 2012) for global and direct solar spectral irradiance at AM1.5, respectively.

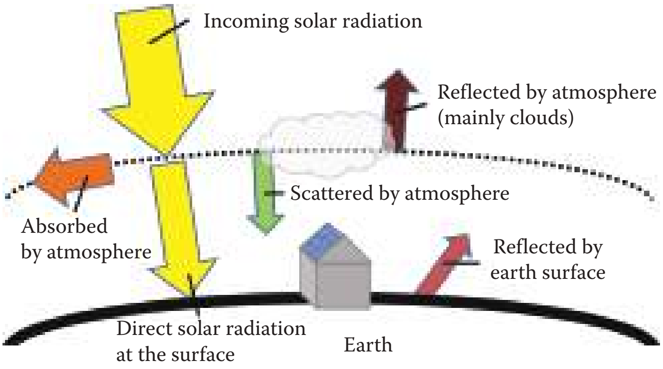

Incoming solar radiation
Reflected by atmosphere
(mainly clouds)
Scattered by atmosphere
Absorbed
by atmosphere
Reflected by
earth surface
Direct solar radiation
at the surface
Earth

FIGURE 39.7 Schematic of solar radiation traveling through the earth's atmosphere.

global, i.e., with direct and diffuse light. The spectrum AM1.5d, which is also shown in Figure 39.6, is only the direct component of the AM1.5g and is used to measure solar cells for concentrated PV (CPV) applications.

39.4 Real Operating Conditions of Non-concentrated Solar Cells and Modules

The output power of a solar system, typically expressed in "Watt-peak" (W_P), and the efficiency of a solar system or solar components like solar cells or solar modules are determined by a standardized measurement of the current–voltage (I-V) characteristic under the "Standard Test Conditions" (STCs), i.e., under a light intensity of 1000 W/m^2 (one-sun), an ambient temperature of 25°C and the standardized spectrum AM1.5g (ASTM G173-03, 2012) which may altogether correspond approximately to a very sunny summer day in the morning on a 37° sun facing tilted surface with low ambient temperature.

Unfortunately, those environmental conditions do not occur very often. During the year and day-time deviations from the reference efficiency under STC are observed and the efficiency of the solar system changes due to a severe number of effects:

- Efficiency decrease due to module heat generation above 25°C
- Efficiency decrease due to low light/partial load condition (e.g., cloudy weather)
- Reflection losses due to incident light under increased angle of incidence (AOI) between the module and the sun
- Decreased or increased efficiency due to the change of the spectrum of the incident light
- Mismatch loss due to interconnection of cells in a module and solar modules in a PV system with different I-V characteristics
- Losses due to soiling and snow
- Losses due to shading
- Losses due to interconnection and (bypass) diode failures
- MPPT failure of the MPP tracker
- Conversion losses and self-consumption of the inverter (stand-by loss)
- Energy yield loss due to differences in real, measured and labeled module power
- Degradation, i.e., efficiency decrease over time due to environmental stress load

The effect of efficiency degradation of solar modules happens mainly due to mechanical stress and moisture resulting in corrosion, discoloration, delamination, breakage, and cracking cells (Jordan and Kurtz, 2013), but also due to solar cell degradation resulting from the formation of defects within the silicon wafers under illumination, e.g., the boron-oxygen-related defect (Schmidt et al., 1997). Another degradation mechanism may occur in solar modules under high potential (Pingel et al., 2010), which is called potential-induced degradation (PID). PID results in a low shunt resistance of the module and can decrease the output power down to zero. Nevertheless, the solar module lifetime exceeds 20–25 years maintaining typically at least 80% of its initial efficiency or for a glass-glass module even more than 30 years. Thus a solar module is one of the longest lasting products for electricity generation that does not need maintenance.

A parameter that considers all the real world influences on a PV system is the performance ratio PR, as defined by

$$PR = \frac{\text{real energy yield}}{\text{ideal energy yield}} = \frac{W_{\text{SP}}}{H_{\text{sun}} \times A_P \times \eta_{\text{STC}}} \tag{39.1}$$

PR is defined as the actual energy yield of a PV system W_{SP} related to the incoming sun irradiation H_{sun} on the array area A_{P} converted by the solar system efficiency under STC η_{STC} over a time interval of months/years. Also a wide spread parameter for PV system assessment is the specific or final system yield Y_{f} in kWh/kWp, as defined by Equation 39.2, which allows to compare the energy production for different systems at similar locations. Typical values are in the range of 500–2500 kWh/kWp.

$$Y_{\text{f}} = \frac{W_{\text{SP}}}{P_{\text{syst}}} \tag{39.2}$$

39.4.1 Irradiation

The frequency with which STC applies in real world is negligible. For a better understanding of which environmental conditions are more typical for a solar system, exemplarily three geographical locations are chosen for further visualization within this chapter, which are situated in middle-Europe (Halle, Germany), southern-Europe (Almeria, Spain) and in tropical climate close to the Equator (Brazzaville, Republic of the Congo).

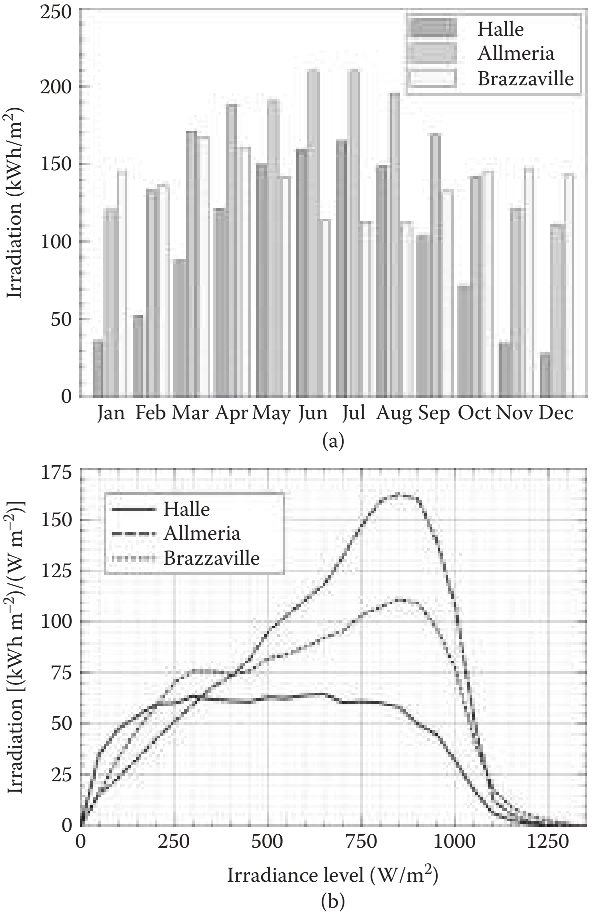


FIGURE 39.8 (a) Comparison of different irradiation conditions (Meteotest, 2008) per month, and lower graph versus in-plane irradiance level. (b) A histogram in steps of 50 W/m^2 per irradiance level bin where the frequency is multiplied by the mean irradiance of each bin.

An overview about the applying in-plane irradiation conditions, i.e., irradiation on the tilted surface, are shown in the upper graph of Figure 39.8 per month and in the lower graph versus irradiance level.

The geometric considerations concerning sun light hitting a tilted surface may be found in references (Duffie and Beckman, 2013; Quaschning, 2005). The yearly solar energy yield, i.e., in-plane irradiation, is 1159 kWh/m^2 for Halle, 1961 kWh/m^2 for Almeria and 1657 kWh/m^2 for Brazzaville for tilted surfaces having tilted angles of 30° for Halle and Almeria, and 10° for Brazzaville, respectively.

Meteorological data may be found worldwide at NASA Surface Meteorology and Solar Energy Data Set (NASA, 2016) and using the software Meteonorm (Meteotest, 2008), for EU at PVGIS EU Joint Research Centre (JRC, 2016), and for US at National Renewable Energy Laboratory (NREL, 2016).

39.4.2 Temperature

The illumination of solar modules does not only generate electricity, but also heat. The main reasons are the heat generation of the solar cells due to thermalization of charge carriers and parasitic absorption within the module of photons with higher wavelength than the bandgap of the semiconductor. A typical solar cell operating temperature increases linearly with the in-plane irradiance G_I above ambient temperature.

As commercial solar modules have a very similar structure as shown in Figure 39.5, they have comparable thermal properties and heat in a similar way. The thermal properties are measured in a standardized way determining the "Nominal Operating Cell Temperature" (NOCT) (IEC 61215, 2005) of the solar module under 800 W/m^2 light intensity, 20°C ambient temperature, AM1.5g spectrum, and 1 m/s wind velocity, which results in typical NOCT = 45°C ± 3°C.

For a better understanding of the system efficiency losses due to different operating temperatures deviating from 25°C (STC), an annual average operating temperature T_{ann} is calculated from the meteorological data set shown in Figure 39.8. The solar cell operating temperature T_{Cell} is calculated by (Ross and Smokler, 1986)

$$T_{\text{Cell}} = T_{\text{amb}} + (\text{NOCT} - 20°\text{C})\frac{G_I}{800\frac{\text{W}}{\text{m}^2}} \tag{39.3}$$

This model represents a standard roof top PV system installation. For an open-rack mounting condition lower T_{Cell} can be expected whereas for building integrated installations higher T_{Cell} is expected (Garcia and Balenzategui, 2004).

Assuming that a PV system only depends on irradiation and different ambient temperatures, the annual average (irradiance weighted) cell operating temperature T_{ann} is calculated by

$$T_{\text{ann}} = \frac{\int^{1a} P(t) \times T_{\text{Cell}}(t)\text{d}t}{\int^{1a} P(t)\text{d}t} = \frac{1}{W_{\text{SP}}} \int^{1a} P(t) \times T_{\text{Cell}}(t)\text{d}t \tag{39.4}$$

Exemplarily, the monthly average operating temperature is calculated and shown in Figure 39.9, assuming a *NOCT* of 45°C and applying Equations 39.3 and 39.4 to the meteorological data sets.

The annual average cell operating temperature for the selected locations are those of 32.0°C for Halle, 42.9°C for Almeria, and 47.7°C for Brazzaville, respectively. Hence Almeria and Brazzaville are close to the *NOCT* = 45°C, which confirms that NOCT is a relevant environmental condition for many locations. Assuming that the relative temperature coefficient of the efficiency of a typical c-Si solar cell is $-0.4\%_{\text{rel}}/°\text{C}$, an energy yield loss compared to STC may be calculated based on the annual average cell operating temperatures. For Halle $(32.0°\text{C} - 25.0°\text{C}) \times -0.4\frac{\%_{\text{rel}}}{°\text{C}} = -2.8\%_{\text{rel}}$ may be expected, for Almeria $-7.2\%_{\text{rel}}$, and for Brazzaville $-9.1\%_{\text{rel}}$, respectively.

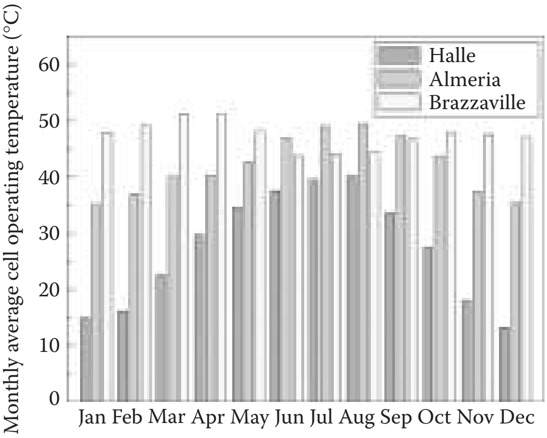


FIGURE 39.9 Monthly average (irradiance weighted) cell operating temperature per month for the locations Halle, Almeria, and Brazzaville.

39.5 Solar Cell Optics

In the field and under real operating conditions, solar cell's optical behavior is mainly determined by the optics of a solar module. In more detail, it strongly depends on the optics of the first medium and its surface structure facing the air (Winter et al., 2014), which is typically the solar glass. During year and day time, sun light appears under different oblique incidence to the module. An angle of incidence (AOI) θ can be defined, which is the angle between the sun's rays and the module normal. Light reaching the module's surface at an angle is spread out over a larger area, thus reduces the (direct) irradiance on the tilted plane following a cosine relation, as shown in Figure 39.10, described by:

$$G_{\mathrm{direct}}(\theta) = G_{\mathrm{direct}}\left(0^\circ\right) \times \cos(\theta) \tag{39.5}$$

The first interface air–glass may be simply described by the Fresnel equations and Beer–Lambert–Bouguer and Snell's laws (Duffie and Beckman, 2013), which are in general wavelength dependent. The angle of refraction θ_r is therefore

$$\theta_r = \arcsin\left(\frac{n_1}{n_2}\sin(\theta)\right) \tag{39.6}$$

where $n_1 = 1$ and $n_2 = 1.52$ are the refractive indices of air and solar glass, respectively. The angle-dependent transmission of such system is approximated by Equation 39.7 considering both reflective losses at the interface and absorption losses within the glass (Duffie and Beckman, 2013).

$$\left(A_{\mathrm{opt}}T_{\mathrm{opt}}\right)(\theta) = A_{\mathrm{opt}}\left(1 - R_{\mathrm{opt}}\right) = e^{-(KL/\cos\theta_r)}\left[1 - \frac{1}{2}\left(\frac{\left(\sin\left(\theta_r - \theta\right)\right)^2}{\left(\sin\left(\theta_r + \theta\right)\right)^2} + \frac{\left(\tan\left(\theta_r - \theta\right)\right)^2}{\left(\tan\left(\theta_r + \theta\right)\right)^2}\right)\right] \tag{39.7}$$

where $K = 4\ \mathrm{m}^{-1}$ may be assumed for the extinction coefficient and $L = 2$ mm for the thickness of the glass (Duffie and Beckman, 2013).

The calculation of the product of optical absorption and transmission through the solar glass ($A_{\mathrm{opt}}T_{\mathrm{opt}}$) approximates here the angle-dependent losses of the solar module, which is defined as the incident angle modifier (IAM). However, *IAM* accounts for all angular and spectral losses due to increased absorption

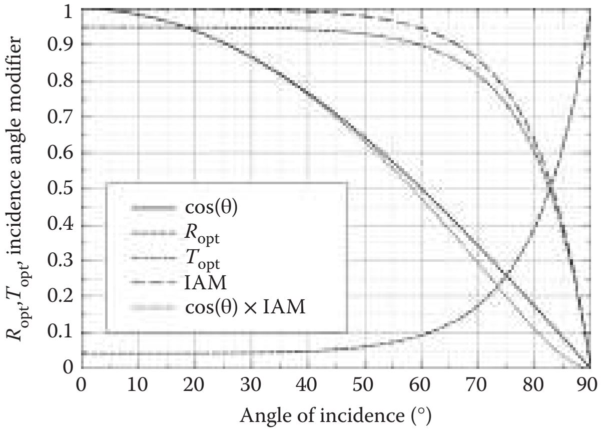


FIGURE 39.10 Cosine, optical reflection, and transmission ($n_1 = 1$ for air and $n_2 = 1.52$ for glass), incidence angle modifier (IAM), angle-dependent losses of a solar module versus angle of incidence.

within the EVA and for geometrically different, wavelength-dependent light propagation under oblique incident, i.e., multiple reflections at the cell or the backsheet. *IAM*, also shown in Figure 39.10, is used to correct the incoming direct irradiance G_{direct} on the tilted module plane as derived in Equation 39.8. The indirect irradiance component may be corrected in a similar approach.

$$G_{\text{direct}}(\theta) = G_{\text{direct}} \times \cos\theta \times IAM(\theta) = G_{\text{direct}} \times \cos\theta \times \frac{\left(A_{\text{opt}}T_{\text{opt}}\right)(\theta)}{\left(A_{\text{opt}}T_{\text{opt}}\right)(0^\circ)} \tag{39.8}$$

Calculations for the angular-dependent irradiation losses over 1 year using Equations 39.5 through 39.8 and the meteorological data sets, as shown in Figure 39.8, result in $-1.3\%_{\text{rel}}$ for Halle, $-1.3\%_{\text{rel}}$ for Almeria, and $-0.9\%_{\text{rel}}$ for Brazzaville.

Latest developments of solar glass lead more and more to the introduction of ARCs and partially deeply structured (textured) glass, which improves the module optics.

The same concepts are also applied for solar cell optic optimization. ARCs for solar cells, as drawn in Figure 39.11, are typically applied as a single layer of silicon-nitride SiN_X films of about 75 nm thickness and a refractive index of about 2, whereas its optimum is slightly lower against air and slightly higher if optimized for the use in a module.

Those thin film layers may be well modeled by the transfer matrix method, which considers oblique incidences, absorbing media, coherent light, and even assemblies of thin films (e.g., double ARC) (Macleod, 2010). The second optimization strategy is to apply a texture on the front side of the solar cell, which increases the chance that reflected light reaches again the solar cell surface a second or even more times as shown in Figure 39.11. Most relevant textures for industrial application are random pyramids for monocrystalline (Cz) cells and isotexture, which forms bowl-like structures, for mc cells. Both are produced by wet chemical etching. Additional benefits of texturing arise because light is coupled into the silicon under an angle. First, this increases the chance of a photon to be absorbed close to the surface, second the path length for absorption is increased within the semiconductor and thirdly such photons may reach the rear side under the critical angle of total internal reflection. This internal light trapping may be even improved further if the rear side has properties of a back reflector, as reflected light of wavelengths close to the bandgap energy gets the chance to be absorbed on the way there and back (at least double length) within the semiconductor. Those back reflectors are, e.g., full area Aluminum layers for a BSF solar cell or dielectric layer stacks for PERC solar cells. Note that semiconductors having a direct bandgap (e.g., GaAs, CIGS, CdTe) may be fabricated with much smaller absorber thickness compared to, e.g., silicon with an indirect bandgap, and are called thin film solar cells.

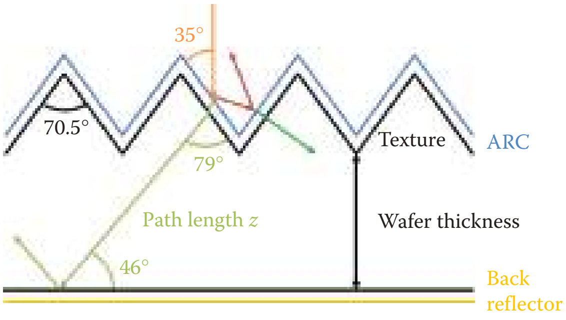


FIGURE 39.11 Schematic of light trapping within a solar cell with antireflection coating (ARC) and a random pyramid texture. Note that the schematic is not to scale, as a typical side length of such pyramid is about 5 μm and a wafer thickness is about 200μm.

39.5.1 Photogenerated Current

A way to model light-induced photocurrent of solar cells, as shown in Figure 39.11, is to use ray-tracing, which is a method to calculate the path of a photon through the solar cell considering its geometry. The light propagation may be calculated at surfaces and interfaces as necessary, e.g., using Fresnel equations. Raytracing software for solar cells are, e.g., Sunrays (Brendel, 1994), TCAD Sentaurus Device (Synopsys, 2016) and for solar modules, e.g., Daidalos (Holst et al., 2013). Also wide spread in the PV community is the usage of the simulation tool PC1D (Rover et al., 1985), which can also analytically calculate solar cell optics with texture and ARC.

The absorption of photons within the active semiconductor leads to the creation of electron–hole pairs, i.e., to a current generation. This photogenerated current density j_{gen} is not influenced by front contact shading and recombination losses and hence can be treated as the maximum current potentially extracted from the solar cell described by

$$j_{\mathrm{gen}} = -q \int_{0}^{\infty} A_{\mathrm{opt}} \Big[E_{\mathrm{ph}} \big) \, \mathrm{d}j_{\mathrm{ph}}(E_{\mathrm{ph}}) \tag{39.9}$$

where q is the elementary charge, E_{ph} is the photon energy and j_{ph} is the photon flux. Assuming $A_{\mathrm{opt}}(E_{\mathrm{ph}} \geq E_{\mathrm{g}}) = 1$, Equation 39.9 solves to

$$j_{\mathrm{gen}} = -q \int_{E_{\mathrm{g}}}^{\infty} \mathrm{d}j_{\mathrm{ph}}(E_{\mathrm{ph}}) \tag{39.10}$$

The optical absorption in silicon is described by its absorption coefficient α which is shown in Figure 39.12.

The maximal achievable photogenerated current density for c-Si is about 46 mA/cm^2 for AM1.5g reference spectrum. Photons with more energy than the bandgap energy E_{g} quickly loose this "extra" energy to the lattice atoms in a process called thermalization. Photons with lower energy than E_{g} are not able to excite any electron from valence to conduction band and are not absorbed under the creation of an electron–hole pair.

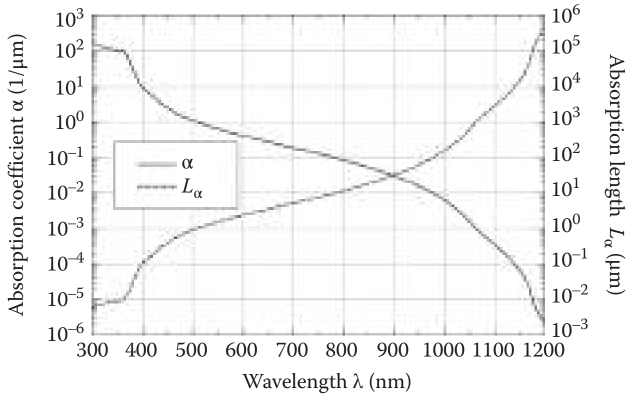


FIGURE 39.12 Absorption coefficient α and absorption length $L_\alpha = 1/\alpha$ for silicon at 300 K.

39.6 Conversion into Electrical Energy

The conversion of the optical energy of the sun light into electrical energy can be regarded in general in terms of particles in a two-level energy system. The particles are excited from their ground state to an excited state by illumination and carry the energy of the photon in the excited state. In general, the excited particles need to be extracted from the absorber into an external electrical circuit; otherwise recombination takes place which is the reverse reaction to excitation, also called generation. This is realized by the usage of selective contacts which only allow the excited particle to flow out at one side of the absorber, preventing the "overflow" of the absorber, which is the driving force. The excited particles flow into the external electrical circuit to perform work and then flow back into the absorber as relaxed particles through the other contact.

A possible realization for selective contacts is to create oppositely situated highly n- and p-type doped regions in the absorbing semiconductor for the respective contact, which is shown in the schematic of an illuminated semiconductor-based solar cell in Figure 39.13. Under illumination-free electrons and holes (excess charge carriers), introduced as excited particles, are generated in the conduction band and valence band, respectively. Thus the quasi-Fermi-levels of electrons E_{Fn} and holes E_{Fp} split in the semiconductor describing the energy distribution of electrons and holes by Fermi–Dirac statistics within the conduction and valence band, respectively. The Fermi levels of the majority carriers join at the surfaces at the respective metal contacts because the surface states exchange electrons very easily with the metal and maintain the same potential. The opposing doping at each contact results in three main effects. First, low minority carrier densities occur at the contacts due to the law of mass action and hence lead to almost zero contact recombination (selective contact). In consequence, only one kind of charge carrier is extracted. Second, it can result in ohmic contact behavior with sufficiently high doping close to the contact, which leads to low resistive losses at the contacts. Third, the creation of p-type and n-type regions in the semiconductor leads to the creation of a p-n junction with a SCR and a built-in potential, also called diffusion voltage, which is typically thought to be a necessity for solar cells, but it is not the "engine" of a solar cell (Wuerfel, 2005). Both carrier types are continuously created under illumination, and, e.g., under short-circuit condition electrons and holes flow out at their respective selective contacts as the reverse direction is blocked for them. Thus an excess of charge carriers is produced and the majority of excess charge carriers do not recombine before they diffuse along their concentration gradients. This creates a charge current through the solar cell. As charge current flows are the result of potential gradients, the behavior of solar cells is described below in this section in terms of potentials in more detail. Under an external forward bias, the difference of E_{Fn} in the n-type region and E_{Fp} in the p-type region correspond to the external voltage V, so that

$$E_{\mathrm{Fn}} - E_{\mathrm{Fp}} = qV \tag{39.11}$$

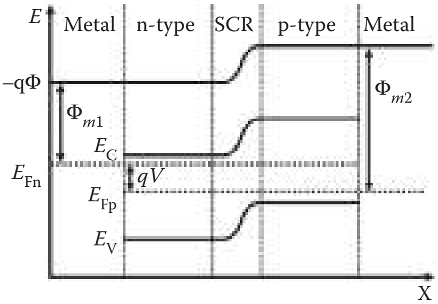


FIGURE 39.13 Schematic of the band diagram of an illuminated solar cell with two selective contacts with their respective metal work functions Φ_m separated by a p-n junction with a space-charge region (SCR) and the electrostatic potential Φ.

where q is the elementary charge. The forward bias reduces the built-in potential of the p-n junction. The driving forces of charge currents in a solar cell are the gradients of the quasi-Fermi-levels E_{Fn}/dx and E_{Fp}/dx, which are the gradients of the electrochemical potentials for electrons and holes, respectively. The electrochemical potential of electrons, which is constant in thermal equilibrium, is determined by the chemical potential $E_{\mathrm{C}} - E_{Fn}$ and the electrostatic potential $-q\Phi$:

$$E_{\mathrm{R}} - E_{Fn} = \left(E_C - E_{Fn}\right) - q\phi \tag{39.12}$$

where E_{R} is the reference potential. The absolute value of E_{R} is not relevant as only gradients of the potentials are of interest. The chemical potential describes the diffusion aspect of carrier transport due to a gradient in the carrier concentration whereas the electrostatic potential describes the force acting on charge carriers due to coulomb interaction and it is the drift aspect of charge carriers transport due to an electric field. Those two transport mechanisms have to be considered together, because diffusing mobile electrons and holes inherently carry a charge.

39.7 Modeling of Ideal *I–V* Characteristics

A PV solar cell can in first-order approximation be described in an equivalent circuit model as a photocurrent generator in parallel with a diode. It corresponds to the single diode equation which will be derived below, in a similar way to the reference (Green, 2003). In order to derive the model, the following assumptions are made: infinite thickness of the absorber, negligible contact resistances, zero contact recombination, zero surface recombination, low injection condition, i.e., the excess charge carrier density is much smaller than the doping density ($\Delta n \ll N_{\mathrm{dop}}$). Under open-circuit condition, there is not a charge carrier flow (no current) into the external circuit resulting in the recombination of all generated charge carriers in the solar cell. Thus the diode equation can be determined from the net recombination rate. The continuity equation for electrons in one dimension under steady-state condition is

$$q\frac{\partial n}{\partial t} = 0 = \nabla \cdot j_{\mathrm{n}} + q\left(G_{\mathrm{n}} - U_{\mathrm{n}}\right) = \nabla \cdot j_{\mathrm{n}} + qG - q\frac{\Delta n}{\tau_{\mathrm{n}}} \tag{39.13}$$

where n is the electron carrier density, j_{n} is the electron current density, G_{n} is the generation rate, U_{n} is the recombination rate and τ_{n} can be interpreted as the average time after which an excess minority carrier recombines. Solving Equation 39.13 for j_{n} leads to

$$j_{\mathrm{n}} = q\int_{0}^{L_{\mathrm{n}}} \left[G_{\mathrm{n}} - q\frac{\Delta n}{\tau_{\mathrm{n}}}\right] \mathrm{d}x = qG_{\mathrm{n}}L_{\mathrm{n}} + q\int_{0}^{L_{\mathrm{n}}} \frac{\Delta n}{\tau_{\mathrm{n}}}\mathrm{d}x \tag{39.14}$$

where L_n is the diffusion lengths for electrons. The relation between carrier lifetime and diffusion length for electrons is determined by the electron diffusion coefficient D_{n}:

$$\tau_{\mathrm{n}} = \frac{L_{\mathrm{n}}^{2}}{D_{\mathrm{n}}} \tag{39.15}$$

Respective equations are derived for holes as minority carriers relevant for the n-type region. The excess charge carrier density Δn may be derived from the law of mass action at all dopant and injection conditions from

$$np = \left(N_{\mathrm{D}} + \Delta n\right)\left(N_{\mathrm{A}} + \Delta n\right) = n_{\mathrm{i,\,eff}}^{2} e^{\left(\frac{E_{\mathrm{Fn}} - E_{\mathrm{Fp}}}{k_{\mathrm{B}}T}\right)} = n_{\mathrm{i,\,eff}}^{2} e^{\left(\frac{qV}{k_{\mathrm{B}}T}\right)} \tag{39.16}$$

where p is the hole concentration, N_{D} is the donor concentration, N_{A} is the acceptor concentration, k_{B} is the Boltzmann constant, T is the temperature and $n_{\mathrm{i,eff}}$ is the effective intrinsic carrier density (Altermatt et al., 2003). The latter depends on the intrinsic carrier concentration n_{i} and the bandgap narrowing (BGN) ΔE_{g}, which is temperature-, doping- and injection-dependent (Schenk, 1998) and is defined by

$$n_{\mathrm{i,eff}}^2 = n_{\mathrm{i}}^2 e^{\left(\frac{\Delta E_{\mathrm{g}}}{k_{\mathrm{B}}T}\right)} \tag{39.17}$$

For a p-type semiconductor under low-level injection, i.e., $\Delta n \ll N_{\mathrm{A}}$, Equation 39.16 reduces to

$$\Delta n \approx \frac{n_{\mathrm{i,eff}}^2}{N_{\mathrm{A}}}\left(e^{\left(\frac{qV}{k_{\mathrm{B}}T}\right)} - 1\right) \tag{39.18}$$

Substituting Equations 39.15 and 39.18 in Equation 39.14 for electrons and holes and integrating over x from L_{n} to L_{p} computes to the total current density:

$$j(V) = j_{\mathrm{n}} + j_{\mathrm{p}} = qG\left[L_{\mathrm{n}} + L_{\mathrm{p}}\right) + \left(\frac{qn_{\mathrm{i,eff}}D_{\mathrm{p}}}{N_{\mathrm{D}}L_{\mathrm{p}}} + \frac{qn_{\mathrm{i,eff}}D_{\mathrm{n}}}{N_{\mathrm{A}}L_{\mathrm{n}}}\right)^{\mathfrak{J}}\left(e^{\left(\frac{qV}{k_{\mathrm{B}}T}\right)} - 1\right) \tag{39.19}$$

This can be simplified to the single diode equation

$$j(V) = j_{\mathrm{SC}} + \left[j_{0\mathrm{e}} + j_{0\mathrm{b}}\right)\left(e^{\left(\frac{qV}{k_{\mathrm{B}}T}\right)} - 1\right) = j_{\mathrm{SC}} + j_0\left(e^{\left(\frac{qV}{k_{\mathrm{B}}T}\right)} - 1\right) \tag{39.20}$$

where j_{SC} is the short-circuit current density, which differs from the photogenerated current density j_{gen}, as not all generated charge carriers are collected because of their finite diffusion lengths. The carriers recombine before they reach the p-n junction. The saturation current density j_0 can be separated in the contribution of the n-type region called emitter saturation current density $j_{0\mathrm{e}}$ and in the contribution of the p-type region called base saturation current density $j_{0\mathrm{b}}$. A more realistic saturation current density of the base for a finite thickness of the solar cell and a non-zero surface recombination is defined by (Goetzberger et al., 1998)

$$j_{0b} = \frac{qn_{\mathrm{i,eff}}D_{\mathrm{n}}}{N_{\mathrm{A}}L_{\mathrm{n}}}G_{\mathrm{Fn}} \tag{39.21}$$

where G_{Fn} is the geometry factor considering the effective surface recombination velocity of the electrons S_n and the thickness of the base (p-type region) W_p:

$$G_{\mathrm{Fn}} = \frac{\cosh\left(\frac{W_{\mathrm{P}}}{L_{\mathrm{n}}}\right) + \frac{D_{\mathrm{n}}}{S_{\mathrm{n}}L_{\mathrm{n}}}\sinh\left(\frac{W_{\mathrm{P}}}{L_{\mathrm{n}}}\right)}{\frac{D_{\mathrm{n}}}{S_{\mathrm{n}}L_{\mathrm{n}}}\cosh\left(\frac{W_{\mathrm{P}}}{L_{\mathrm{n}}}\right) + \sinh\left(\frac{W_{\mathrm{P}}}{L_{\mathrm{n}}}\right)} \tag{39.22}$$

In Figure 39.14, an ideal solar cell I-V and power characteristic are shown, which are calculated using Equation 39.20 assuming $j_{\mathrm{SC}} = -40$ mA/cm^2, $j_0 = 100$ fA/cm^2, and $T = 25$°C.

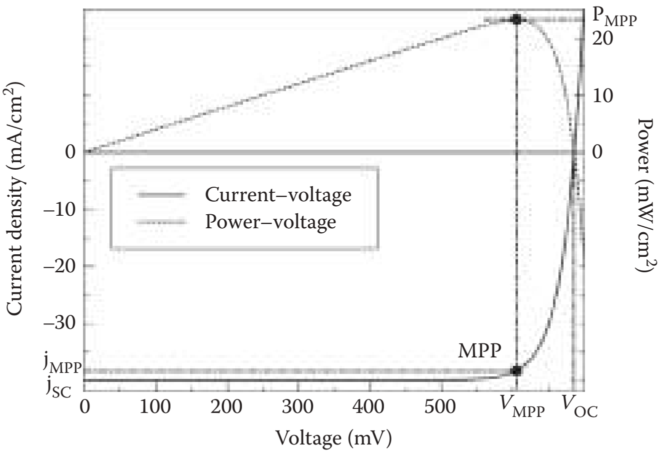


FIGURE 39.14 Example of solar cell current–voltage (I–V) and power characteristics. (MPP, maximum power point.)

39.7.1 *I–V* Parameters and Quantum Efficiency

In Figure 39.14, important output parameters of I–V and P–V characteristics can be extracted as there are: short-circuit current density $j_{SC}(V = 0\text{ V})$, open-circuit voltage V_{OC} ($j(V_{OC}) = 0\text{ mA/cm}^2$) and the maximum power point (MPP) ($P_{MPP}(V) = P_{max}$) with the current density and voltage at MPP j_{MPP} and V_{MPP}, respectively.

The short-circuit current density j_{SC} is the part of photogenerated current density j_{ph} which is collected by the p-n junction due to limited carrier diffusion length within the emitter and the base. A measure of this difference is the internal quantum efficiency (IQE). It is the relation between extracted charge carriers from the solar cell and generated charge carriers in the solar cell. The wavelength-dependent IQE(λ) can be derived from the measurable external quantum efficiency EQE(λ):

$$\text{IQE}(\lambda) = \frac{\text{extracted charge carriers}}{\text{generated charge carriers}} = \frac{\text{EQE}(\lambda)}{(1 - R_{opt}(\lambda) - A_{opt,par}(\lambda) - T_{opt}(\lambda))} \tag{39.23}$$

The EQE also considers optical losses due to reflection, transmission, and parasitic absorption $A_{opt,par}$ in, e.g., dielectric layers at front and rear side. It can be defined by the ratio:

$$\text{EQE}(\lambda) = \frac{\text{extracted charge carriers}}{\text{incident photons}} \tag{39.24}$$

The fill factor (FF) and the efficiency η are derived from the I–V characteristic. The FF is defined by

$$\text{FF} = \frac{j_{MPP}V_{MPP}}{j_{SC}V_{OC}} = \frac{P_{MPP}}{j_{SC}V_{OC}} \tag{39.25}$$

FF is always below 1 and is introduced to describe nonideal losses in the solar cell arising, e.g., due to series resistance losses as discussed in Section 39.8. The efficiency is calculated considering the irradiance G_I and the cell area A_{Cell}:

$$\eta = \frac{P_{MPP}}{G_I A_{Cell}} = \frac{\text{FF}\, j_{SC}V_{OC}}{G_I A_{Cell}} \tag{39.26}$$

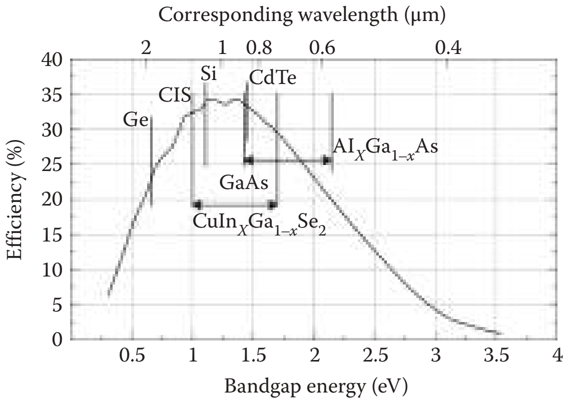


FIGURE 39.15 Dependence of a maximal achievable efficiency on the bandgap energy under an AM1.5g spectrum (1 sun) and a cell operating temperature of 300 K. Optimal bandgap energies of semiconductors are in the range of 1.0–1.5 eV. Established semiconductor absorber materials for solar cells are silicon (Si), Gallium-Arsenide (GaAs), Aluminum-Gallium-Arsenide ($Al_xGa_{1-x}As$), Germanium (Ge), CdTe, Copper-Indium-Diselenide (CIS), Copper-Indium-Gallium-Diselenide (CIGS).

39.7.2 Maximum Efficiency of Ideal Single Junction Solar Cells

For ideal single-junction solar cells, a maximal reachable efficiency may be calculated depending on their bandgap energy E_g following the reference (Wuerfel, 2005) which is very similar to the so-called Shockley–Queisser limit calculation (Shockley and Queisser, 1961). Two loss mechanisms are solely considered. First, nonabsorption losses due to photons with energy lower than the bandgap energy of the semiconductor are considered by Equation 39.10. Second, radiative recombination losses of the semiconductor occur assuming a one-sided emission of only photons of a blackbody spectrum with a temperature of 300 K higher than the bandgap energy, which is described by

$$j_{0,\text{rad}} = qj_{\text{ph,emission}}^{300\text{K}} = q\int_{E_g}^{\infty} \mathrm{d}j_{\text{ph}}^{300\text{K}}(E_{\text{ph}}) \qquad (39.27)$$

Equation 39.27 is derived in Wuerfel (2005) assuming a constant distribution of charge carriers over the absorber thickness, large diffusion lengths, and perfected optics where all photons with the bandgap energy or lower are reabsorbed. Using the single diode model described by Equation 39.20 and substituting 39.26 and 39.27 computes a bandgap energy-dependent maximal efficiency $\eta(E_g)$ for a single p-n junction which is shown in Figure 39.15 for an AM1.5g reference spectrum.

39.8 Nonideal *I–V* Characteristics

39.8.1 Series and Parallel Resistance

Solar cells are large diodes. They suffer under real operation from resistance losses. One differentiates two resistances in solar cells: series resistance r_S and parallel resistance r_P. Note that the small letters indicate area-weighted quantities in $\Omega \cdot \text{cm}^2$ similar to the current density j in mA/cm^2. Series resistance losses arise from current transport in the front metal busbars and metal fingers as well as in front contacts, emitter, base, rear contact, and rear metallization which is visualized by a sketch of a BSF solar cell in Figure 39.16.

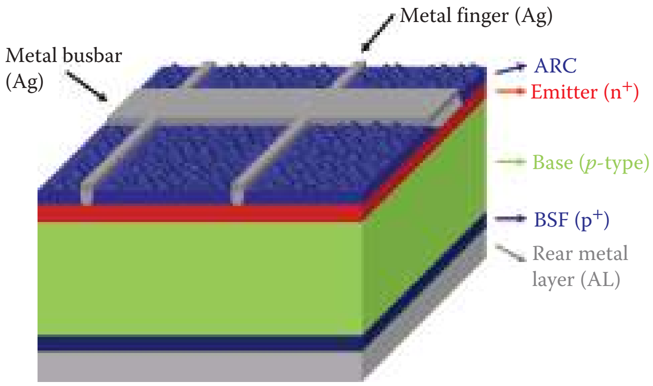


FIGURE 39.16 Sketch of back-surface-field (BSF) solar cell with an indicated random pyramid texture. The screen-printed front and rear metallization consist mainly of silver and aluminum, respectively. During solar cell fabrication, the p-type boron-doped wafer is phosphorus-diffused creating a highly n-type doped emitter. The BSF is formed by an alloying process during a fast firing step while creating the front and rear contact at the same time.

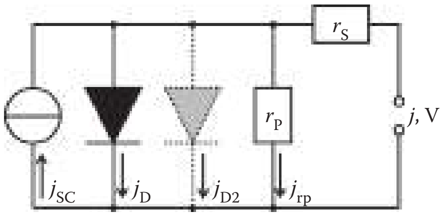


FIGURE 39.17 Equivalent circuit diagram of the single diode and double diode (dotted second diode) model of a solar cell.

As all these parts are included, r_s is called lumped series resistance. Parallel resistance losses, also called shunt resistance losses, arise when the p-n junction is short-circuited, e.g., by grown-in defects of the material (Breitenstein et al., 2004). A corresponding single diode equivalent circuit model is set up as shown in Figure 39.17.

It is described by the implicit Equation 39.28 and examples for analytically calculated I-V characteristics are shown in Figure 39.18.

$$j(V) = j_{SC} + j_D + j_{rp} = j_{SC} + j_0\left(e^{\left(\frac{q(V+j(V)r_S)}{k_B T}\right)} - 1\right) + \frac{V + j(V)\,r_S}{r_P} \tag{39.28}$$

39.8.2 Recombination

As r_S cannot be zero and r_P cannot be infinite, both reduce the FF and thus also the efficiency compared to the ideal case. The consideration of series and parallel resistance is in most cases not sufficient to fully describe I–V characteristics. Additional influences which are the root causes for deviations from the ideal I–V curve are the following:

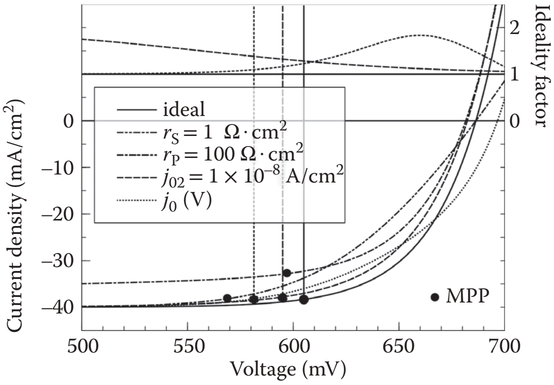


FIGURE 39.18 Solar cell I–V characteristics: ideal, suffering r_S, suffering r_P, suffering nonidealities modeled by the double diode model (compare Equation 39.30) and by a voltage-dependent saturation current density $j_0(V)$ in the voltage range from 500 mV to 700 mV, which is the relevant range for maximum power point (MPP). For details on $j_0(V)$ please refer to the text in Section 39.8.2.

- Distributed series resistance losses (Araújo et al., 1986; Fong et al., 2011) leading to a voltage-dependent change of current flow pattern in solar cells (Altermatt et al., 1996) which results in a voltage-dependent series resistance $r_S(V)$
- Injection-dependent lifetimes, i.e., voltage-dependent recombination within the volume, e.g., due to the boron-oxygen-related defect (Schmidt et al., 2001), or injection-dependent surface recombination velocities, e.g., at thermal oxides (Aberle et al., 1993)
- Edge recombination (Grove and Fitzgerald, 1966; Kuehn et al., 2000)
- Recombination in the space-charge region (SCR) (Wuerfel, 2005)

An injection-dependent lifetime within the absorber, i.e., voltage-dependent recombination, may be modeled by a voltage-dependent saturation current density $j_0\,(V)$ (Müller, 2016) which is typically experienced for Shockley–Read–Hall (SRH) recombination with a single defect with asymmetric capture cross sections, as, e.g., for the boron-oxygen-related defect (Schmidt et al., 1997). A $j_0\,(V)$ may be derived by Equation 39.29 due to a change of the injection density Δn in the solar cell with voltage as described by Equation 39.29. Assuming an excess carrier lifetime for a SRH midgap defect with the electron lifetime parameter $\tau_{n0} = 100$ μs and the hole lifetime parameter $\tau_{p0} = \tau_{n0} \times 10 = 1000$ μs for the parameters $j_{0e} = 10$ fA/cm^2, $N_A = 7.2 \times 10^{15}$cm^{-3}, $W_A = 160$ μm, a corresponding I–V curve is calculated with the single diode Equation 39.20 and is shown in Figure 39.18.

$$j_0\,(V) = j_0\,(\Delta n) = j_{0e} + j_{0b}\,(\Delta n) = j_{0e} + \frac{q W_A n_{i,\mathrm{eff}}^2 (N_A)}{N_A \tau_{\mathrm{eff}}(\Delta n)} \tag{39.29}$$

Recombination in the SCR can be described by the double diode model which is given by

$$j(V) = j_{SC} + j_D + j_{D2} + j_{rp} = j_{SC} + j_0\left(e^{\left(\frac{q(V+jr_S)}{k_B T}\right)} - 1\right) + j_{02}\left(e^{\left(\frac{q(V+jr_S)}{2k_B T}\right)} - 1\right) + \frac{V + jr_S}{r_P} \tag{39.30}$$

Its corresponding equivalent circuit model is also shown in Figure 39.17 and an example for an I–V characteristic is shown in Figure 39.18. The second diode is analytically derived with an ideality factor $n = 2$ (Wuerfel, 2005) for recombination via defects in the space-charge region, compared to $n = 1$ for the first diode. However, SCR recombination is typically not the main contribution to nonideal behavior, i.e., the saturation current density of the second diode j_{02} is mainly a fit parameter, when applying the double diode model to fit dark *I*-*V* characteristics of solar cells. For crystalline wafer-based silicon solar cells, the root cause is typically different and a fitted j_{02} accounts for all effects listed earlier.

Another possibility for the description of nonideal *I*–*V* characteristics is the usage of a voltage-dependent ideality factor $n(V)$ (Rhoderick and Williams, 1988), which can be calculated from dark or light *I*–*V* characteristics by Equation 39.31 if $V > 3k_B T/q$:

$$\frac{1}{n} = \frac{k_B T}{q} \frac{\mathrm{d}(\ln j)}{\mathrm{d}V} \tag{39.31}$$

Exemplarily deriving the voltage-dependent ideality factor $n(V)$, the double diode model is now compared to the voltage-dependent saturation current density $j_0(V)$ using Equation 39.29, which is shown in in the upper part of Figure 39.18. The second diode model has as expected an ideality factor of $n = 2$ for voltages below 400 mV and it decreases toward $n = 1$ almost at V_{OC}. At MPP it is in-between at $n = 1.33$ for the shown example. For $j_0(V)$ the ideality factor increases in this example from $n = 1$ to $n = 1.1$ at MPP to a maximum of about $n = 1.83$ and then decreases again toward $n = 1$ while it is $n = 1.25$ at V_{OC}.

Thus, in general, nonideal *I*–*V* behavior can be expected for any real solar cell and the $n(V)$ evaluation may give hints for its origin (McIntosh, 2001). When analyzing *I*–*V* characteristics, it is necessary to firstly derive a voltage-dependent series resistance with preferable a multiple-light-level determination method (Fong et al., 2011) because current flow patterns within a solar cell remain similar and injection-dependent effects are reduced when changing the illumination only by 10%–20%. In a second step, a series resistance free *I*-*V* curve is derived for the $n(V)$ evaluation (Steingrube et al., 2011).

As an alternative and quicker approach to analyze *I*–*V* characteristics, the *FF* can be evaluated in terms of series resistance induced and recombination-induced losses. The *FF* is compared to the pseudo fill factor (*pFF*) described by Equation 39.32 which accounts only for recombination driven *FF* changes if r_P is sufficiently high ($r_P > 10\ \mathrm{k\Omega \cdot cm^2}$).

$$\mathrm{pFF} = \mathrm{FF} - \Delta\mathrm{FF}_{rs} = \mathrm{FF} - \frac{j_{MPP}^2 r_S(V_{MPP})}{j_{SC} V_{OC}} \tag{39.32}$$

39.9 Numerical Simulation and Models for Crystalline Silicon Solar Cells

The diode model-based approaches are very often not sufficient to describe solar cells; therefore numerical device simulations are carried out. The advantage of numerical simulations compared to analytical solutions is an increased accuracy and provides a deeper insight into device physics and solar cell losses. Analytical solutions are restricted to particular assumptions such as constant photogeneration, constant doping or low injection conditions, which often only approximate real device behavior. Typical problems that demand numerical solar cell simulation are the following:

- Local line front contacts desire a two-dimensional simulation
- Local point rear contacts desire a three-dimensional simulation
- Inhomogeneous distributed wafer quality in case of mc silicon
- Injection-dependent SRH recombination in the bulk
- Injection-dependent SRH recombination at surfaces
- Consideration of Auger recombination in the emitter or BSF with their doping profiles

39.9.1 Numerical Simulation

The physics for the numerical simulation of solar cells are implemented in various simulators. Most developed and widely used simulators are PC1D (Rover et al., 1985), PC1Dmod (Haug et al., 2015) for one-dimensional problems and TCAD Sentaurus Device (Synopsys, 2016), Atlas (Silvaco, 2016), Quokka (Fell, 2013), and AFORS-HET (Varache et al., 2015) for two and three-dimensional simulations. An in-depth introduction into numerical device simulation of crystalline silicon solar cells can be found in Altermatt (2011) which is also used as the source of the most important facts and models discussed below in this section. An update will be given below for the intrinsic recombination model and the incomplete ionization model.

For numerical solar cell simulation a coupled set of differential equations, which are the basic equations of semiconductor device physics, have to be solved by an iterative procedure. This set consists of the Poisson Equation 39.33, which describes the relationship between the electrostatic potential Φ and the charge distribution, the continuity equations for electrons and holes (Equations 39.13, 39.34, 39.35), which conserve the quantity charge carrier, and the current density equations in the drift-diffusion approach for electrons (Equation 39.36) and holes (Equation 39.37) in one dimension:

$$\nabla \cdot (\varepsilon \nabla \Phi) = -q\left(p - n + N_{\mathrm{don}}^{+} - N_{\mathrm{acc}}^{-}\right) \tag{39.33}$$

$$q\frac{\partial n}{\partial t} = \nabla \cdot j_{\mathrm{n}} + q(G - U) \tag{39.34}$$

$$q\frac{\partial p}{\partial t} = \nabla \cdot j_{\mathrm{p}} + q(G - U) \tag{39.35}$$

$$j_{\mathrm{n}} = -q\mu_{\mathrm{n}} n \nabla \Phi + qD_{\mathrm{n}} \nabla n \tag{39.36}$$

$$j_{\mathrm{p}} = -q\mu_{\mathrm{p}} p \nabla \Phi - qD_{\mathrm{p}} \nabla p \tag{39.37}$$

In Equations 39.33 through 39.37, ε is the permittivity of the material, p is the hole carrier density, N_{D}^{+} is the ionized donor concentration and N_{A}^{-} is the ionized acceptor concentration, μ_{n} and μ_{p} are the mobilities for electrons and holes, respectively and D_{p} is the hole diffusion coefficient. The mobilities and diffusion coefficients are related through the Einstein relationships $D_{\mathrm{n}} = \frac{k_{\mathrm{B}}T}{q}\mu_{\mathrm{n}}$ and $D_{\mathrm{p}} = \frac{k_{\mathrm{B}}T}{q}\mu_{\mathrm{p}}$ for a nondegenerated semiconductor. For a degenerated semiconductor (e.g., doping densities higher than $1 \times 10^{19}\ \mathrm{cm}^{-3}$), Fermi–Dirac statistics are valid and the electron transport equation is extended by the additional term $-nk_{\mathrm{B}}T\nabla\left(\ln\left(\frac{n}{N_{\mathrm{C}}}e^{(E_{\mathrm{C}}-E_{\mathrm{Fn}})/(k_{\mathrm{B}}T)}\right)\right)$ and the hole transport equation is extended by $-pk_{\mathrm{B}}T\nabla\left(\ln\left(\frac{p}{N_{\mathrm{V}}}e^{\left(E_{\mathrm{Fp}}-E_{\mathrm{V}}\right)/(k_{\mathrm{B}}T)}\right)\right)$, respectively. N_{C} and N_{V} are the effective density of states for electrons in the conduction band and for holes in the valence band, respectively. At those doping densities higher than $1\times10^{19}\ \mathrm{cm}^{-3}$, Boltzmann statistics overestimates the product of electron and hole density and may lead to a distorted loss analysis between Auger recombination losses in the highly doped region and surface recombination. Thus Fermi–Dirac statistics should be used, especially as the widely used PC1D software is updated using Fermi–Dirac statistics nowadays (Haug et al., 2015).

39.9.2 Models for Crystalline Silicon Solar Cells

A doping and carrier concentration-dependent as well as temperature-dependent mobility model (Klaassen, 1992a,b) is used to calculate Figure 39.19.

For highly doped semiconductors, the bandgap decreases, which is called bandgap narrowing (BGN). It occurs due to carrier-dopant and carrier–carrier interactions, which may be considered for silicon using a BGN model (Schenk, 1998) which is used to calculate Figure 39.20.

The effective intrinsic carrier concentration $n_{\mathrm{i,eff}}$ changes correspondingly with the doping density (Altermatt et al., 2003) using Equation 39.17 and $n_{\mathrm{i,eff}}$ is shown in Figure 39.21.

For medium-doped semiconductors in the range $1\times10^{18}\mathrm{cm}^{-3}$ to $1\times10^{19}\mathrm{cm}^{-3}$ the effect of incomplete ionization occurs, which is caused by the Fermi-level approaching the donor or acceptor level, resulting

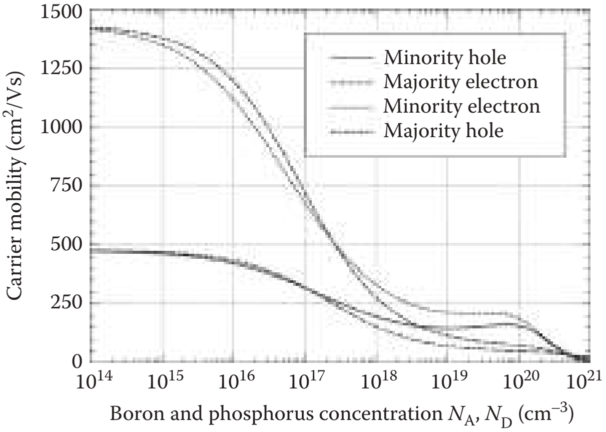


FIGURE 39.19 Klaassen's parametrization of the minority and majority carrier mobility in boron- and phosphorus-doped silicon at 300 K.

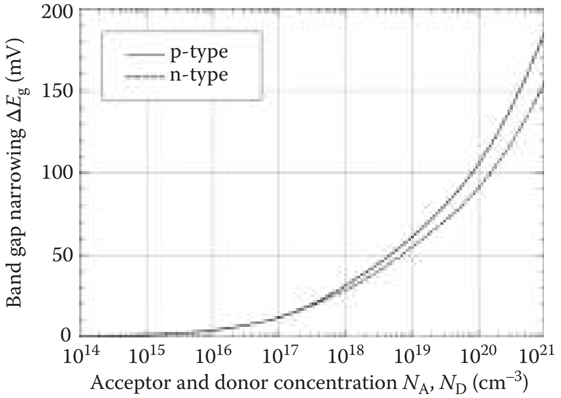


FIGURE 39.20 Bandgap narrowing in p- and n-type silicon calculated with the Schenk model at 300 K.

in the restriction of electrons or holes to localized donor or acceptor states, which prevents free electron or hole movement, respectively. The consideration of incomplete ionization is recommended for the BSF solar cell and a parametrization may be found in reference (Steinkemper et al., 2015).

In order to include the geometrical properties of the solar cell and to decrease the simulation time geometrical considerations have to be performed to derive a well-suited geometry of the device for simulations. Most solar cells have an area of about 15.6×15.6 cm^2 and are about 150–200 μm thick. Smallest features are, e.g., front contacts with a width down to 10 μm. However, it is not necessary to simulate the complete solar cell because its structure is highly symmetric. The simulation domain can be reduced to a geometrically irreducible *standard domain* as shown in Figure 39.22.

This domain is very often two-dimensional where its width is half the spacing between the front metal fingers. Note that the front surface is typically textured, which it is not accounted for in the flat device domain due to the fact that, e.g., random pyramids would result in a huge three-dimensional domain. For simulation, the device domain is discretized on a mesh as shown in Figure 39.22. The mesh should be fine

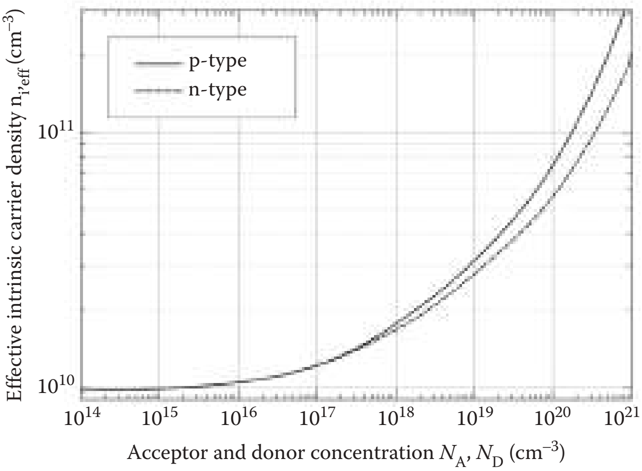


FIGURE 39.21 Influence of the doping density on the effective intrinsic carrier concentration in p- and n-type silicon at 300 K.

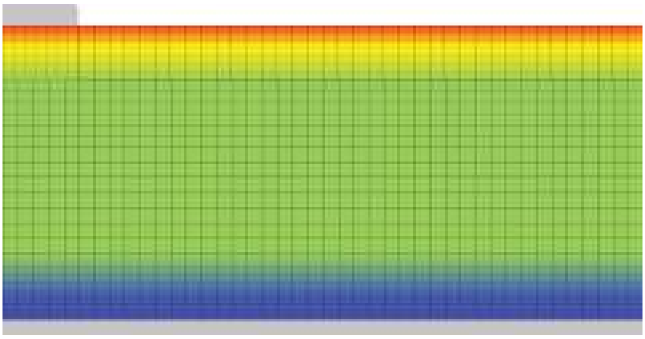


FIGURE 39.22 Geometrically irreducible two-dimensional *standard domain* for a BSF solar cell. Mesh refinements near the front surface, in diffused layers and close to contacts are indicated.

enough where any input or output parameter varies strongly with distance, which is for solar cells close to the front surface where the blue part of the sunlight is absorbed within nanometers, and where a front diffusion layer is typically situated. Besides near metal contacts the mesh should be refined. The generation of charge carriers as described in Section 39.4 is typically considered by an one-dimensional generation profile $G(x)$. The reverse process of intrinsic recombination may be included by a combined model for the radiative recombination and the Auger recombination (Richter et al., 2012), which is used to calculate Figure 39.23.

At surfaces, surface recombination occurs via surface states within the bandgap and additionally fixed charges are present which lead to local band bending (Kingston and Neustadter, 1955).

The most interesting result of a simulation is the sweep of an $I-V$ curve which is performed changing the applied voltage at the contacts as boundary condition. For a full solar cell $I-V$ characteristic, the standard domain I-V curves can be interconnected in a circuit by ohmic resistances, in, e.g., a SPICE network simulation, which represent the front metal fingers. A sketch of such a network is shown in Figure 39.24.

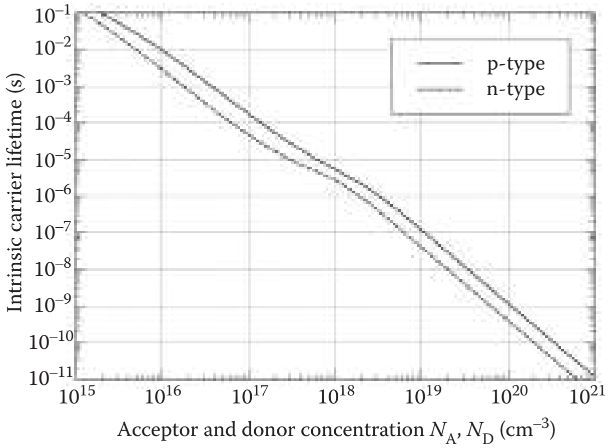


FIGURE 39.23 Influence of the doping density on radiative and Auger recombination in silicon (intrinsic recombination) in p- and n-type silicon at 300 K.

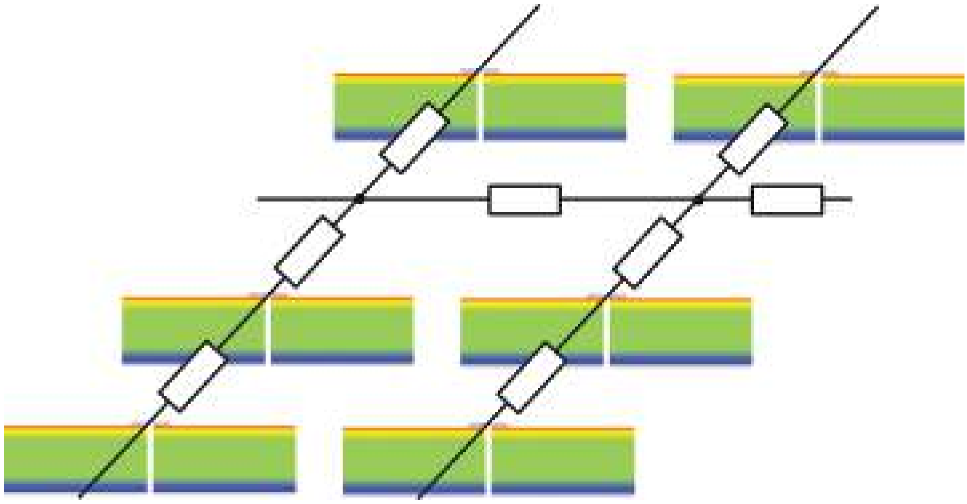


FIGURE 39.24 Sketch of a network of resistances and standard domains which allow simulating a full solar cell I–V characteristic by device and circuit simulation. Compare Figure 39.16 to the sketch of the same solar cell part.

39.10 Real Working Behavior of Solar Cells

Under real operating conditions, main changes of solar cell performance arise due to its intensity and temperature behavior.

39.10.1 Light Intensity Behavior

The intensity dependence is mainly governed by the decrease of irradiance G which results in a decrease of the short-circuit current density j_{SC}. For typical crystalline silicon solar cells under nonconcentrated sunlight conditions, the output power is mainly determined by the linear relationship $G \propto j_{SC}$. Deviations may occur for irradiance conditions below 0.2 suns, where injection-dependent effects, e.g., injection-dependent carrier lifetime, become apparent. The general change of the open-circuit voltage V_{OC} with irradiance G may be derived from the ideal-behavior solving the single diode Equation 39.20 to V_{OC} for $j(V_{OC}) = 0\ \text{mA/cm}^2$:

$$V_{OC} = \frac{k_B T}{q} \ln\left(1 + \frac{j_{SC}}{j_0}\right) \tag{39.38}$$

The change of *FF* with irradiance G can be approximated by Equation 39.39 (Green, 1982) for the ideal-behavior, which depends on V_{OC}, and is called *ideal fill factor* FF_0.

$$FF_0\left(v_{OC}\right) = \frac{v_{OC} - \ln\left(v_{OC} + 0.72\right)}{v_{OC} + 1}, \text{with} v_{OC} = \frac{q}{k_B T} V_{OC} \tag{39.39}$$

An example of a calculated *I-V* characteristic under low illumination using the single diode Equation 39.20 is shown in Figure 39.25, assuming $j_{SC} = 20\ \text{mA/cm}^2$ and $j_0 = 100\ \text{fA/cm}^2$. The almost halved j_{SC} results in nearly half of the maximum power.

The so-called low light behavior of a solar cell is shown in Figure 39.26 using the single diode equivalent circuit model (Equation 39.28) with j_{SC} (STC) $= -40\ \text{mA/cm}^2$ and $j_0 = 100\ \text{fA/cm}^2$. Additionally, the relative efficiency normalized to its STC value is calculated and compared:

$$\eta_{rel}(G, T) = \frac{\eta(G, T)}{\eta(1000 \frac{W}{m^2}, 25°C)} \tag{39.40}$$

For the ideal-behavior ($r_S = 0\ \Omega \cdot \text{cm}^2$) in Figure 39.26, it is deduced that the efficiency decreases with decreasing irradiance, which is mainly due to the V_{OC} decrease and only slightly due to the *FF* decrease. The variation of the lumped series resistance in Figure 39.26 shows that its influence governs

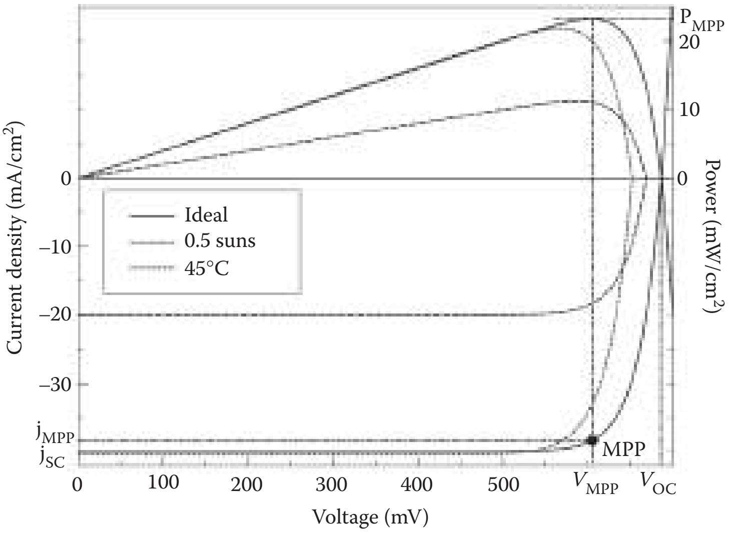


FIGURE 39.25 Comparison of solar cell I-V and power characteristics under STC, under low illumination of 0.5 suns, and under increased temperature of $T = 45°C$. (MPP, maximum power point.)

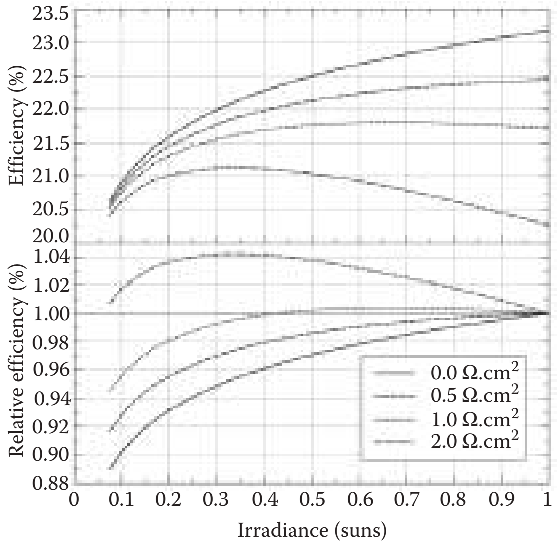


FIGURE 39.26 Relative efficiency versus irradiance for different lumped series resistances r_s, where $r_S = 0\ \Omega \cdot \text{cm}^2$ represents the ideal behavior, whereas $r_S = 0.5\ \Omega \cdot \text{cm}^2$ is a typical value for c-Si solar cells and r_S between 1 and $2\ \Omega \cdot \text{cm}^2$ is typical for solar modules due to the cell interconnection, e.g., the tabbing ribbons cause additional series resistances.

within the irradiance range from 200 W/m^2 to 1000 W/m^2. In this range, the current in the solar cell is sufficiently high and η saturates or even reduces toward STC with increasing r_S. This results in an increasing relative efficiency under low light condition for solar cells and modules with high r_S (compare lower part of Figure 39.26), which seems beneficial under real operating condition. However, the series resistance solely limits the STC efficiency (compare upper part of Figure 39.26) and the solar module will stay behind its potential. In conclusion, r_S should be low for maximum energy yield.

39.10.2 Temperature Behavior

The temperature behavior is determined by several effects (Dupré et al., 2015). It increases the effective intrinsic carrier density $n_{\text{i,eff}}$ (Misiakos and Tsamakis, 1993), which decreases the quasi-Fermi-level splitting and thus at higher temperature the voltage at the same injection density is reduced (compare Equations 39.17 and 39.18). A detrimental effect, but smaller, is the increase of the bandgap with increasing temperature, which results in a slight increase in absorbed photon current density. Thus j_{SC} is slightly increased. The change of *FF* with temperature depends on a variety of influences, e.g., temperature-dependent conductivity of the front and rear metallization, as well as front and rear contacts. For silicon, the *FF* typically decreases with higher temperature. Surprisingly, the temperature behavior of j_{SC}, V_{OC} and η is typically linearly under relevant operating conditions in the range of 0°C–75°C. Therefore, it is expressed in terms of their absolute temperature coefficients $TC_{\text{abs}}(j_{SC})$, $TC_{\text{abs}}(V_{OC})$ and $TC_{\text{abs}}(\eta)$ or more often in terms of their relative counterparts $TC_{\text{rel}}(j_{SC})$, $TC_{\text{rel}}(V_{OC})$ and $TC_{\text{rel}}(\eta)$ normalized to STC similar to Equation 39.40. The relevant temperature coefficient of the efficiency $TC_{\text{rel}}(\eta)$ is in the range of −0.5%/K to −0.25%/K for crystalline silicon solar cells. An example of a calculated *I-V* curve at 45°C using the single diode model is shown in Figure 39.25 assuming a $j_{SC}(45°\text{C}) = 40.4\ \text{mA/cm}^2$ and $j_0\ (45°\text{C}) = 191\ \text{fA/cm}^2$. This corresponds to a $TK_{\text{rel}}(j_{SC}) = 0.05\%/\text{K}$, i.e., $\Delta j_{SC}\ (\Delta T) = j_{SC}\ (25°\text{C}) \times TC_{\text{rel}}\ (j_{SC}) \times \Delta T$, and a j_0 increase which is approximated by the proportionality to $T^3 \times e^{-\frac{E_g(T)}{k_B T}}$ (Wolf et al., 1977). The increased temperature leads to decreased V_{OC} and thus to slightly lower P_{MPP}.

39.10.3 Energy Yield

The real working behavior of solar cells and solar modules, respectively, is regarded in terms of their delivered energy yield below to complete the information given in this chapter.

The main influences on solar cell performance under real operating conditions are changing irradiance, cell temperature, angle of incidence, and spectral variations compared to the STC condition. The spectral variation is below 1% for c-Si solar cells (Huld et al., 2011) and hence will not be considered in the further discussion. For a typical solar module, the energy yield is calculated for the three locations introduced in Section 39.3 using the Equations 39.3, 39.5 through 39.8, 39.26 and 39.28 assuming c-Si solar module relevant parameters $j_0 = 100\ \text{fA/cm}^2$, $r_S = 1\Omega \cdot cm^2$, $r_P = \infty$, $TC_{\text{rel}}(\eta) = 0.4\%/°C$, $K = 4\ \text{m}^{-1}$ and $L = 2\ \text{mm}$ and $NOCT = 45°\text{C}$. The specific yields per month are shown in the lower part of Figure 39.27 using Equation 39.2 and the specific yields over 1 year are $Y_f = 1120$ kWh/kWp for Halle, $Y_f = 1823$ kWh/kWp for Almeria, and $Y_f = 1504$ kWh/kWp for Brazzaville.

In a first-order approximation, the specific yield follows the available solar irradiation as shown in Figure 39.8 which results, e.g., in low Y_f for winter months in Halle or similar Y_f over the whole year for Brazzaville near the equator. To separate the irradiation influence, the performance ratio has been introduced in Equation 39.1 and here *PR* is shown in the upper part of Figure 39.27. It is mostly below 100% due to increased cell operating temperature above 25°C in summer months for all locations and due to the warmer climate in Almeria and Brazzaville. Performance ratios over 1 year are $PR = 96.6\%$ for Halle, $PR = 92.9\%$ for Almeria and $PR = 90.8\%$ for Brazzaville, which are very similar to the estimated temperature losses derived from the annual average cell operating temperature in Section 39.3.

The calculations are carried out assuming no additional cell-to-module efficiency changes while module series resistance and optics under oblique incidence are included. Additional *PR* losses arise in reality due to losses in the inverter (up to 5%), due to cell-to-cell and module-to-module mismatches (about 1%), due to shading (depends on location), thus good *PR* are in the range of 85%–90% in the first years. Additionally, efficiency degradation occurs over time of about $0.5\%_{\text{rel}}$/year (Jordan and Kurtz, 2013), which further decreases the *PR* of about 6.25% over the system lifetime of about 25 years.

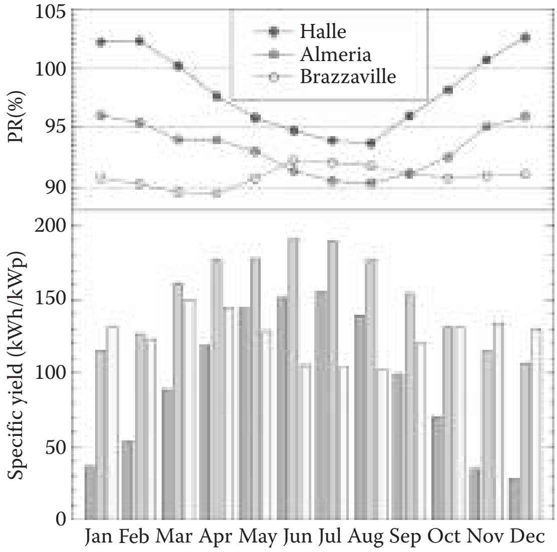


FIGURE 39.27 Specific yield (lower graph) and performance ratio (upper graph) per month for the three locations Halle, Almeria, and Brazzaville.

39.11 Conclusion and Outlook

PV systems became already an integral part of our world in terms of energy supply and also in our landscapes. The levelized cost of electricity (LCOE) of PV is approaching that of conventional electricity power plants (IEA, 2014). Thus, a major share of electricity is expected to be PV electricity in the next decades. However, it is a fluctuating energy resource and storage is getting more and more important.

Wafer-based silicon solar cells became the market leading technology and at the moment a change is unlikely to happen. The competitive thin-film solar cells on CIGS and CdTe absorber material benefit from reduced material consumption, but are still built in solar module with front glass for long lifetime which nowadays lowers significantly the inherent cost advantage. Organic and dye-sensitized (inorganic) thin-film absorbers are not showing a high efficiency potential yet (Green et al., 2015).

New solar cell structures have to be designed right away in cost competitive solar modules with a long lifetime. For example, the recently emerging perovskite-based solar cells (Green et al., 2014) have shown an incredible efficiency increases in the past 5 years, but they still lack stability against UV light and will face similar module cost disadvantage as thin-films solar cells do. Efficiency matters, but costs and stability have to be competitive in order to gain economic success.

Besides the single-junction solar cells for 1 sun illumination, there are several other developing technologies and approaches. Concentrating photovoltaics (CPV) is close to enter the market. CPV uses optical elements to increase the irradiance on a smaller solar cell size for increased efficiency and reduced material consumption. However, for higher concentration (>20 suns) the PV system needs to track the sun, which results in higher complexity and higher maintenance requirements. Increased efficiency of solar cells may be reached by so-called third generation PVs (Green, 2003). Examples are multijunction solar cells which are stacks of semiconductor absorber materials of different bandgap energies reducing thermalization losses which have shown great potential (Green et al., 2015), whereas hot carrier solar cells, multiband solar cells, and photon up- and down-converting materials are still in focus of basic research.

Acknowledgments

I would like to thank Eric Schneiderlöchner, Pietro Altermatt, Byungsul Min, and Julia Müller for their great support.

References

Aberle A G, Robinson S J, Wang A, Zhao J, Wenham S R and Green M A (1993) High-efficiency silicon solar cells: Fill factor limitations and non-ideal diode behaviour due to voltage-dependent rear surface recombination velocity. *Prog. Photovolt Res. Appl.* 1(2):133–143.

Altermatt P P (2011) Models for numerical device simulations of crystalline silicon solar cells—A review. *J. Comput. Electron.* 10(3):314–330.

Altermatt P P, Heiser G, Aberle A G, Wang A, Zhao J, Robinson S J, Bowden S and Green M A (1996) Spatially resolved analysis and minimization of resistive losses in high-efficiency Si solar cells. *Prog. Photovolt Res. Appl.* 4:399–414.

Altermatt P P, Schenk A, Geelhaar F and Heiser G. (2003) Reassessment of the intrinsic carrier density in crystalline silicon in view of band-gap narrowing. *J. Appl. Phys.* 93(3):1598–1604.

Araújo G L, Cuevas A and Ruiz J M (1986) The effect of distributed series resistance on the dark and illuminated current-voltage characteristics of solar cells. *IEEE Trans. Electron Dev*. 33(3):391–401.

ASTM E490-00a (2014) *Standard Solar Constant and Zero Air Mass Solar Spectral Irradiance Tables*. West Conshohocken, PA: ASTM International. www.astm.org.

ASTM G173-03 (2012) *Standard Tables for Reference Solar Spectral Irradiances: Direct Normal and Hemispherical on* 37° *Tilted Surface*. West Conshohocken, PA: ASTM International. www.astm.org.

Becquerel A E (1839) Recherches sur les effets de la radiation chimique de la lumière solair au moyen des courants électriques. *C. R. Acad. Sci.* 9:561–567.

Blakers A, Wang A, Milne A, Zhao J and Green M A (1989) 22.8% efficient silicon solar cell. *Appl. Phys. Lett.* 55:1363–1365.

Breitenstein O, Rakotoniaina J P, Al Rifai M H and Werner M (2004) Shunt types in crystalline silicon solar cells. *Prog. Photovolt Res. Appl.* 12(7):529–538.

Brendel R (1994) Sunrays: A versatile ray tracing program for the photovoltaic community. In: *Proceedings of the 12th European Photovoltaic Solar Energy Conference*, Amsterdam, Netherlands, pp. 1339–1342.

Chapin D M, Fuller C S and Pearson G O (1954) A new silicon p–n junction photocell for converting solar radiation into electrical power. *J. Appl. Phys.* 25:676–677.

Duffie J A and Beckman W A (2013) *Solar Engineering of Thermal Processes* (4th edition). Hoboken, NJ: John Wiley & Sons Inc.

Dupré O, Vaillon R and Green M A (2015) Physics of the temperature coefficients of solar cells. *Sol. Energ. Mat. Sol. Cells.* 140:92–100.

Fell A (2013) A free and fast three-dimensional/two-dimensional solar cell simulator featuring conductive boundary and quasi-neutrality approximations. *IEEE Trans. Electron Dev.* 60(2):733–738.

Fong K C, McIntosh K R and Blakers A W (2011) Accurate series resistance measurement of solar cells. *Prog. Photovolt Res. Appl.* 21(4):490–499.

Fraunhofer ISE (2013) *Photovoltaics Report*, 7 November 2013. http://www.ise.fraunhofer.de/en/downloads-englisch/pdf-files-englisch/photovoltaics-report-slides.pdf

Garcia M C A and Balenzategui J L (2004). Estimation of photovoltaic module yearly temperature and performance based on Nominal Operation Cell Temperature calculations. *Renew. Energy* 29:1997–2010.

Goetzberger A, Knobloch J and Voß B (1998) *Crystalline Silicon Solar Cells*. Chichester: John Wiley & Sons Ltd.

Green M (1982) *Solar Cells: Operating Principles, Technology, and System Applications*. Englewood Cliffs, NJ: Prentice-Hall, Inc.

Green M A (2003) *Third Generation Photovoltaics–Advanced Solar Energy Conversion.* Berlin Heidelberg: Springer-Verlag.

Green M A, Emery K, Hishikawa Y, Warta W and Dunlop E D (2015). Solar cell efficiency tables (Version 45). *Prog. Photovolt Res. Appl.* 23(1):1–9.

Green M A, Ho-Baillie A and Snaith H J (2014) The emergence of perovskite solar cells. *Nat. Photonics.* 8(7):506–514.

Grove A S and Fitzgerald D J (1966) Surface effects on p-n junctions: Characteristics of surface space-charge regions under non-equilibrium conditions. *Solid State Electron.* 9:783–806.

Haug H, Greulich J, Kimmerle A and Marstein E S (2015) PC1Dmod 6.1—State-of-the-art models in a well-known interface for improved simulation of Si solar cells. *Sol. Energ. Mat. Sol. Cells.* 142:47–53.

Holst H, Winter M, Vogt M R. Bothe K, Kontges M, Brendel R and Altermatt P (2013) Application of a new ray tracing framework to the analysis of extended regions in Si solar cell modules. *Energy Procedia* 38:86–93.

Huld T, Friesen G, Skoczek A, Kenny R P, Sample T, Field M and Dunlop E D (2011) A power-rating model for crystalline silicon PV modules. *Sol. Energ. Mat. Sol. Cells.* 95(12):3359–3369.

IEC 61215 (2005) *Crystalline Silicon Terrestrial Photovoltaic (PV) Modules—Design Qualification and Type Approval.* Geneva, Switzerland: International Electrotechnical Commission (IEC). http://www.iec.ch.

International Energy Agency (IEA) (2014) *Technology Roadmap: Solar Photovoltaic Energy* (2014 edition). Accessed January 12, 2016. https://www.iea.org/media/freepublications/technologyroadmaps/solar/TechnologyRoadmapSolarPhotovoltaicEnergy_2014edition.pdf

ITRPV (July 2015) *International Technology Roadmap for Photovoltaic: Results 2014*, 6th edition. Accessed January 2016. http://www.itrpv.net/.cm4all/iproc.php/Reports%20downloads/ITRPV_Roadmap_2015_Rev1_July_150722.pdf?cdp=a

Joint Research Center (JRC) (2016) Photovoltaic Geographical Information System (PVGIS). Accessed 18 January 2016. http://re.jrc.ec.europa.eu/pvgis/

Jordan D C and Kurtz S R (2013) Photovoltaic degradation rates—An analytical review. *Prog. Photovolt Res. Appl.* 21(1):12–29.

Kingston R H and Neustadter S F (1955) Calculation of the space charge, electric field, and free carrier concentration at the surface of a semiconductor. *J. Appl. Phys.* 26(6):718–720.

Klaassen D B M (1992a) A unified mobility model for device simulation—I. Model equations and concentration dependence. *Solid State Electron.* 35(7):953–959.

Klaassen D B M (1992b) A unified mobility model for device simulation—II. Temperature dependence of carrier mobility and lifetime. *Solid State Electron.* 35(7):961–967.

Kopp G and Lean J L (2011) A new, lower value of the total solar irradiance: Evidence and climate significance. *Geophys. Res. Lett.* 38:L01706.

Kuehn R, Fath P and Bucher E (2000) Effects of pn junctions bordering on surfaces investigated by means of 2D-modeling. In: *Proceedings of the 28th IEEE Photovoltaic Specialists Conference*, Anchorage, AK, pp. 116–119.

Lammert M and Schwartz R (1977) The interdigitated back contact solar cell: A silicon solar cell for use in concentrated sunlight. *IEEE Trans. Electron Dev.* 24(4):337–342.

Macleod H A (2010) *Thin-Film Optical Filters* (4th edition). Boca Raton, FL: CRC Press, Taylor & Francis.

Mandelkorn J and Lamneck J (1972) Simplified fabrication of back surface electric field silicon cell and novel characteristics of such cells. In *Proceeding 9th IEEE Photovoltaic Specialists Conference*, Silver Springs, MD.

McIntosh K R (2001) *Lumps, Humps and Bumps: Three Detrimental Effects in the Current-Voltage Curve of Silicon Solar Cell.* PhD thesis, University of New South Wales, Sydney, Australia.

Meteotest (2008) *Meteonorm Version 6.1.0.0.* Accessed January 12, 2016. www.meteotest.com.

Misiakos K and Tsamakis D (1993) Accurate measurements of the silicon intrinsic carrier density from 78 to 340 K. *J. Appl. Phys.* 74(5):3293–3297.

Müller M (2016) Reporting effective lifetimes at solar cell relevant injection densities. *Energy Proc.* 92:138—144.

National Aeronautics and Space Administration (NASA) (2016) Surface meteorology and Solar Energy Accessed January 18, 2016. https://eosweb.larc.nasa.gov/sse/.

National Renewable Energy Laboratory (NREL) (2016) *Renewable Resource Data Center (RReDC).* Accessed January 18, 2016. http://www.nrel.gov/rredc/.

Pingel S, Frank O, Winkler M, Daryan S, Geipel T, Hoehne H and Berghold J (2010) Potential induced degradation of solar cells and panels. In: *Proceedings 35th IEEE Photovoltaic Specialists Conference*, Honolulu, HI, 2010, pp. 2817–2822.

PV Lighthouse (2016) PV Lighthouse. Accessed 12 January 2016. https://www.pvlighthouse.com.au

PV*EDUCATION.ORG* (2016) PVEducation Accessed January 12, 2016. http://pveducation.org

pvXchange Trading GmbH (2016) Market price data. Accessed January 12, 2016. http://www.pvxchange.com/priceindex/Default.aspx?langTag=de-DE

Quaschning V (2005) *Understanding Renewable Energy Systems.* London, UK: Earthscan.

Rhoderick E H and Williams R H (1988) *Metal-Semiconductor Contacts* (2nd edition). New York, NY: Oxford University Press.

Richter A, Glunz S W, Werner F, Schmidt J and Cuevas A (2012) Improved quantitative description of Auger recombination in crystalline silicon. *Phys. Rev. B.* 86(16):165202.

Ross R G Jr and Smokler M I (1986). Flat-Plate Solar Array Project–Final Report, Vol. VI Engineering Sciences and Reliability, JPL Pub. No. 86–31, http://ntrs.nasa.gov/search.jsp?R=19870011218.

Rover D T, Basore P A, Thorson G M (1985) PC-1D version 2: Enhanced numerical solar cell modelling. In: *Proceedings of the 18th IEEE Photovoltaic Specialists Conference*, Las Vegas, NV, p. 703.

Schenk A (1998) Finite-temperature full random-phase approximation model of band gap narrowing for silicon device simulation. *J. Appl. Phys.* 84(7):3684–3695.

Schmidt J, Aberle A G and Hezel R (1997) Investigation of carrier lifetime instabilities in Cz-grown silicon. In: *Proceedings of the 26th IEEE Photovoltaic Specialists Conference*, New York, NY, pp. 13–18.

Schmidt J, Cuevas A, Rein S and Glunz S W (2001) Impact of light-induced recombination centres on the current-voltage characteristic of czochralski silicon solar cells. *Prog. Photovolt Res. Appl.* 9(4):249–255.

Shockley W and Queisser H J (1961) Detailed balance limit of efficiency of p-n junction solar cells. *J. Appl. Phys*. 32(3):510–519.

Silvaco (2016) *Atlas*. Santa Clara, CA: Silvaco. http://www.silvaco.com.

Steingrube S, Wagner H, Hannebauer H, Gatz S, Chen R, Dunham S T, Dullweber T, Altermatt P P, Brendel R (2011) Loss analysis and improvements of industrially fabricated Cz-Si solar cells by means of process and device simulations. *Energy Procedia* 8:263–268.

Steinkemper H, Rauer M, Altermatt P P, Heinz F D, Schmiga C and Hermle M (2015) Adapted parameterization of incomplete ionization in aluminum-doped silicon and impact on numerical device simulation. *J. Appl. Phys.* 117(7):074504.

Synopsys (2016) *TCAD Sentaurus Device version 2015.06*. Mountain View, CA: Synopsys. http://www.synopsys.com.

Tanaka M, Taguchi M, Matsuyama T, Sawada T, Tsuda S, Nakano S, Hanafusa H and Kuwano Y (1992) Development of new a-Si/c-Si heterojunction solar cells: ACJ-HIT (artificially constructed junction-heterojunction with intrinsic thin-layer). *Jpn. J. Appl. Phys.* 31:3518–3522.

Varache R, Leendertz C, Gueunier-Farret M E, Haschke J, Muñoz D and Korte L (2015) Investigation of selective junctions using a newly developed tunnel current model for solar cell applications. *Sol. Energ. Mat. Sol. Cells*. 141:14–23.

Winter M, Vogt M R, Holst H, Altermatt P P (2014) Combining structures on different length scales in ray tracing: Analysis of optical losses in solar cell modules. In: *Proceedings of the 14th International Conference on Numerical Simulation of Optoelectronic Devices,* Palma de Mallorca, Spain.

Wolf M, Noel G T and Stirn R J (1977) Investigation of the double exponential in the current-voltage characteristics of silicon solar cells. *IEEE Trans. Electron Dev.* 24(4):419–428.

Wuerfel P (2005) *Physics of Solar Cells*. Weinheim, Germany: Wiley-VCH Verlag GmbH. KGaA.

Zhao J, Wang A and Green M A (1999) 24.5% efficiency silicon PERT cells on MCZ substrates and 24.7% efficiency PERL cells on FZ substrates. *Prog. Photovolt Res. Appl.* 7:474–474.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

40

Multijunction Solar Cells

40.1 Introduction 415
40.2 Device Model and Simulation Workflow 418
40.3 Mesh Discretization 419
40.4 Optics 420
Transfer Matrix Method • Complex Refractive Index • Absorption and Quantum Yield
40.5 Drift-Diffusion Calculation 423
Band Structure and Intrinsic Carrier Concentration • Recombination, Lifetimes, and Surface/Interface Recombination Velocities • Tunnel Diodes
40.6 Current–Voltage Calculation 431
40.7 Quantum Efficiency Calculation 433
40.8 Photon Recycling and Luminescent Coupling 434
40.9 Summary 437

Matthew Wilkins
and
Karin Hinzer

40.1 Introduction

Multijunction solar cells (MJSCs) currently hold the records for greatest power conversion efficiency of any solar cell, under one-sun conditions (Spectrolab, 38.8% [1]) and under concentration (Soitec/Fraunhofer, 46.0% at 508 suns). MJSCs based on III–V materials such as GaAs and GaInP find applications in aerospace applications where size and weight take precedence over cost, and in terrestrial concentrating photovoltaic (CPV) systems where concentrating optics focus the incident sunlight onto very small (< 5 × 5 mm) cells. MJSCs based on material systems, such as amorphous silicon/silicon and II-VI compounds, and those using III–Vs grown on alternative substrates, have potential to combine high efficiency with lower costs suitable for nonconcentrating terrestrial applications.

Single-junction solar cells have two major loss mechanisms: thermalization and transmission. Photons of energy less than the semiconductor bandgap are transmitted through the material without being absorbed (transmission), while photons of energy in excess of the bandgap convert that excess energy to heat. Single junctions, then, must be designed to minimize transmission (requiring a low bandgap E_g), and minimizing thermalization (requiring a large E_g). For typical solar spectra, this leads to an optimal bandgap of ~1.3–1.4 eV [2,3].

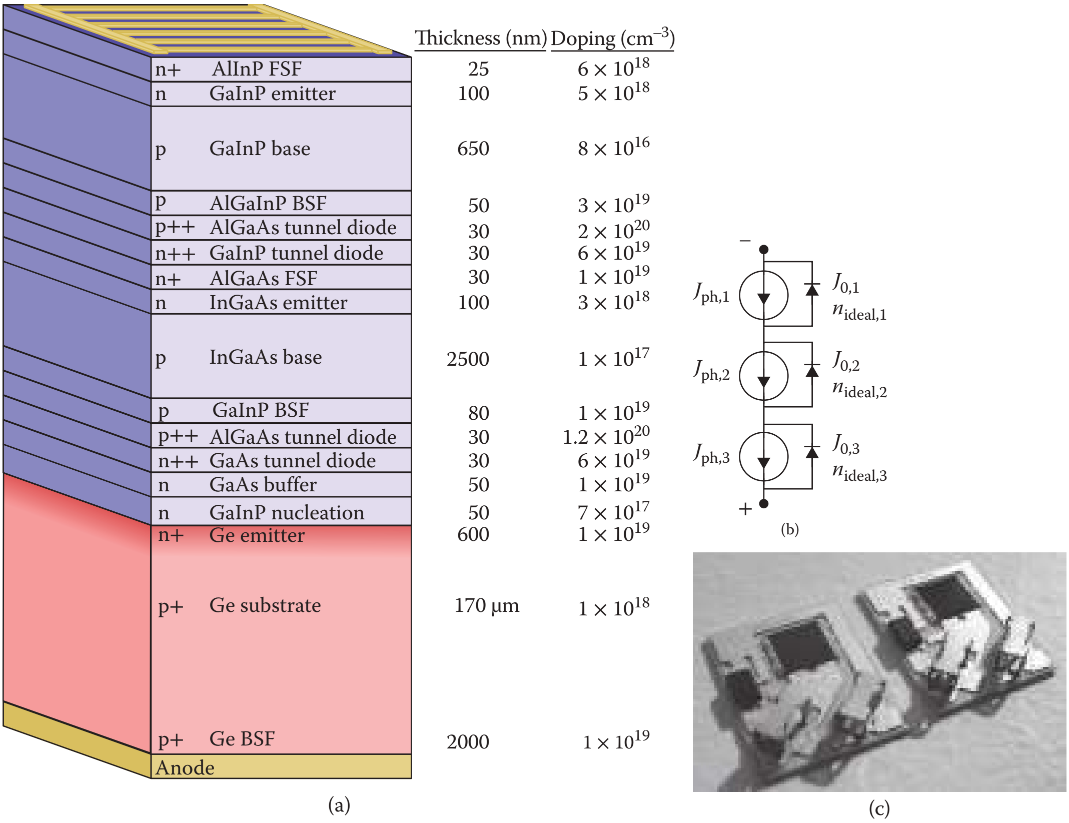
MJSCs consist of a stack of *n*-*p* semiconductor diodes of different materials, each chosen to absorb a different part of the solar spectrum. As light passes through the stack, the short wavelengths are absorbed in the top junction which has a large E_g, and longer wavelengths are absorbed in each of the subsequent junctions. This arrangement minimizes the impact of transmission and thermalization losses by providing materials that are capable of absorbing the vast majority of photons within the solar spectrum, and by ensuring that each photon is absorbed in a junction with E_g similar to the photon energy. Typically, the junctions are all grown epitaxially in a monolithic device, with all junctions connected in series; this implies that the current through all junctions must be the same. Less commonly, some devices are designed such that separate contacts are made to each junction so that they can each be operated independently.

Since MJSCs are more expensive to produce than single-junction solar cells, terrestrial applications of MJSCs typically involve concentrating optics that focus light from a large aperture onto a small area of solar cell. In these systems, the sun's intensity is concentrated by a factor of 10 to 2000. Besides the cost savings from using smaller cells, there is also a significant increase in cell efficiency under concentration. We can write the efficiency η as

$$\eta = \frac{P_{\max}}{XE_{1-\text{sun}}} = \frac{J_{SC}\, V_{OC}\, FF}{XE_{1-\text{sun}}}, \tag{40.1}$$

where $E_{1-\text{sun}}$ is the incident light intensity under standard 1-sun conditions, X is the concentration factor, $P_{\max}$ is the output electrical power density at the optimal voltage bias, J_{SC} is the short-circuit current density, V_{OC} is the open-circuit voltage, and FF is the fill factor.

A simple equivalent circuit model for the series-connected triple-junction solar cell is shown in Figure 40.1b; each junction i is modeled with a current source providing the 1-sun photocurrent for that


Thickness (nm)
Doping (cm^{-3})
n+ AlInP FSF 25 6×10^{18}
n GaInP emitter 100 5×10^{18}
p GaInP base 650 8×10^{16}
p AlGaInP BSF 50 3×10^{19}
p++ AlGaAs tunnel diode 30 2×10^{20}
n++ GaInP tunnel diode 30 6×10^{19}
n+ AlGaAs FSF 30 1×10^{19}
n InGaAs emitter 100 3×10^{18}
p InGaAs base 2500 1×10^{17}
p GaInP BSF 80 1×10^{19}
p++ AlGaAs tunnel diode 30 1.2×10^{20}
n++ GaAs tunnel diode 30 6×10^{19}
n GaAs buffer 50 1×10^{19}
n GaInP nucleation 50 7×10^{17}
n+ Ge emitter 600 1×10^{19}
p+ Ge substrate 170 μm 1×10^{18}
p+ Ge BSF 2000 1×10^{19}
Anode
−
$J_{ph,1}$
$J_{0,1}$
$n_{ideal,1}$
$J_{ph,2}$
$J_{0,2}$
$n_{ideal,2}$
$J_{ph,3}$
$J_{0,3}$
$n_{ideal,3}$
+

(a) (b) (c)

FIGURE 40.1 (a) Layer structure of a triple-junction GaInP/(In)GaAs/Ge solar cell. (b) Simple circuit model for the triple-junction cell. (c) Two 5 × 5-mm MJSCs mounted on ceramic carriers.

subcell, $J_{\mathrm{ph},i}$. A parallel diode provides the exponential diode behavior of each cell, so the device current density

$$J = XJ_{\mathrm{ph},i} - J_{0,i}\left(\exp\frac{qV_i}{n_{\mathrm{ideal},i}kT} - 1\right) \quad \forall\, i, \tag{40.2}$$

where kT/q is the thermal voltage, $n_{\mathrm{ideal},i}$ is the diode ideality factor of each subcell and $J_{0,i}$ is the saturation current density. It is useful to model each junction with two parallel diodes with different ideality factors, and also add parasitic series and shunt resistances. The diodes effectively model the Shockley–Read–Hall (SRH), radiative, and Auger recombination mechanisms, which are active in each junction.

If all three photocurrents $J_{\mathrm{ph},i}$ are equal, then the device's short-circuit current density $J_{\mathrm{SC}} = J_{\mathrm{ph},i}$ and the current density through the diodes is zero. If they are not equal, then $J_{\mathrm{SC}} = \min(J_{\mathrm{ph},i})$, and any excess photocurrent is dissipated through the diodes.

Under open-circuit conditions, the current at the device terminals, and at the interconnections between subcells, is zero. Therefore, all photogenerated current recombines via the diodes.

The short-circuit current density J_{SC} increases linearly with concentration, but there is also an increase in open-circuit voltage V_{OC}, which comes about due to the higher concentration of carriers in the semiconductor and larger separation of the quasi-Fermi levels. Under a concentration factor X, the open-circuit voltage is [4]

$$V_{\mathrm{OC}} = \frac{kT}{q}\left[\sum_i n_{\mathrm{ideal},i}\ln\left(\frac{J_{\mathrm{ph},i}}{J_{0,i}}\right) + \ln X\sum_i n_{\mathrm{ideal},i}\right]. \tag{40.3}$$

If we neglect effects due to parasitic resistances, the fill factor also increases with concentration; consequently, MJSCs generally gain in efficiency as concentration is increased. Of course, this relation does not extend to infinitely high concentration; eventually the assumptions used to derive the diode-like behavior of Equation 40.2 break down, while resistive terms become increasingly important. In such cases, a more complete physics-based device simulation can be useful to study these limits.

The most common commercially available MJSC is a lattice-matched design with triple junctions consisting of $Ga_{0.49}In_{0.51}P$, $In_{0.01}Ga_{0.99}As$, and Ge, respectively [5]. The GaInP and (In)GaAs junctions are clad with a front surface field and a back surface field of a larger bandgap, which helps to confine minority carriers within the junction region. The triple *n-p* junctions (or subcells) are connected in series with *p*++/*n*++ tunnel diodes, which provide an optically transparent, ohmic connection between the *p*-type base of one subcell and the *n*-type emitter of the next subcell. The structure is grown by metal-organic chemical vapor deposition (MOCVD) on a germanium substrate, with the germanium junction being formed by diffusion of group V dopants into the top surface of the substrate during the initial growth of epitaxial material [6,7]. The remaining subcell and tunnel-diode layers are formed by conventional MOCVD growth, requiring ~4 μm of epitaxial material in ~20 layers. A typical layer structure is shown in Figure 40.1a, and some typical 5 × 5-mm cells are shown in Figure 40.1c. This basic structure will be used as an example throughout this chapter.

MJSCs have been simulated with good results using a variety of approaches including lumped-parameter models [2], distributed parameter models [8], and drift-diffusion-based simulators in 1D, 2D, and 3D. In this chapter, we focus primarily on drift-diffusion-based methods. The simulation results shown here were generated using the Sentaurus TCAD package from Synopsys, Inc.; example simulation projects for MJSC devices are available from most of the major simulation software vendors [9–11] and should provide a good starting point.

One of the greatest challenges in simulating these devices is managing the large number of material parameters that are required for each of the materials; this is particularly difficult for ternary and quaternary materials such as AlGaAs and InGaAs for example, where the material composition may be subject to optimization and so material parameters are needed for a wide variety of compositions.

40.2 Device Model and Simulation Workflow

The process of setting up and running a simulation begins by defining each of the material regions (Figure 40.2). In the case of MJSC, this will be a simple structure of uniform layers for the most part; the layers are defined in Figure 40.1a. The exception in our example structure is the germanium subcell, which is formed by diffusion of dopants rather than by epitaxial growth of the emitter. For the germanium subcell, we model the p-type substrate as a single region, with a smoothly varying "error function" profile of n-type doping added to the emitter area. The back surface field at the bottom of the substrate is also modeled in the same way.

Each of the regions is discretized with a mesh, material parameters are assigned to each region, and ohmic contact boundary conditions are defined at the cathode and anode.

From this point, the process differs for QE or for J–V curve calculations. For a QE calculation, additional "virtual contacts" are be defined between each junction, enabling the current from each junction to be extracted separately. An optical calculation is done to determine the optical generation rate as a function of position within the device at the starting wavelength. The Poisson equation is solved to find the potential profile at equilibrium, and then the coupled Poisson and electron and hole continuity equations

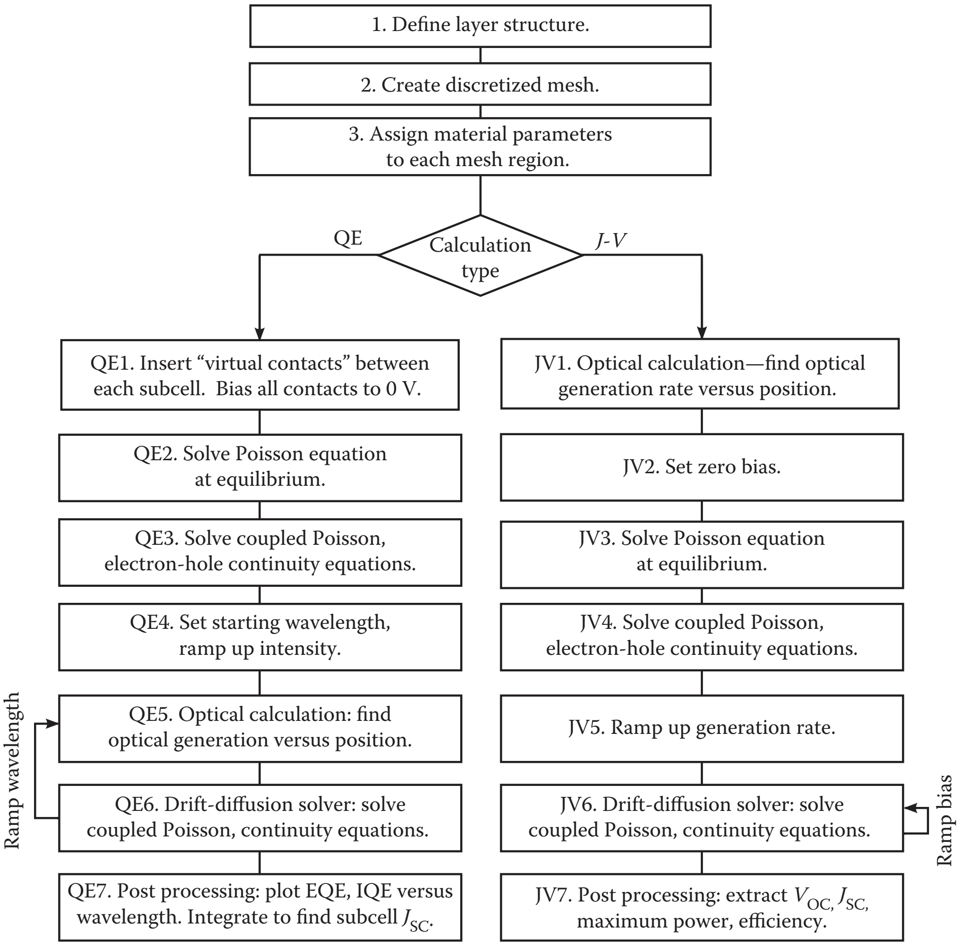

1. Define layer structure.
2. Create discretized mesh.
3. Assign material parameters to each mesh region.
QE
Calculation type
J-V
QE1. Insert "virtual contacts" between each subcell. Bias all contacts to 0 V.
QE2. Solve Poisson equation at equilibrium.
QE3. Solve coupled Poisson, electron-hole continuity equations.
QE4. Set starting wavelength, ramp up intensity.
Ramp wavelength
QE5. Optical calculation: find optical generation versus position.
QE6. Drift-diffusion solver: solve coupled Poisson, continuity equations.
QE7. Post processing: plot EQE, IQE versus wavelength. Integrate to find subcell J_{SC}.
JV1. Optical calculation—find optical generation rate versus position.
JV2. Set zero bias.
JV3. Solve Poisson equation at equilibrium.
JV4. Solve coupled Poisson, electron-hole continuity equations.
JV5. Ramp up generation rate.
JV6. Drift-diffusion solver: solve coupled Poisson, continuity equations.
Ramp bias
JV7. Post processing: extract V_{OC}, J_{SC}, maximum power, efficiency.

FIGURE 40.2 Typical simulation workflow for calculation of quantum efficiency (QE) and current–voltage (J–V) characteristics of multijunciton solar cells.

are solved as the incident optical intensity (and the resulting optical generation rate) is ramped up to the desired value. Finally, the incident wavelength is ramped over the range of interest, with the optical problem being recalculated at each wavelength, and the coupled Poisson-continuity problem is also solved at each wavelength to find the short-circuit current of each junction. This wavelength and short-circuit current data are postprocessed to produce plots of QE.

In the case of a *J*–*V* curve calculation, the optical problem is only solved once, but with an appropriate broadband spectrum. The simulation then starts with a Poisson solution at zero bias. Intensity is gradually ramped up as was the case for the QE calculation, and then the voltage bias is ramped to produce a *J*–*V* curve. Postprocessing is done to identify the maximum power point and extract key parameters, such as open-circuit voltage (V_{OC}), short-circuit current (J_{SC}), fill factor (*FF*), and power conversion efficiency (η).

40.3 Mesh Discretization

Depending on the physical phenomena which are being studied and the symmetry of the device, the region of interest can be discretized with a one-dimensional mesh [12], a 2D mesh representing a small symmetry element such as a half-gridline pitch [13], an entire half of the device, or a 3D model of a portion of the device [14]. The mesh must resolve steps in potential at material heterointerfaces, so ideally mesh vertices along the interfaces should be duplicated with a boundary condition relating the solution variables on each side of the interface (Figure 40.3).

In the direction normal to the interfaces, the minimum required mesh density is typically driven by a need to resolve potential gradients near the interfaces and *p*–*n* junctions. Tunnel diodes in particular have depletions regions of perhaps 10 nm in thickness, and these must be resolved in order for a nonlocal tunneling model to work properly. Far from the interfaces and junctions, the minimum mesh density is driven by the need to resolve standing waves in the optical intensity profile. In the lateral direction, mesh elements can usually be allowed to grow quite large, except at the edges of gridlines and at the device perimeter.

For many studies, a 1D model or a 2D half-gridline pitch option is sufficient to study the phenomena of interest; however, certain phenomena will not be accurately represented. Table 40.1 summarizes the

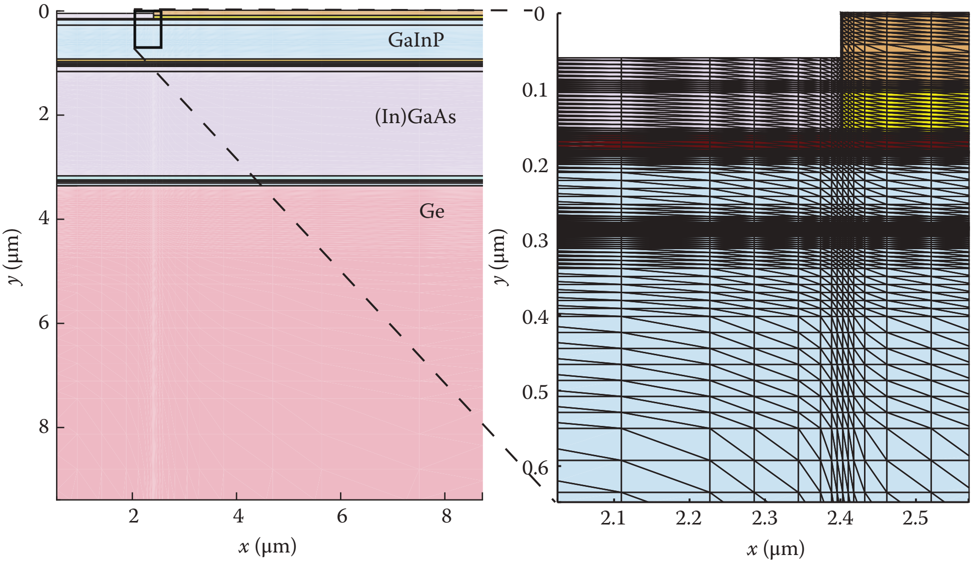


FIGURE 40.3 (a) A 2D, half-grid pitch mesh for the triple-junction solar cell. The full simulation region extends to 60 × 170 μm. (b) Detail of the area near the top contact gridline.

TABLE 40.1 Limitations of Reduced Model Geometries.

	1D	2D Half-Grid Pitch	2D Half-Device	3D Distributed Parameter Model Calibrated Using 2D Simulations	3D
p-n junctions and tunnel diodes	•	•	•	•	•
ARC/window interface		•	•	•	•
Emitter sheet resistance		•	•	•	•
Shading due to gridlines		•	•	•	•
Perimeter recombination			Partial	•	•
Grid ohmic losses				•	•
Nonuniform bias due to grid resistance				•	•
Nonuniform illumination			•	•	•
Busbar regions			Partial	•	•

ARC, anti-reflection coating

limitations of various reduced-dimensional representations. The full 2D and 3D options may be very computationally intensive and will likely be impractical using standard desktop computer hardware; however Létay et al. have demonstrated a method of addressing this by modeling small portions of the device representing the perimeter and the center of the device, linked through an external circuit simulator [13]. Alternatively, a 2D half-grid pitch model can be used to characterize the illuminated I–V performance of a small cell element, and that information can be used to calibrate the parameters of a distributed circuit model [15].

40.4 Optics

Several types of solvers exist which can calculate the carrier generation rate as a function of position within the device layer stack for a given illumination. Raytracing and finite-difference time domain methods [14] can be used, but here we focus on the transfer matrix method (TMM), which is widely used in simulating MJSCs because they have a planar structure and minimal surface texture.

40.4.1 Transfer Matrix Method

In a 2D or 3D simulation including gridlines, we can apply the TMM to calculate the optical generation rate as a function of depth within the layer stack. The propagation of electromagnetic waves through the layer stack is represented by a series of matrices. For each layer there is a matrix describing the reflections from the top interface, and a second matrix representing the phase change and attenuation through the thickness of the layer.

For a stack of N layers, we define phasors representing the amplitude and phase of the y-component of the electric field for downward- and upward-propagating waves at the top and bottom of each layer. The vector

$$\begin{bmatrix} E^{+}_{m,\text{top}} \\ E^{-}_{m,\text{top}} \end{bmatrix} \tag{40.4}$$

contains the phasors describing the electric field at the top of layer m. The symbol + refers to downward propagation (into the cell) and − refers to upward propagation (reflected waves). Layer 0 is the medium above the cell, and layer $N + 1$ is the medium below the cell, which represents the opaque back contact, or air in the case of a bifacial solar cell.

Note that for a plane wave of wavelength λ_0 with an angle of incidence θ_0 with respect to the surface normal, the angle of propagation within any layer m can be found from Snell's equation, $n_0 \sin\theta_0 = n_m \sin\theta_m$. The y-component of the wave vector is then

$$k_{y,m} = \frac{2\pi}{\lambda_0}\frac{n_m}{n_0}\cos\theta_m. \tag{40.5}$$

The phasors at the top and bottom interfaces of a layer are related by a propagation matrix $\mathbf{P}_m$,

$$\begin{bmatrix} E^{+}_{m,\text{top}} \\ E^{-}_{m,\text{top}} \end{bmatrix} = \begin{bmatrix} e^{-jk_{y,m}d_m} & 0 \\ 0 & e^{jk_{y,m}d_m} \end{bmatrix} \begin{bmatrix} E^{+}_{m,\text{bot}} \\ E^{-}_{m,\text{bot}} \end{bmatrix}, \tag{40.6}$$

where j is $\sqrt{-1}$, and d_m is the thickness of layer m.

The phasors at bottom of one layer and the top of the next are are related by

$$\begin{bmatrix} E^{+}_{m-1,\text{bot}} \\ E^{-}_{m-1,\text{bot}} \end{bmatrix} = \frac{1}{t_{m-1,m}} \begin{bmatrix} 1 & r_{m-1,m} \\ r_{m-1,m} & 1 \end{bmatrix} \begin{bmatrix} E^{+}_{m,\text{top}} \\ E^{-}_{m,\text{top}} \end{bmatrix}, \tag{40.7}$$

where $t_{m-1,m}$ and $r_{m-1,m}$ are the Fresnel coefficients for reflection and transmission at the interface. We denote this transmission matrix $\mathbf{T}_{m-1,m}$.

A transfer matrix for the entire layer stack can then be assembled as

$$M = \prod_{m=1}^{N} \left(\mathbf{T}_{m-1,m}\mathbf{P}_m\right)\mathbf{T}_{N,N+1}. \tag{40.8}$$

And finally, we set the electric field incident on the top of the cell, $E^{+}_{0,\text{bot}}$, to any convenient value and write

$$\begin{bmatrix} E^{+}_{0,\text{bot}} \\ E^{-}_{0,\text{bot}} \end{bmatrix} = M \begin{bmatrix} E^{+}_{N+1,\text{top}} \\ 0 \end{bmatrix}. \tag{40.9}$$

This can be solved for the reflected field, $E^{-}_{0,\text{bot}}$, and from that point the total electric field at any point in the structure can be found straightforwardly by assembling a transfer matrix to relate the fields at the top surface with the fields at the point of interest.

This type of calculation is often done assuming that unpolarized light is incident perpendicular to the device, but is also valid for other angles of incidence. The calculation is done as a 1D problem and the resulting carrier generation rate profile is applied to all regions that are not shaded by the grid (Figure 40.4b). Where a broad spectrum illumination is being studied, the TMM is solved repeatedly at each wavelength and the carrier generation rate is summed, weighted according to the incident spectral irradiance [16]. Similarly, illumination from a wide range of angles can be modeled by summing the optical generation calculated from many TMM calculations.

40.4.2 Complex Refractive Index

Solar cell simulations require data on the complex refractive index (refractive index n and extinction coefficient k) over a wide range of wavelengths. This can be a challenge as published data are often limited to a certain range of wavelengths [17]; it may be necessary to measure properties of some materials

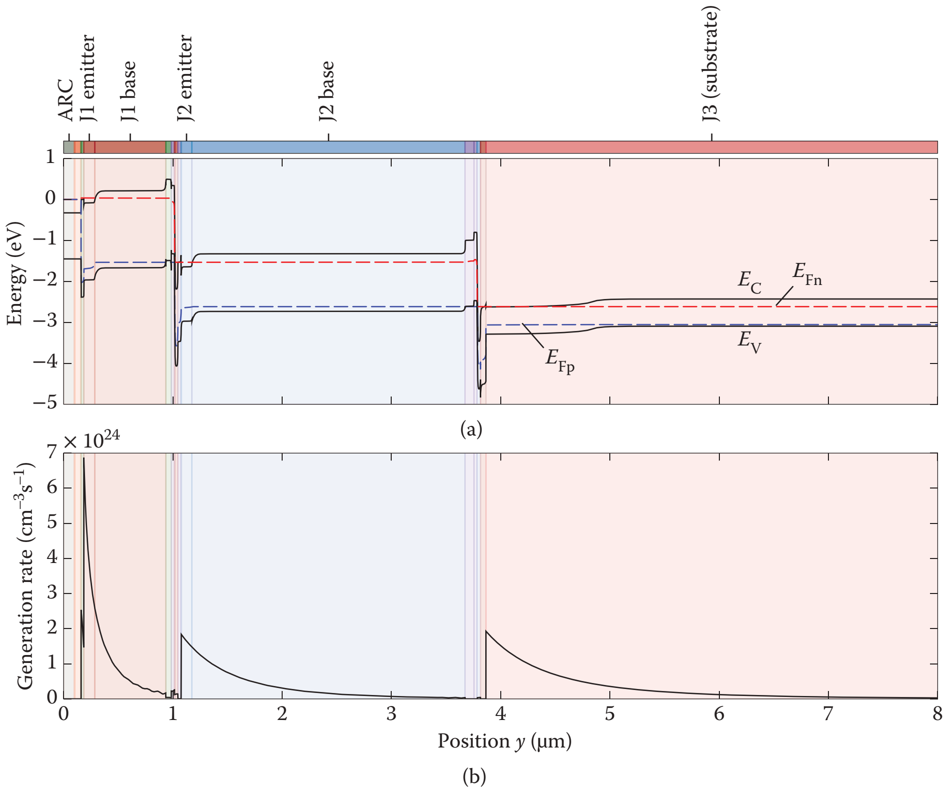


FIGURE 40.4 (a) Simulated band diagram at maximum power point for the triple-junction solar cell. (b) Profile of optical generation rate through the depth of the triple-junction solar cell, as calculated using the TMM. Illumination is broadband (AM1.5D spectrum). (ARC, antireflection coating.)

using ellipsometry or to combine data from several studies to yield a useable dataset. There can be further difficulties due to the use of ternary and quaternary alloys where data are only available for specific compositions.

In some cases, research has been done to model the dielectric function of a material in terms of a model dielectric function (MDF) consisting of a series of Lorentzian oscillators and other terms representing the various band-to-band transitions [18–20]. If the oscillator parameters are estimated as a function of alloy composition, this can be an effective method to generate n,k data for arbitrary alloy compositions. For use in simulating solar cells, care must be taken because these MDF formulations usually yield excessively large k values at photon energies below the bandgap.

40.4.3 Absorption and Quantum Yield

Once the electric field and optical intensity as a function of position within the structure is calculated using the TMM (or any other method), the local absorbed power density is the negative derivative of the optical intensity, $A = -\nabla(\frac{1}{2}\epsilon|\mathbf{E}|^2)$, where ϵ is the dielectric permittivity. If we assume that the quantum yield is unity (i.e., all absorbed photons result in generated carriers), then the local carrier generation rate is

$$G = -\frac{\lambda}{hc}\nabla(\frac{1}{2}\epsilon|\mathbf{E}|^2). \tag{40.10}$$

The assumption of unity quantum yield generally works very well in III–V heterostuctures, but in some materials such as highly doped silicon, free-carrier absorption is significant and will result in a quantum yield less than unity.

The resulting optical generation profile is plotted in Figure 40.4b, for the case of broadband illumination with the AM1.5D spectrum as defined by ASTM G173-03. We can note that within each subcell, there is strong carrier generation near the top of the junction, which then decays quickly through the thickness of the junction. This is to be expected since each subcell is designed to absorb nearly all of the incident light of energy greater than its bandgap, so the generation decays nearly to zero at the bottom of each junction.

40.5 Drift-Diffusion Calculation

Having determined the optical generation rate as a function of position, we then proceed to simulation of the electrical device performance. The electrostatic potential within the device follows the Poisson equation,

$$\nabla \cdot (\epsilon \nabla \phi + \vec{P}) = -q(p - n - N_{\mathrm{A}} + N_{\mathrm{D}}), \tag{40.11}$$

where ϕ is the electrostatic potential, $\vec{P}$ is the polarization which is zero in semiconductors with a zincblende crystal structure, p and n are the free hole and electron concentrations respectively, and N_{A} and N_{D} are the acceptor and donor concentrations. ϵ is the permittivity of the material.

The electron and hole concentrations also follow the continuity equations,

$$\begin{aligned} \nabla \cdot \vec{J}_n &= qR_{\mathrm{net}} + q\frac{\partial n}{\partial t}, \\ -\nabla \cdot \vec{J}_p &= qR_{\mathrm{net}} + q\frac{\partial p}{\partial t}. \end{aligned} \tag{40.12}$$

$\vec{J}_n$ and $\vec{J}_p$ are the electron and hole current densities; R_{net} is the net recombination rate, i.e., the sum of contributions from all recombination and generation processes. At a minimum this will include SRH recombination, Auger recombination, and radiative recombination, as well as the optical generation calculated using the TMM. It can also include additional terms such as band-to-band tunneling and coupled generation, as we discuss below. For studies of solar cells, we are generally interested in steady-state solutions so the time derivaties are zero. The current densities $\vec{J}_n$ and $\vec{J}_p$ can be further expanded into drift and diffusion terms, which depend on the gradient of the local electrical potential and the gradients of the carrier concentrations, the same set of unknowns as in Equation 40.11. Using the Einstein relation, the current densities are [21]

$$\begin{aligned} \vec{J}_n &= -nq\mu_n \nabla E_{\mathrm{F}n}, \\ \vec{J}_p &= -pq\mu_p \nabla E_{\mathrm{F}p}, \end{aligned} \tag{40.13}$$

where μ_n and μ_p are the electron and hole mobilities and $E_{\mathrm{F}n}$ and $E_{\mathrm{F}n}$ are the electron and hole quasi-Fermi levels. Relating the conduction band energy and valence band energy E_{C} and E_{V} to the material properties and the electrostatic potential, we have

$$\begin{aligned} E_{\mathrm{C}} &= -\chi - q(\phi - \phi_{\mathrm{ref}}), \\ E_{\mathrm{V}} &= -\chi - E_{\mathrm{g,eff}} - q(\phi - \phi_{\mathrm{ref}}), \end{aligned} \tag{40.14}$$

where χ is the electron affinity of the material, $E_{\mathrm{g,eff}}$ is the effective bandgap (including the effect of bandgap narrowing), and ϕ_{ref} is an arbitrarily chosen reference potential.

Using Fermi–Dirac statistics for the electron and hole concentrations and making the assumptions that the carrier populations and the crystal are at a constant temperature throughout the device and that the conduction and valence bands have parabolic band structures, the electron and hole quasi-Fermi levels are [21]

$$\begin{aligned} E_{\mathrm{F}n} &= kT\mathcal{F}_{1/2}^{-1}\left(\frac{n}{N_{\mathrm{C}}}\right) + E_{\mathrm{C}}, \\ E_{\mathrm{F}p} &= -kT\mathcal{F}_{1/2}^{-1}\left(\frac{p}{N_{\mathrm{V}}}\right) + E_{\mathrm{V}}, \end{aligned} \tag{40.15}$$

where k is the Boltzmann constant, T is the device temperature, $\mathcal{F}_{1/2}^{-1}$ is the inverse Fermi-Dirac integral of order 1/2, and N_{C} and N_{V} are the effective densities of states in the conduction and valence bands, respectively. By substituting Equations 40.13 through 40.15 into Equation 40.12, we arrive at a formulation of the three coupled differential equations for steady-state problems, stated solely in terms of the material properties and functions of the three variables n, p, and ϕ:

$$\nabla \cdot \epsilon\nabla\phi = -q(p - n - N_{\mathrm{D}} - N_{\mathrm{A}}), \tag{40.16}$$

$$\begin{aligned} \nabla \cdot \left[n\mu_n\nabla\left(kT\mathcal{F}_{1/2}^{-1}\left(\frac{n}{N_{\mathrm{C}}}\right) - \chi - q(\phi - \phi_{\mathrm{ref}})\right)\right] &= -R_{\mathrm{net}}, \\ \nabla \cdot \left[p\mu_p\nabla\left(kT\mathcal{F}_{1/2}^{-1}\left(\frac{p}{N_{\mathrm{V}}}\right) + \chi + E_{\mathrm{g,eff}} + q(\phi - \phi_{\mathrm{ref}})\right)\right] &= -R_{\mathrm{net}}. \end{aligned} \tag{40.17}$$

Typically, we need to solve the coupled set of Equations 40.16 and 40.17 to find the local potential and carrier concentrations throughout the device. Once those quantities are known, then the local electron and hole current densities and the quasi-Fermi levels are also known, and the voltage and current at each electrode can also be found. We are dealing with heterostructures with abrupt changes in the valence and conduction band energies and other quantities at each material interface. To handle these discontinuities, the coupled equations are solved separately in each material region with a set of boundary conditions relating the solution variables on each side of the interfaces. The boundary conditions allow for thermionic emission across or tunneling through small potential barriers at the interfaces [22,23].

It can be difficult to achieve convergence of the coupled Poisson and continuity equations, so we typically begin first with a solution of the Poisson equation under equilibrium conditions (i.e., with all currents and net recombination/generation equal to zero). This provides an initial guess for the solution of the coupled Poisson and continuity equations, again under equilibrium. Next, we gradually ramp up the illumination intensity, leading to a solution at nonequilibrium, and then sweep either the bias at the contacts (to produce a J–V curve) or the illumination wavelength (to produce a QE plot). In the case of QE, the optical generation must be recalculated at each wavelength.

In order to provide physically accurate results, the set of physical models and material parameters must be carefully chosen. Given the wide range of materials and compositions that are present in a single device, managing this database of material parameters is probably the single most difficult part of a MJSC device simulation. Altermatt provides an excellent guide to the parameters needed to do good simulations of silicon solar cells [24]; similar studies are needed for each of the materials in the MJSC structure. Wherever possible, it is advisable to validate simulations of single junctions of each material before attempting to simulate a multijunction device.

40.5.1 Band Structure and Intrinsic Carrier Concentration

While material bandgaps are often well known to a high degree of accuracy, Altermatt makes the point that it is more important to have a self-consistent set of parameters (i.e., bandgap, effective masses, and doping-induced bandgap narrowing) that lead to an accurate value of the effective intrinsic carrier concentration, than to have any one of these individually correct.

If Fermi–Dirac statistics are being used in the simulation, then the experiments used to determine these parameters must also have been based on the same statistics. For the III–V materials commonly used in MJSCs, [25] is a good starting point.

Ternary materials in the $(\mathrm{Al,Ga})_{0.5}\mathrm{In}_{0.5}\mathrm{P}$ system can have varying degrees of Cu:Pt ordering, where Al or Ga atoms and In atoms are found on separate planes of the crystal structure [26]. This means that a range of bandgap values are possible for the same material composition. For example, $\mathrm{Ga}_{0.5}\mathrm{In}_{0.5}\mathrm{P}$ can have a bandgap in the range between 1.8 and 1.9 eV.

In Tables 40.2 through 40.4, we provide the material parameters used for the drift-diffusion-based simulation of the triple-junction solar cell as presented in this chapter. The effective bandgap, $E_{\mathrm{g,eff}}$ at a temperature T is described using a Varshni model [25] for the temperature dependence,

$$E_{\mathrm{g,eff}} = E_{\mathrm{g,ref}} - \frac{\alpha T^2}{T+\beta} + \frac{\alpha T_{\mathrm{ref}}^2}{T_{\mathrm{ref}}+\beta} - \Delta E_{\mathrm{g}}, \tag{40.18}$$

where $E_{\mathrm{g,ref}}$ is the bandgap at a reference temperature T_{ref}, and α and β are the Varshni parameters. ΔE_{g} is a doping-dependent bandgap narrowing term, which we specify using a table of data points. Similarly, the electron affinity χ (difference between the vacuum energy level and the conduction band energy) is described by

$$\chi = \chi_{\mathrm{ref}} + \frac{\alpha T^2}{T+\beta} - \frac{\alpha T_{\mathrm{ref}}^2}{T_{\mathrm{ref}}+\beta} + \Delta E_{\mathrm{g}} \cdot (\Delta\chi/\Delta E_{\mathrm{g}}), \tag{40.19}$$

where the ratio $\Delta\chi \,/\, \Delta E_{\mathrm{g}}$ is specified as a constant.

We calculate the conduction band and valence band densities of states, N_{C} and N_{V}, from the electron and hole effective masses m_n and m_p [21],

$$N_{\mathrm{C}} = 2.5094 \times 10^{19} \left(\frac{m_n}{m_0} \cdot \frac{T}{300\ \mathrm{K}}\right)^{3/2} \mathrm{cm}^{-3}, \tag{40.20}$$

$$N_{\mathrm{V}} = 2.5094 \times 10^{19} \left(\frac{m_p}{m_0} \cdot \frac{T}{300\ \mathrm{K}}\right)^{3/2} \mathrm{cm}^{-3}. \tag{40.21}$$

The effective intrinsic carrier density is then calculated from the band edge densities of states and the effective bandgap,

$$n_{\mathrm{i,eff}} = \sqrt{N_{\mathrm{C}} N_{\mathrm{V}}}\, \exp \frac{-E_{\mathrm{g,eff}}}{2kT}. \tag{40.22}$$

Carrier mobilities μ_n and μ_p are adjusted for temperature- and doping-dependent effects. One of two parameterizations, the “Masetti model” [27] or the “Arora model” [28] are used depending on what data is available and/or which model provides the best fit to data. The Masetti model takes the form

$$\mu = \mu_{\mathrm{min1}} \exp\left(\frac{-P_{\mathrm{c}}}{N_{\mathrm{A}}+N_{\mathrm{D}}}\right) + \frac{\mu_{\mathrm{max}}(T/300\ \mathrm{K})^{\zeta} - \mu_{\mathrm{min2}}}{1+\left[(N_{\mathrm{A}}+N_{\mathrm{D}})/C_{\mathrm{r}}\right]^{\alpha}} - \frac{\mu_1}{1+\left[C_{\mathrm{s}}/(N_{\mathrm{A}}+N_{\mathrm{D}})\right]^{\beta}}, \tag{40.23}$$

where N_{D} and N_{A} are the concentrations of donors and acceptors respectively. The remaining parameters are listed in the parameter tables.

TABLE 40.2 Table of Material Parameters for the GaInP/GaInAs/Ge Solar Cell Example.

	Symbol	**Units**	**$Ga_{0.49}In_{0.51}P$**	**$In_{0.01}Ga_{0.99}As$**	**Germanium**
	ϵ/ϵ_0		12.005	12.91	16.0
			[29]	[29]	[30,31]
	n, k		[20]	[32]	[33]
Band gap	$E_{g,0}$	eV	1.88	1.40	0.744
	α	eV/K	1.82×10^{-4}	5.405×10^{-4}	4.77×10^{-4}
	β	K	81	204	235
	T_{ref}	K	300	300	0
	χ_0	eV	3.924	4.07	3.960
	ΔE_g	eV	[29]	[34]	[29]
	$\Delta\chi/\Delta E_g$		0.5	0.5	0.5
			[35]	[25]	[31]
Density of states	m_n/m_0		0.08515	0.06553	0.5438
	$N_{C,300}$	cm^{-3}	6.235×10^{17}	4.210×10^{17}	1.0063×10^{19}
	m_p/m_0		0.7125	0.5236	0.3406
	$N_{V,300}$	cm^{-3}	1.5093×10^{19}	9.509×10^{18}	4.9883×10^{18}
			[29,35]	[25,35]	[29,35]
Mobility (Masetti model)	μ_{max}	cm^2/V.s		(9400, 491.5)	
	ζ			(2.1, 2.2)	
	μ_{min1}	cm^2/V.s		(500.0, 20.0)	
	μ_{min2}	cm^2/V.s		(500.0, 20.0)	
	μ_1	cm^2/V.s		(0, 0)	
	P_c	cm^{-3}		(0, 0)	
	C_r	cm^{-3}		$(6 \times 10^{16}, 1.48 \times 10^{17})$	
	C_s	cm^{-3}		(0, 0)	
	α			(0.394, 0.38)	
	β			(0, 0)	
				[36]	
Mobility (Arora model)	A_{min}	cm^2/V.s	(400.0, 15.0)		(850.0, 300.0)
	α_m		(0, 0)		(0, 0)
	A_d	cm^2/V.s	(3900, 135)		(2950, 1500)
	α_d		(0, 0)		(0, 0)
	A_N	cm^{-3}	$(2 \times 10^{16}, 1.5 \times 10^{17})$		$(2.6 \times 10^{17}, 1 \times 10^{17})$
	α_N		(1.955, 1.47)		(0, 0)
	A_a		(0.7, 0.8)		(0.56, 1.0)
	α_a		(0, 0)		(0, 0)
			[36]		[31]
Recombination	B_{rad}	cm^3/s	1×10^{-10}	2×10^{-10}	6.4×10^{-14}
	τ_{max}	s	$(2 \times 10^{-8}, 2 \times 10^{-8})$	$(1 \times 10^{-7}, 2 \times 10^{-8})$	(0.001, 0.001)
	τ_{min}	s	(0, 0)	(0, 0)	(0, 0)
	N_{ref}	cm^{-3}	$(1 \times 10^{19}, 1 \times 10^{19})$	$(1 \times 10^{16}, 2 \times 10^{18})$	$(1 \times 10^{16}, 1 \times 10^{16})$
	γ		(1.0, 1.0)	(1.0, 3.0)	(1.0, 1.0)
	T_α	K	(0, 0)	(0, 0)	(0, 0)
	T_{coeff}		(0, 0)	(0, 0)	(0, 0)
	A_A	cm^6s^{-1}	$(3 \times 10^{-30}, 3 \times 10^{-30})$	$(5 \times 10^{-30}, 5 \times 10^{-30})$	$(1 \times 10^{-30}, 1 \times 10^{-30})$
	B_A	cm^6s^{-1}	(0, 0)	(0, 0)	(0, 0)
	C_A	cm^6s^{-1}	(0, 0)	(0, 0)	(0, 0)
	H		(0, 0)	(0, 0)	(0, 0)
	N_0	cm^{-3}	$(1 \times 10^{18}, 1 \times 10^{18})$	$(1 \times 10^{18}, 1 \times 10^{18})$	$(1 \times 10^{18}, 1 \times 10^{18})$
			[35]	[35]	[35]
Nonlocal tunneling	g		(0.21, 0.4)	(0.21, 0.4)	
	m_t/m_0		(0.24, 0.48)	(0.085, 0.34)	
			[37]	[38]	

Note: Paired values in parentheses are specified for electrons and for holes, respectively.

TABLE 40.3 Table of Material Parameters for the GaInP/GaInAs/Ge Solar Cell Example

	Symbol	**Units**	**$Al_{0.5}In_{0.5}P$**	**$Al_{0.3}Ga_{0.7}As$**	**$Al_{0.2}Ga_{0.8}As$**
	ϵ/ϵ_0		11.355	12.055	12.34
			[29]	[29]	[29]
	n, k		[20]	[19]	[19]
Band gap	$E_{g,0}$	eV	2.382	1.7976	1.6726
	α	eV/K	1.76×10^{-4}	4.98×10^{-4}	5.12×10^{-4}
	β	K	134.2	142.8	163.2
	T_{ref}	K	300	300	300
	χ_0	eV	3.75	3.756	3.877
	ΔE_g	eV	[29]	[34]	[34]
	$\Delta\chi/\Delta E_g$		0.5	0.5	0.5
			[25]	[25,35]	[25,35]
Density of states	m_n/m_0		0.23	0.0879	0.0796
	$N_{C,300}$	cm^{-3}	2.768×10^{18}	6.540×10^{17}	5.636×10^{17}
	m_p/m_0		0.36	0.5842	0.5593
	$N_{V,300}$	cm^{-3}	5.4203×10^{18}	1.1206×10^{19}	1.0496×10^{19}
			[25]	[25,35]	[25,35]
Mobility (Masetti model)	μ_{max}	$cm^2/V.s$	(150.0, 180.0)	(3721.1, 240)	(5896.6, 308.21)
	ζ		(1.0, 1.0)	(1.74, 1.83)	(1.81, 1.9)
	μ_{min1}	$cm^2/V.s$	(0, 0)	(195.09, 5.0)	(313.6, 8.8571)
	μ_{min2}	$cm^2/V.s$	(0, 0)	(195.09, 5.0)	(313.6, 8.8571)
	μ_1	$cm^2/V.s$	(0, 0)	(0, 0)	(0, 0)
	P_c	cm^{-3}	(0, 0)	(0, 0)	(0, 0)
	C_r	cm^{-3}	$(5 \times 10^{17}, 2.75 \times 10^{17})$	$(1.16 \times 10^{17}, 1.00 \times 10^{17})$	$(9.33 \times 10^{16}, 1.07 \times 10^{17})$
	C_s	cm^{-3}	(0, 0)	(0, 0)	(0, 0)
	α		(0.436, 0.397)	(0.5758, 0.324)	(0.5152, 0.33425)
	β		(0, 0)	(0, 0)	(0, 0)
			[36]	[36]	[36]
Mobility (Arora model)	A_{min}	$cm^2/V.s$			
	α_m				
	A_d	$cm^2/V.s$			
	α_d				
	A_N	cm^{-3}			
	α_N				
	A_a				
	α_a				
Recombination	B_{rad}	cm^3/s	1×10^{-10}	1×10^{-10}	1×10^{-10}
	τ_{max}	s	$(2.5 \times 10^{-8}, 2.5 \times 10^{-8})$	$(1 \times 10^{-6}, 1 \times 10^{-8})$	$(1 \times 10^{-6}, 1 \times 10^{-8})$
	τ_{min}	s	$(5 \times 10^{-10}, 5 \times 10^{-10})$	(0, 0)	(0, 0)
	N_{ref}	cm^{-3}	$(1 \times 10^{16}, 1 \times 10^{16})$	$(1 \times 10^{16}, 2 \times 10^{18})$	$(1 \times 10^{16}, 2 \times 10^{18})$
	γ		(1.0, 1.0)	(1.0, 3.0)	(1.0, 3.0)
	T_α	K	(0, 0)	(0, 0)	(0, 0)
	T_{coeff}		(0, 0)	(0, 0)	(0, 0)
	A_A	cm^6s^{-1}	$(3 \times 10^{-30}, 3 \times 10^{-30})$	$(1 \times 10^{-30}, 1 \times 10^{-30})$	$(1 \times 10^{-30}\ 1 \times 10^{-30})$
	B_A	cm^6s^{-1}	(0, 0)	(0, 0)	(0, 0)
	C_A	cm^6s^{-1}	(0, 0)	(0, 0)	(0, 0)
	H		(0, 0)	(0, 0)	(0, 0)
	N_0	cm^{-3}	$(1 \times 10^{18}, 1 \times 10^{18})$	$(1 \times 10^{18}, 1 \times 10^{18})$	$(1 \times 10^{18}, 1 \times 10^{18})$
			[35]	[35]	[35]
Nonlocal tunneling	g			(0.21, 0.4)	
	m_t/m_0			(0.09, 0.37)	
				[38]	

Note: Paired values in parentheses are specified for electrons and for holes, respectively.

TABLE 40.4 Table of Material Parameters for the GaInP/GaInAs/Ge Solar Cell Example.

	Symbol	**Units**	**$Al_{0.25}Ga_{0.25}In_{0.5}P$**	**MgF**	**TiOx**
	ϵ/ϵ_0		11.84	5.3	6
			[29]	[39]	[40]
	n, k		[20]	[41]	[42,43]
Band gap	$E_{g,0}$	eV	1.984		
	α	eV/K	1.76×10^{-4}		
	β	K	134.2		
	T_{ref}	K	300		
	χ_0	eV	3.997		
	ΔE_g	eV	[29]		
	$\Delta\chi/\Delta E_g$		0.5		
			[25]		
Density of states	m_n/m_0		0.1214		
	$N_{C\,300}$	cm^{-3}	1.061×10^{18}		
	m_p/m_0		0.6244		
	$N_{V\,300}$	cm^{-3}	1.24×10^{19}		
			[29,35]		
Mobility (Masetti model)	$\mu_{\max}$	$cm^2/V.s$			
	ζ				
	$\mu_{\min1}$	$cm^2/V.s$			
	$\mu_{\min2}$	$cm^2/V.s$			
	μ_1	$cm^2/V.s$			
	P_c	cm^{-3}			
	C_r	cm^{-3}			
	C_s	cm^{-3}			
	α				
	β				
Mobility (Arora model)	$A_{\min}$	$cm^2/V.s$	(400.0, 15.0)		
	α_m		(0, 0)		
	A_d	$cm^2/V.s$	(3900, 135)		
	α_d		(0, 0)		
	A_N	cm^{-3}	$(2 \times 10^{16}, 1.5 \times 10^{17})$		
	α_N		(1.955, 1.47)		
	A_a		(0.7, 0.8)		
	α_a		(0, 0)		
			[36]		
Recombination	B_{rad}	cm^3/s	1×10^{-10}		
	$\tau_{\max}$	s	$(1 \times 10^{-9}, 1 \times 10^{-9})$		
	$\tau_{\min}$	s	(0, 0)		
	N_{ref}	cm^{-3}	$(1 \times 10^{19}, 1 \times 10^{19})$		
	γ		(1.0, 1.0)		
	T_α	K	(0, 0)		
	T_{coeff}		(0, 0)		
	A_A	cm^6s^{-1}	$(3 \times 10^{-30}, 3 \times 10^{-30})$		
	B_A	cm^6s^{-1}	(0, 0)		
	C_A	cm^6s^{-1}	(0, 0)		
	H		(0, 0)		
	N_0	cm^{-3}	$(1 \times 10^{18}, 1 \times 10^{18})$		
			[35]		
Nonlocal tunneling	g				
	m_t/m_0				

Note: Paired values in parentheses are specified for electrons and for holes, respectively.

TABLE 40.5 Table of Interface Parameters for the GaInP/GaInAs/Ge Solar Cell Example.

	Symbol	Units	TiOx/AlInP	AlInP/GaInP	GaAs/AlGaAs	GaInP/Ge
Interface recombination	S_0	cm/s	0 fitted	$(2 \times 10^5, 2 \times 10^5)$ [44]/fit to data	(200, 200) [44]	$(1 \times 10^5,\ 1 \times 10^5)$ [6]/fit to data

Note: Paired values in parentheses are specified for electrons and for holes respectively.

The Arora model for mobility uses a set of temperature-corrected constants,

$$N_0 = A_N \left(\frac{T}{300\ \mathrm{K}}\right)^{\alpha_N} ; A^* = A_\mathrm{a} \left(\frac{T}{300\ \mathrm{K}}\right)^{\alpha_\mathrm{a}} \tag{40.24}$$

$$\mu_\mathrm{min} = A_\mathrm{min} \left(\frac{T}{300\ \mathrm{K}}\right)^{\alpha_\mathrm{m}} ; \mu_\mathrm{d} = A_\mathrm{d} \left(\frac{T}{300\ \mathrm{K}}\right)^{\alpha_\mathrm{d}} . \tag{40.25}$$

Mobility is then calculated with

$$\mu = \mu_\mathrm{min} + \frac{\mu_\mathrm{d}}{1 + \left[(N_\mathrm{A} + N_\mathrm{D})/N_0\right]^{A^*}} . \tag{40.26}$$

40.5.2 Recombination, Lifetimes, and Surface/Interface Recombination Velocities

Having defined some of the basic material properties, we can now look at the recombination rates, each of which have their own material parameters. Here we present the rates as they are implemented in the Sentaurus software [21]; similar formulas can be found in texts on device simulation such as Palankovski [31]. The radiative recombination rate can be written as

$$R_\mathrm{rad} = B_\mathrm{rad}(np - \gamma_n \gamma_p n_\mathrm{i,eff}^2), \tag{40.27}$$

where γ_n and γ_p are correction factors which are needed for simulations using Fermi–Dirac statistics, and are defined by

$$\gamma_n = \frac{n}{N_\mathrm{C}} \exp \frac{E_\mathrm{C} - E_{\mathrm{F}n}}{kT}, \tag{40.28}$$

$$\gamma_p = \frac{p}{N_\mathrm{V}} \exp \frac{E_{\mathrm{F}p} - E_\mathrm{V}}{kT}. \tag{40.29}$$

The SRH recombination rate is somewhat more complicated. We begin by finding the doping-dependent minority carrier lifetimes for the SRH process,

$$\tau_\mathrm{SRH} = \tau_\mathrm{min} + \frac{\tau_\mathrm{max} - \tau_\mathrm{min}}{1 + \left(\frac{N_\mathrm{A} + N_\mathrm{D}}{N_\mathrm{ref}}\right)^{\gamma}}. \tag{40.30}$$

This calculation will be done separately using different parameters for the electron and hole minority carrier lifetimes. Then we can write the expression for the SRH recombination rate,

$$R_\mathrm{SRH} = \frac{np - \gamma_n \gamma_p n_\mathrm{i,eff}^2}{\tau_{\mathrm{SRH},p}(n + \gamma_n n_1) + \tau_{\mathrm{SRH},n}(p + \gamma_p p_1)}, \tag{40.31}$$

where n_1 and p_1 are

$$n_1 = n_{\mathrm{i,eff}} \exp\left(\frac{E_{\mathrm{trap}}}{kT}\right); \quad p_1 = n_{\mathrm{i,eff}} \exp\left(\frac{-E_{\mathrm{trap}}}{kT}\right). \tag{40.32}$$

Note that we have not included a temperature dependence of the lifetimes, or any enhancement of the SRH process due to electric field, but these could be significant in some devices. E_{trap} is the position of the trap level relative to the intrinsic level; we generally take it to be zero.

Finally, we write the Auger recombination rate,

$$R_{\mathrm{Auger}} = (C_n n + C_p p)(np - n_{\mathrm{i,eff}}^2). \tag{40.33}$$

Due to the long nonradiative lifetimes in the bulk of many III–V materials and the relatively thin layers, interface recombination can be a significant factor in the performance of these devices. The values of surface recombination velocity can vary widely due to various growing conditions, and so this along with the SRH lifetime can be treated as a “fitting parameter” rather than a known material property [44]. The surface recombination rate is treated with an SRH model,

$$R_{\mathrm{surf}} = \frac{np - \gamma_n \gamma_p n_{\mathrm{i,eff}}^2}{(n + \gamma_n n_1)/S_p + (p + \gamma_p p_1)/S_n}, \tag{40.34}$$

where S_n and S_p are the electron and hole surface recombination velocities.

40.5.3 Tunnel Diodes

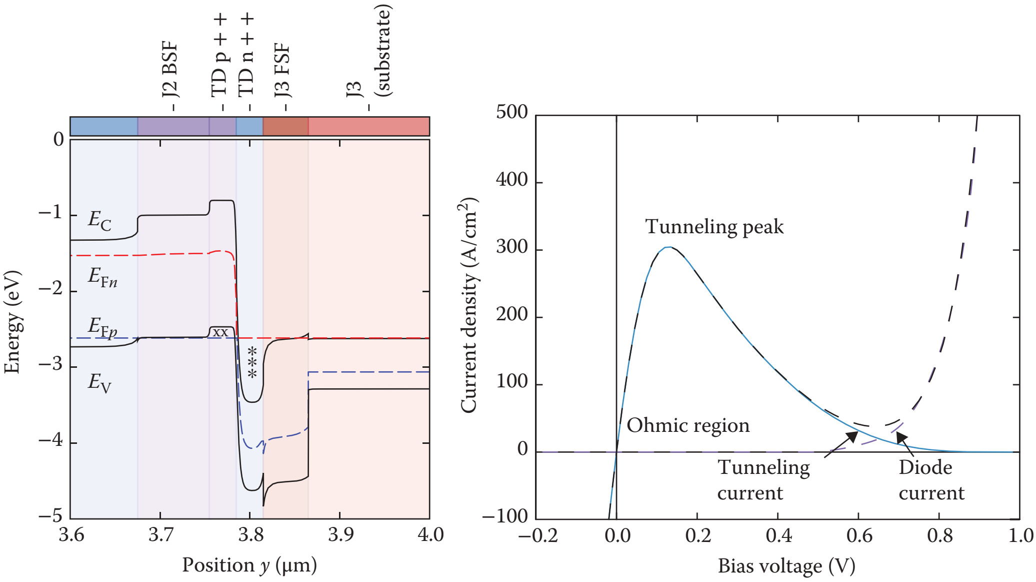
The tunnel diodes that allow for the flow of carriers between subcells in MJSCs operate via band-to-band tunneling across a degenerately doped p++/n++ junction [45]. The depletion width in this case will be extremely narrow (~5 nm), with carrier populations at similar energy levels on both sides of the junction (Figure 40.5a). When a small bias is applied, the free electrons become aligned in energy with holes on the opposite side, and can tunnel across the junction, creating a net current.

A typical J–V curve of a tunnel diode is shown in Figure 40.5b. Under normal conditions, the tunnel diode should remain in the ohmic region, where large currents can be conducted with minimal potential difference across the tunnel diode. This requires that the peak tunneling current density should be significantly greater than the solar cell’s short-circuit current density J_{SC}. If the bias voltage is increased beyond the ohmic region, the free electrons and holes are no longer aligned in energy and the tunneling current decreases. Eventually, the diode will be forward biased and will conduct via normal mechanisms.

Many drift-diffusion-based simulators allow for modeling band-to-band tunneling using approaches derived from the work of Kane et al. [46] and others [47,48]. These models can reproduce all of the behavior of a tunnel junction [37,38,49] including the ohmic region, tunneling peak and excess current region of the tunnel diode J–V curve [30].

Tunnel diodes typically have minimal impact on the performance of MJSCs, except in pathological cases or where the device is operated at very high current density. For many studies, a simplistic band-to-band tunneling model based only on local quantities will be sufficient. For example, in Synopsys Sentaurus, there is a model that adds a band-to-band recombination term to the continuity equation following the work of Hurkx [47],

$$R^{\mathrm{bb}} = A \, \frac{np - n_{\mathrm{i,eff}}^2}{(n + n_{\mathrm{i,eff}})(p + n_{\mathrm{i,eff}})} \left(\frac{|F|}{1\mathrm{V/cm}}\right)^{\sigma} \exp\left(-\frac{F_0}{|F|}\right), \tag{40.35}$$



(a) (b)

FIGURE 40.5 (a) Detail of the bottom tunnel-diode portion of the cross-sectional band diagram. The tunnel diode is short circuited. "xx" symbols represent holes and "*" symbols represent electrons that are available to tunnel across the junction into the opposite band. (b) J–V curve of a tunnel diode. Contributions from band-to-band tunneling and the diode diffusion current are shown.

where n and p are the electron and hole concentrations, $n_{\mathrm{i,eff}}$ is the effective intrinsic carrier concentration, and F is the local electric field strength. σ has a value of 2 for direct transitions and 5/2 for indirect transitions. A and F_0 are material-dependent parameters that must be determined experimentally. This type of model should be adequate to give good results in cases where tunnel diodes are not dominating the behavior of an MJSC, but because it does not consider the variation of the band profile and material properties across the tunneling path, it will not be physically accurate.

A more accurate, but also more computationally intensive, approach uses a nonlocal band-to-band tunneling model combined with trap-assisted tunneling [50]. A "nonlocal" model is one where the tunneling current at a given position is calculated based on quantities at other locations, such as the potential profile across the tunneling path, and carriers are effectively removed from one location in the structure and reinserted at a different location. The use of this type of nonlocal model is described in detail in [38] and [51]. Implementation in Sentaurus is shown in Reference [52]. Figure 40.5b was simulated using nonlocal models for band-to-band tunneling as well as SRH recombination enhanced by trap-assisted tunneling. The individual contributions from each of these two mechanisms is shown in the figure. Some typical values of the tunneling masses m_t and effective Richardson constants g are included in Tables 40.2 and 40.3. Readers should refer to references [38,50–52] for interpretation of these parameters.

It is our experience that, while it it very possible to fit tunnel-diode parameters to a measured J–V curve, the resulting model tends not to be predictive for varying tunnel junction designs (i.e., varying doping levels or material compositions). The peak tunneling current of devices can vary by several orders of magnitude for relatively small changes in doping or material composition.

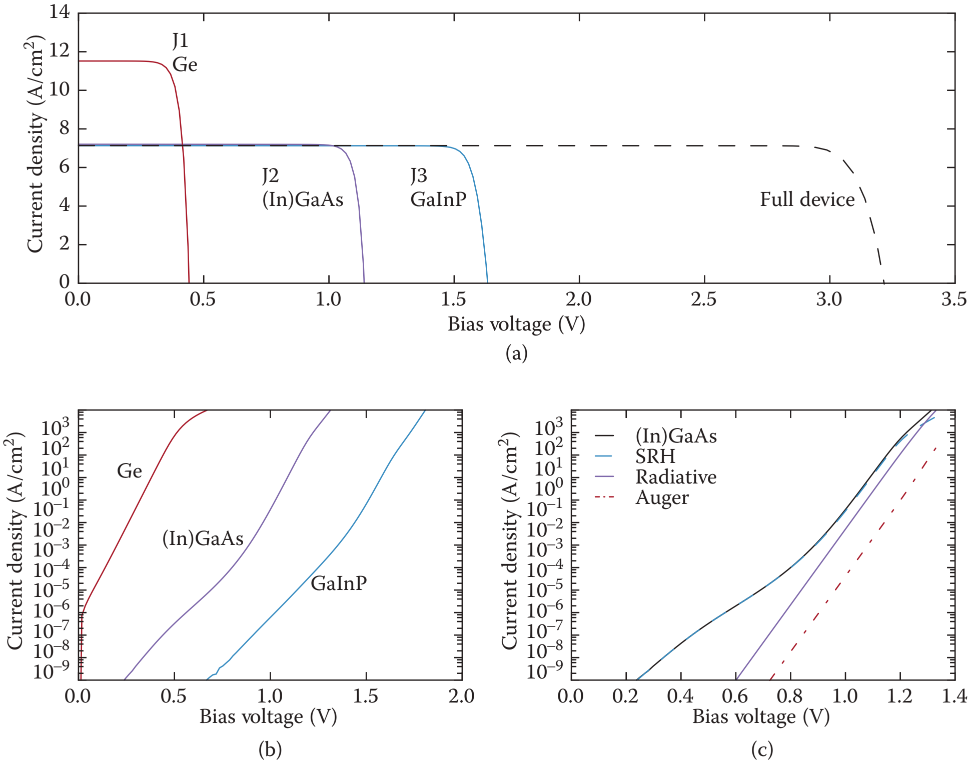
40.6 Current–Voltage Calculation

A J–V curve calculation is performed by first calculating the optical generation as a function of position for the appropriate illumination conditions, then solving the drift-diffusion problem while ramping bias at the contacts. It can also be useful to generate *J*–*V* curves for individual subcells. Additional “virtual” ohmic contact boundary conditions are defined at locations in between the subcells in order to bias a subcell individually. The current through a given subcell is then the sum of currents at all contacts above it (Figure 40.6).

It’s very informative to study the subcell dark *J*–*V* curves on a log-current axis, as the contributions from various physical processes can be distinguished [2]. In Figure 40.6c, the dark *J*–*V* curve of the (In)GaAs subcell is plotted along with the recombination currents due to SRH, radiative, and Auger mechanisms. The recombination currents are calculated as

$$J_i = \frac{q}{A} \int_V R_i \, \mathrm{d}V, \tag{40.36}$$

where q is the electron charge, R is the recombination rate per unit volume, A is the cell area and i refers to the recombination process in question. The integral is evaluated over the volume of a subcell. Comparing simulated dark-current curves against measured single-junction devices can be a good first step to validating multijunction simulations. When plotted on a log axis, the slope of the curve is directly related to the diode ideality factor. In particular, the SRH mechanism usually produces a region where the ideality factor $n_{\text{ideal}} \simeq 2$; this gives an excellent method to verify that the SRH current is modeled accurately independent of other factors. In contrast, fitting the SRH lifetimes based on V_{OC} measured at high intensity (1 sun or greater) would not be effective as the cell is not strongly dominated by the SRH process at that level of bias.



(a)

(b)

(c)

FIGURE 40.6 (a) *J*–*V* curves of the subcells and the triple-junction device under 500 suns, AM1.5D illumination. (b) Dark current curves of each of the subcells. (c) Dark *J*–*V* curve of (In)GaAs with the contributions from individual processes.

40.7 Quantum Efficiency Calculation

Experimentally, QE of multijunction cells is measured by using bias lights to ensure that the subcell of interest is limiting the overall device current, as described in Reference [53]. In this way, it is possible to probe each of the subcells individually. This experiment can also be replicated in simulation by adding bias illumination sources, or simply adding a fixed amount of carrier generation to the non-limiting subcells, in addition to a monochromatic source, which is scanned in wavelength over the range of interest. The bias light intensity must be adjusted such that the short-circuit currents of the non-probed junctions are greater than the current produced by the probed junction at any wavelength; this requires some iteration to find the correct bias intensities.

An alternative approach that can be implemented in simulation is to simply define ohmic “virtual contacts” to the front surface field and back surface field layers of each subcell. This provides a more direct control over the voltage bias being applied to each subcell than the light-biasing method, but should generally yield identical results. It also has advantages over the light-biasing method in that there is no iteration required to find the correct setup, and where the light-biasing method requires a separate sweep over wavelength for each junction, with the “virtual contact” method the short-circuit currents and QE plots of all triple junctions can be collected in one wavelength sweep. The simulated QE for the triple-junction solar cell, calculated using this method, is presented in Figure 40.7. The simulation will output short-circuit current density as a function of wavelength λ for subcell i, $J_{SC,i}$. From the optical calculation, we also know the reflectivity from the top of the layer stack, $R(\lambda)$ and transmissivity through the layer stack, $T(\lambda)$. The external quantum efficiency (EQE) of a given subcell is defined as the ratio of short-circuit current density to incident photon flux density $\Phi(\lambda)$,

$$EQE_i(\lambda) = \frac{J_{SC,i}(\lambda)}{q\Phi(\lambda)}, \tag{40.37}$$

or in terms of the optical intensity $E(\lambda)$,

$$EQE_i(\lambda) = \frac{hc}{q\lambda}\frac{J_{SC,i}(\lambda)}{E(\lambda)}, \tag{40.38}$$

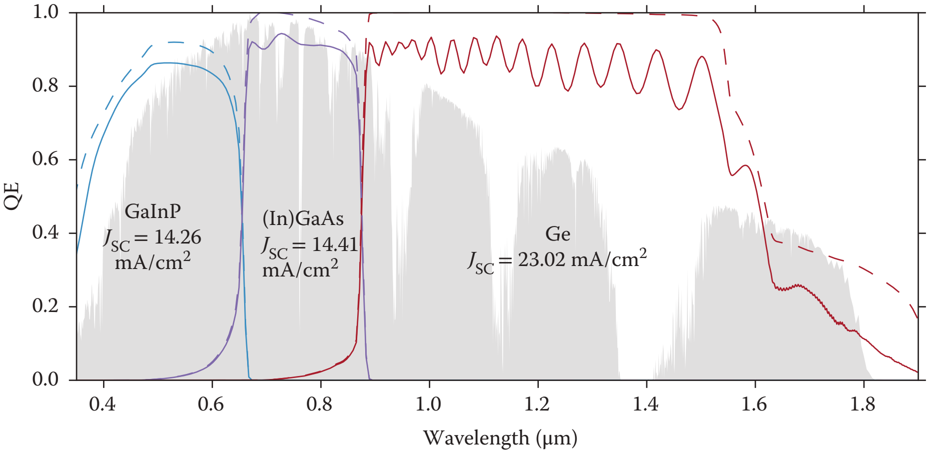


FIGURE 40.7 External (solid) and internal (dashed) quantum efficiency (EQE and IQE) of the triple-junction solar cell. The normalized AM1.5D solar spectral irradiance is shown filled in gray.

where h is Planck's constant, c is the speed of light and q is the electron charge. Internal quantum efficiency (IQE) is given by

$$IQE_i(\lambda) = \frac{EQE_i(\lambda)}{1 - R(\lambda) - T(\lambda)}. \tag{40.39}$$

QE curves as calculated in software often show Fabry–Pérot type oscillations at long wavelengths which do not appear in physical measurements. Frequently these oscillations are physically correct for a perfectly collimated beam, but in an experimental setup with a finite beam divergence and monochromator resolution these fine details cannot be resolved. In order to make meaningful comparisons between measured and simulated data, it can therefore be useful to convolve the simulated QE with a spectral function approximating the monochromator lineshape, thus giving a representation of the simulated QE as it would be measured by an instrument. We can write this effective QE as

$$EQE_{\text{eff}}(\lambda) = \int_{-\infty}^{\infty} EQE(\lambda + l)w(l)\mathrm{d}l, \tag{40.40}$$

where $EQE(\lambda)$ is the simulated QE, $w(\lambda)$ is a function approximating the spectral lineshape of the monochromator which is centered at $\lambda = 0$ and integrates to unity, and l is a variable of integration.

Besides calculating the experimentally measurable quantities R, T, EQE, and IQE, simulation also provides the opportunity for introspection – examining quantities, which would not be measurable. We can integrate various recombination rates over the volume of the device to quantify each of the loss mechanisms and explain any nonideality in the simulation. For example, integrating the SRH recombination rate over the volume of subcell i, we find a loss component

$$L_{\text{SRH},i}(\lambda) = \frac{1}{A\Phi(\lambda)} \int_{V_i} R_{\text{SRH}}(\lambda)\mathrm{d}V. \tag{40.41}$$

Similarly for interface recombination at the interfaces above and below subcell i,

$$L_{\text{surf},i}(\lambda) = \frac{1}{A\Phi(\lambda)} \left(\int_{A_{\text{em},i/\text{fsf},i}} R_{\text{surf}}\mathrm{d}A + \int_{A_{\text{base},i/\text{bsf},i}} R_{\text{surf}}\mathrm{d}A \right), \tag{40.42}$$

where the integrals are evaluated over the area of the front surface field/emitter and base/back surface field interfaces of the subcell in question. In general, it is possible to account for all EQE losses in this way, in which case the sum of all subcell EQEs and all loss components should be unity (Figure 40.8). Just as the EQE can be integrated with incident spectrum to give J_{SC} of each subcell, the loss components can be integrated to indicate the amount of photocurrent lost via a particular mechanism.

40.8 Photon Recycling and Luminescent Coupling

Some MJSCs show significant effects due to photon recycling, which can enhance the voltage of individual subcells, and luminescent coupling which can redistribute photocurrent between junctions and reduce sensitivity to spectrum [54,55].

To understand this, we can modify the circuit model of Figure 40.1b with two diodes per subcell, one of which represents the nonradiative recombination current and one the radiative current, $J_{\text{rad},i}$ (Figure 40.9). A portion of the radiatively emitted current results in generation within other junctions, according to

$$J_{\text{LC},i} = \sum_j K_{i,j} J_{\text{rad},j}, \tag{40.43}$$

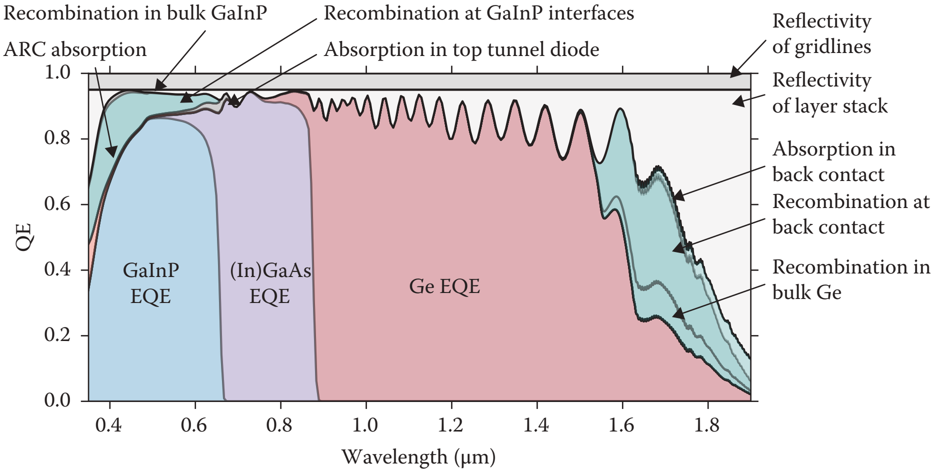


FIGURE 40.8 Quantum efficiency plot showing the breakdown of losses. (ARC, antireflection coating; EQE, external quantum efficiency.)

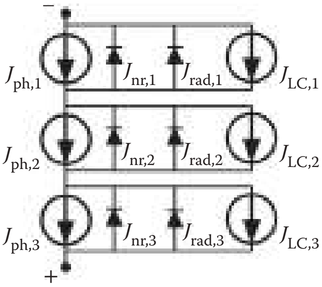


FIGURE 40.9 Equivalent circuit model of a triple-junction cell, modified to distinguish radiative and nonradiative diode currents and to include coupled generation.

where $K_{i,j}$ is a junction-to-junction coupling coefficient. If any of the photocurrents $J_{ph,i}$ are in excess of the device current J, that excess current will recombine through the nonradiative and radiative diodes in Figure 40.9. At sufficiently high bias, GaAs and GaInP junctions have predominantly (>90%) radiative recombination. These junctions are thick enough that the self-coupling coefficient $K_{i,i}$ can also be >90%, meaning that carriers can cycle repeatedly through the process of radiative recomination, photon emission, and reabsorption within the same junction. This process is called "photon recycling" and can be responsible for an increase in carrier concentrations.

A related term, "luminescent coupling," refers to the coupling between different junctions. If one of the bottom junctions is current limiting (i.e., it has the samllest J_{ph} of any subcell), then the excess current in the upper junctions can be radiatively coupled into the bottom junction, allowing a J_{SC} greater than the photocurrent in the limiting junction.

Luminescent coupling can also be observed in experimental measurements of multijunction cell QE. When bias lights are applied to the top two subcells, some of that bias light is coupled into the subcell that is being probed, interfering with the measurement [56].

These effects can be modeled in a drift-diffusion device simulator with the addition of a coupled generation term to the continuity equation [57,58], or through a process of successive approximations to the optical generation function [59].

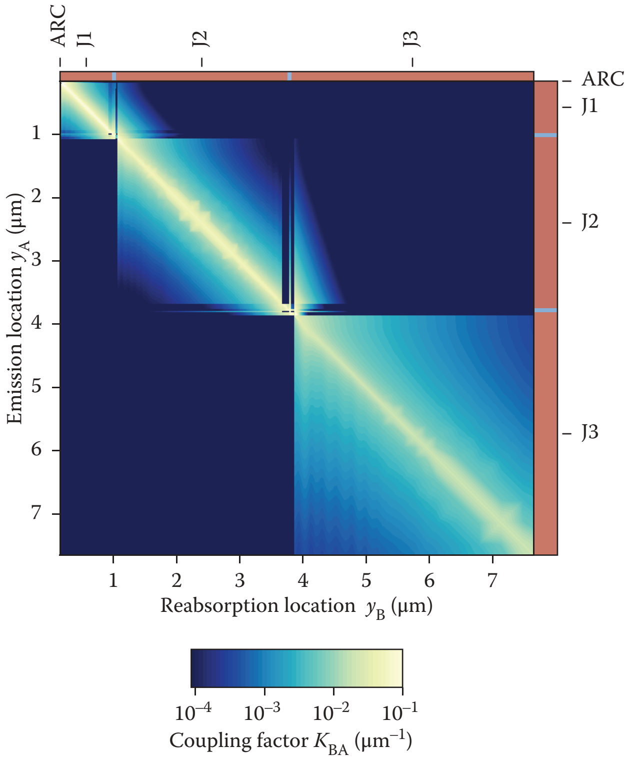


FIGURE 40.10 Luminescent coupling matrix for the triple-junction GaInP/(In)GaAs/Ge example structure. (ARC, antireflection coating.)

We discuss the first option here, and begin by modifying the process flow outlined in Figure 40.2 and insert a step after the assignment of material parameters for calculation of the coupling matrix, which gives the probability that a photon emitted from a position A will be reabsorbed at another position, B (shown in Figure 40.10). Naturally, photons have a great probability of being reabsorbed near the point where they are emitted (i.e., along the main diagonal of the matrix), or in layers of smaller bandgap such as a lower subcell. Indeed, the figure shows significant absorption in the GaAs junction (J2) due to emissions from J1, and also absorption in J3 from emissions in J2. This coupled transfer is unidirectional; there is no coupling of emitted photons from the lower subcells into the upper ones. Also, it should be noted that Figure 40.10 indicates only the efficiency with which emitted photons are reabsorbed. Some layers may not emit strongly at all depending on the local carrier concentration and the ratio of radiative and nonradiative lifetimes.

Once the coupling matrix has been calculated, the simulation can proceed as normal with a modification to the electron and hole continuity equations to add the coupled generation term which is a function of radiative recombination rates at all positions within the device. While the simulator used here, Synopsys Sentaurus, does not directly provide a method to add this term, it does provide a facility to define custom generation/recombination terms through a C++ programming interface. The custom C++ code is executed at each iteration of the nonlinear equation solver, and has access to quantities such as carrier concentrations (and by extension, recombination rates) at all mesh vertices within the simulation domain. The technique is described in detail in the appendix to Reference [58].

The result of this calculation can be seen in Figure 40.11; the "plateau" in QE between 650 and 880 nm is due to light that is radiatively coupled from the (In)GaAs junction into the germanium junction, and is an undesirable effect as it interferes with measuring the actual germanium photocurrent $J_{ph,3}$.

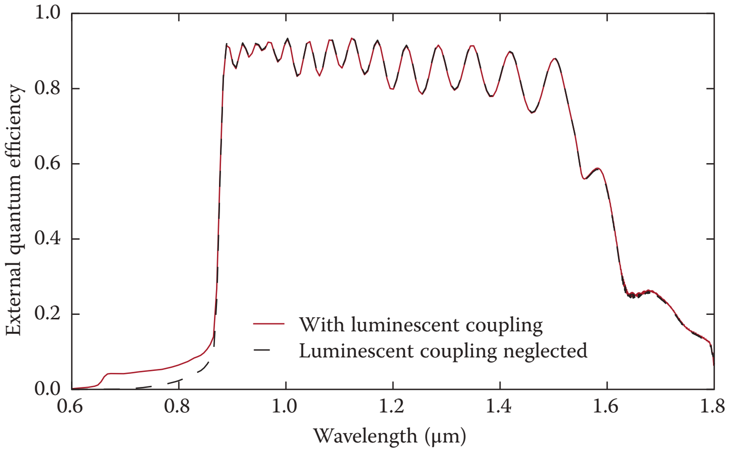


FIGURE 40.11 Calculated quantum efficiency of the germanium junction including coupled generation from light emitted by the (In)GaAs junction.

In multijunction phototransducers, which are devices closely related to MJSCs but are designed to operate under single-wavelength illumination, the luminescent coupling causes a significant broadening of the wavelength response. This broadening effect has also been modeled using this technique[58,60].

In contrast with the circuit-model representation, this treatment of luminescent coupling using a drift-diffusion simulation provides some advantages. The radiative emission rate, and the resulting coupled generation, is calculated as a function of position and reflects local changes in carrier density. The rates of emission from the emitter, the space charge region, and the base will each vary depending on the bias conditions, and the simulation will handle these changes appropriately. Areas that are shaded by gridlines and busbars will also be treated correctly. Furthermore, it is straightforward to consider coupling between all layer combinations in the device structure, not just those layers that comprise the photovoltaic junctions.

40.9 Summary

MJSCs are relatively complex devices, with typically 25 or more heterointerfaces and a number of photovoltaic junctions and tunnel diodes. With sufficient calibration, models can be developed that provide a very accurate representation of the device performance and can indicate the relative importance of various loss mechanisms. Once the model is calibrated, a detailed design of experiments can be done to pinpoint optimal design for a given requirement. In order to do accurate simulations, an extensive set of material parameters must be collected for each of the semiconductor materials. In particular, the complex refractive index, the minority carrier lifetimes, and the tunneling parameters tend to be difficult to find in the literature and need to be found through experiment.

In many cases, lateral current spreading and uneven bias across the surface of the cell have significant impacts on efficiency, which cannot be represented in a simple equivalent circuit model. In concentrating systems that have chromatic aberrations, the illumination of each subcell will have a different, nonuniform distribution over the the cell aperture. This leads to local mismatch in the subcell current densities and is an interesting problem to study through device simulation. Simulations based on drift-diffusion enable the device to be simulated including transport across all heterojunctions, without any presupposition of diode-like behavior, and hence this type of model can be used to study the effects of extreme operating conditions or manufacturing errors where the diode model is not applicable.

References

1. M. A. Green, K. Emery, Y. Hishikawa,W. Warta, and E. D. Dunlop. Solar cell efficiency tables (version 47). *Prog. Photovolt. Res. Appl.*, 24(1):3–11, 2016.
2. J. Nelson. *Physics of Solar Cells*. London: Imperial College Press, 2003.
3. C. H. Henry. Limiting efficiencies of ideal single and multiple energy gap terrestrial solar cells. *J. Appl. Phys.*, 51(8):4494, 1980.
4. R. R. King, D. Bhusari, A. Boca, D. Larrabee, X. Liu, W. Hong, C. M. Fetzer, D. C. Law, and N. H. Karam. Band gap-voltage offset and energy production in next-generation multijunction solar cells. *Prog. Photovolt. Res. Appl.*, 19:797–812, 2011.
5. R. R. King, N. H. Karam, J. H. Ermer, N. Haddad, P. Colter, T. Isshiki, H. Yoon, H. L. Cotal, D. E. Joslin, D. D. Krut, R. Sudharsanan, K. Edmondson, B. T. Cavicchi, and D. R. Lillington. Next-generation, high-efficiency III-V multijunction solar cells. In *Conference Record 28th IEEE Photovoltaic Specialists Conference*., pages 998–1001, Anchorage, AK, 2000. IEEE.
6. D. J. Friedman and J. M. Olson. Analysis of Ge junctions for GaInP/GaAs/Ge three-junction solar cells. *Prog. Photovolt. Res. Appl.*, 9(3):179–189, 2001.
7. D. J. Friedman, J. M. Olson, S. Ward, T. Moriarty, K. Emery, S. Kurtz, A. Duda, R. R. King, H. L. Cotal, D. R. Lillington, J. H. Ermer, and N. H. Karam. Ge concentrator cells for IIIV multijunction devices. In *Conference Record 28th IEEE Photovoltaic Specialists Conference*, Anchorage, AK, pages 965–967. IEEE, 2000.
8. I. García, C. Algora, I. Rey-Stolle, and B. Galiana. Study of non-uniform light profiles on high concentration III-V solar cells using quasi-3D distributed models. In *2008 33rd IEEE Photovolatic Specialists Conference*, San Diego, CA, pages 1–6. IEEE, May 2008.
9. Synopsys Inc. Simulation of a GaAs / GaInP Dual-Junction Solar Cell, Synopsys Inc., Mountain View, CA, 2012.
10. Solarex17.in: EQE of III-V Tandem Cell. Silvaco Inc., Santa Clara, CA, 2015. http://www.silvaco.com/examples/tcad/section44/example17/index.html
11. Z. Q. Li, Y. G. Liao, and Z. M. S. Li, Modelling of multi-junction solar cells by Crosslight APSYS, High and Low Concentration for Solar Electric Applications, *Proceedings of the SPIE*, vol. 6339, pp. 633909–15, San Diego, CA, 2006.
12. W. E. McMahon, J. M. Olson, J. F. Geisz, and D. J. Friedman. An examination of 1D solar cell model limitations using 3D SPICE modeling. In *38th IEEE Photovoltaic Specialists Conference*., pp. 002088–002091. Austin, TX, Jun 2012.
13. G. Létay, M. Hermle, and A. W. Bett. Simulating single-junction GaAs solar cells including photon recycling. *Prog. Photovolt. Res. Appl.*, 14(8):683–696, 2006.
14. T. Rahman and K. Fobelets. Efficient tool flow for 3D photovoltaic modelling. *Comput. Phys. Commun.*, 193:124–130, 2015.
15. P. Sharma, A. W. Walker, J. F. Wheeldon, K. Hinzer, and H. Schriemer. Enhanced efficiencies for high concentration, multijunction PV systems by optimizing grid spacing under nonuniform illumination. *Int. J. Photoenergy*, 2014:582083, 2014.
16. ASTM International. ASTM G173-03: Standard Tables for Reference Solar Spectral Irradiances: Direct Normal and Hemispherical on 37 degree Tilted Surface. Technical report, ASTM International, 2012.
17. D. E. Aspnes, S. M. Kelso, R. A. Logan, and R. Bhat. Optical properties of AlxGa1−xAs. *J. Appl. Phys.*, 60(2):754, 1986.
18. T. Kim, T. Ghong, Y. Kim, S. Kim, D. Aspnes, T. Mori, T. Yao, and B. Koo. Dielectric functions of InxGa1−xAs alloys. *Phys. Rev. B.*, 68(11):115323, 2003.
19. A. B. Djurisic, A. D. Rakic, P. C. K. Kwok, E. H. Li, M. L. Majewski, and J. M. Elazar. Modeling the optical constants of Al(x)Ga(1-x)As alloys. *J. Appl. Phys.*, 86(1):445–451, 1999.

20. M. Schubert, J. A. Woollam, G. Leibiger, B. Rheinlander, I. Pietzonka, T. Sab, and V. Gottschalch. Isotropic dielectric functions of highly disordered AlxGa1$-x$InP (0 <= x <= 1) lattice matched to GaAs. *J. Appl. Phys.*, 86(4):2025–2033, 1999.
21. Sentaurus Device User Guide, version K-2015. Synopsys Inc., Mountain View, CA. 1446p, 2015.
22. D. Schroeder. *Modelling of Interface Carrier Transport for Device Simulation*. Wien: Springer-Verlag, 1994.
23. K. Horio and H. Yanai. Numerical modeling of heterojunctions including the thermionic emission mechanism at the heterojunction interface. *IEEE Trans. Electron Dev.*, 37(4):1093–1098, 1990.
24. P. P. Altermatt. Models for numerical device simulations of crystalline silicon solar cells—a review. *J. Comput. Electron.*, 10(3):314–330, 2011. Multi-Junction Solar Cells 40–43
25. I. Vurgaftman, J. R. Meyer, and L. R. Ram-Mohan. Band parameters for III-V compound semiconductors and their alloys. *J. Appl. Phys.*, 89(11):5815, 2001.
26. D. H. Levi. Effects of ordering on the optical properties of $GaInP_2$. *Proc. SPIE.*, 5530:326–337, 2004.
27. G. Masetti, M. Severi, and S. Solmi. Modeling of carrier mobility against carrier concentration in arsenic-doped, phosphorus-doped, and boron-doped silicon. *IEEE Trans. Electron Dev.*, 30(7):764–769, 1983.
28. N. D. Arora, J. R. Hauser, and D. J. Roulston. Electron and hole mobilities in silicon as a function of concentration and temperature. *IEEE Trans. Electron Dev.*, 29(2):292–295, 1982.
29. J. Piprek. *Semiconductor Optoelectronic Devices: Introduction to Physics and Simulation*. San Diego, CA: Academic Press, 2003.
30. S. M. Sze and K. Ng Kwok. *Physics of Semiconductor Devices*, 3rd edition. Hoboken, NJ: Wiley-Interscience, 2007.
31. V. Palankovski and R. Quay. *Analysis and Simulation of Heterostructure Devices*. Vienna: Springer-Verlag, 2004.
32. S. Adachi. *Optical Constants of Crystalline and Amorphous Semiconductors*. Boston, MA: Springer, 1999.
33. E. D. Palik, editor. *Handbook of Optical Constants of Solids*. Cambridge, MA: Academic Press, 1998.
34. E. F. Schubert. *Physical Foundations of Solid-State Devices*. Rensselaer Polytechnic Institute, Troy, NY, 273 pages, 2006.
35. M. Levinshtein, S. Rumyantsev, and M. Shur, editors. *Handbook Series on Semiconductor Parameters*, vol. 2: Ternary And Quaternary III-V Compounds. Singapore: World Scientific, 1999.
36. M. Sotoodeh, A. H. Khalid, and A. A. Rezazadeh. Empirical low-field mobility model for III-V compounds applicable in device simulation codes. *J. Appl. Phys.*, 87(6):2890, 2000.
37. J. F. Wheeldon, C. E. Valdivia, A. W. Walker, G. Kolhatkar, A. Jaouad, A. Turala, B. Riel, D. Masson, N. Puetz, S. Fafard, R. Arès, V. Aimez, T. J. Hall, and K. Hinzer. Performance comparison of AlGaAs, GaAs and InGaP tunnel junctions for concentrated multijunction solar cells. *Prog. Photovolt. Res. Appl.*, 19(4):442–452, 2010. 40–44 Book title goes here.
38. A. W. Walker, O. Thériault, M. M. Wilkins, J. F. Wheeldon, and K. Hinzer. Tunnel-junction- limited multijunction solar cell performance over concentration. *IEEE J. Sel. Top. Quantum Electron.*, 19(5):4000508, 2013.
39. J. Fontanella, C. Andeen, and D. Schuele. Low-frequency dielectric constants of α-quartz, sapphire, MgF2, and MgO. *J. Appl. Phys.*, 45(7):2852–2854, 1974.
40. D. R. Lide, editor. *CRC Handbook of Chemistry and Physics,* 86th edition. Boca Raton, FL: CRC Press, 2005.
41. J. M. Siqueiros, R. Machorro, and L. E. Regalado. Determination of the optical constants of MgF(2) and ZnS from spectrophotometric measurements and the classical oscillator method. *Appl. Opt.*, 27(12):2549–2553, 1988.
42. J. R. Devore. Refractive indices of rutile and sphalerite. *J. Opt. Soc. Am.*, 41(6):416, 1951.

43. J. Kischkat, S. Peters, B. Gruska, M. Semtsiv, M. Chashnikova, M. Klinkmüller, O. Fedosenko, S. Machulik, A. Aleksandrova, G. Monastyrskyi, Y. Flores, and W. T. Masselink. Mid-infrared optical properties of thin films of aluminum oxide, titanium dioxide, silicon dioxide, aluminum nitride, and silicon nitride. *Appl. Opt.*, 51(28):6789–6798, 2012.
44. S. R. Kurtz, J. M. Olson, D. J. Friedman, J. F. Geisz, and A. E. Kibbler, Passivation of interfaces in high-efficiency photovoltaic devices, In *MRS Online Proceedings Library Archive*, vol. 573, pp. 95–106, San Francisco, CA, 1999.
45. W. Guter and A. W. Bett. IV-Characterization of Devices Consisting of Solar Cells and Tunnel Diodes. In *4th World Conference Photovoltaic Energy Conversion.*, pages 749–752. Waikoloa, HI, 2006.
46. E. O. Kane. Theory of tunneling. *J. Appl. Phys.*, 32(1):83, 1961.
47. G. A. M. Hurkx, D. B. M. Klaassen, and M. P. G. Knuvers. A new recombination model for device simulation including tunneling. *IEEE Trans. Electron Dev.*, 39(2):331–338, 1992.
48. A. Ajoy. Complex bandstructure of direct bandgap III-V semiconductors: application to tunneling. In M. Katiyar, B. Mazhari, and Y. N. Mohapatra, editors, *International Workshop on Physics of Semiconductor Devices*, Kanpur, India, 2012.
49. J. F. Wheeldon, C. E. Valdivia, A. Walker, G. Kolhatkar, T. J. Hall, K. Hinzer, D. Masson, S. Fafard, A. Jaouad, A. Turala, R. Ares, and V. Aimez. AlGaAs tunnel junction for high efficiency multi-junction solar cells: Simulation and measurement of temperature-dependent operation. In *34th IEEE Photovoltaic Specialists Conference*, Philadelphia, PA, pages 000106–000111, 2009.
50. M. Baudrit and C. Algora. Tunnel diode modeling, including nonlocal trap-assisted tunneling: A focus on III–V multijunction solar cell simulation. *IEEE Trans. Electron Dev.*, 57(10):2564–2571, 2010.
51. M. Hermle, G. Létay, S. P. Philipps, and A. W. Bett. Numerical simulation of tunnel diodes for multi-junction solar cells. *Prog. Photovolt. Res. Appl.*, 16:409–418, 2008.
52. Synopsys Inc. Simulation of GaAs Tunnel Diode for Multijunction Solar Cells, Synopsys, Inc., Mountain View, CA, 9 pages, 2012.
53. K. Emery, M. Meusel, R. Beckert, A. W Bett, and W. Warta. Procedures for evaluating multijunction concentrators. In *Conference Record 28th IEEE Photovoltaic Specialists Conference.*, pages 1126–1130. Anchorage, AK, 2000.
54. D. J. Friedman, J. F. Geisz, and M. A. Steiner. Effect of luminescent coupling on the optimal design of multijunction solar cells. *IEEE J. Photovolt.*, 4(3):986–990, 2014.
55. D. J. Friedman, J. F. Geisz, and M. A. Steiner. Analysis of multijunction solar cell current voltage characteristics in the presence of luminescent coupling. *IEEE J. Photovolt.*, 3(4):1429–1436, 2013.
56. S. H. Lim, J.-J. Li, E. H. Steenbergen, and Y.-H. Zhang. Luminescence coupling effects on multijunction solar cell external quantum efficiency measurement. *Prog. Photovolt. Res. Appl.*, 21(3):344–350, 2013.
57. M. Wilkins, A. M. Gabr, A. H. Trojnar, H. Schriemer, and K. Hinzer. Effects of luminescent coupling in single- and 4-junction dilute nitride solar cells. In *Proceeding 40th Photovoltaic Specialists Conference*, pages 6–9, Denver, CO, 2014.
58. M. Wilkins, C. E. Valdivia, A. M. Gabr, D. Masson, S. Fafard, and K. Hinzer. Luminescent coupling in planar opto-electronic devices. *J. Appl. Phys.*, 118(14):143102, 2015.
59. A. W. Walker, O. Hohn, D. N. Micha, L. Wagner, H. Helmers, A. W. Bett, and F. Dimroth. Impact of photon recycling and luminescence coupling on III - V photovoltaic devices. In *Proc. SPIE*, vol. 9358, San Francisco, CA, pages 11–13, 2015.
60. M. Wilkins, C. E. Valdivia, S. Chahal, M. Ishigaki, D. P. Masson, S. Fafard, and K. Hinzer. Performance impact of luminescent coupling on monolithic 12-junction phototransducers for 12 V photonic power systems. In *Proc. of SPIE*, vol. 9743, San Francisco, CA, 2016.

41

Nanostructure Solar Cells

41.1 Introduction: Nanostructures in Photovoltaics 441
41.2 Physical Mechanisms of NSSC Device Operation 444
Photogeneration of Electron–Hole Pairs • Charge Separation and Extraction • Recombination
41.3 Modeling of NSSC Device Characteristics 454
Global Detailed Balance Theories for Ideal Systems • Hybrid Models: Semiclassical Transport with Detailed Balance Rates • Fully Microscopic Quantum-Kinetic Models
41.4 Case Study: Single Quantum Well p-i-n SC 461
States • Dynamics • Characteristics
41.5 Summary 465

Urs Aeberhard

41.1 Introduction: Nanostructures in Photovoltaics

The operation of solar cell (SC) devices is governed by a large variety of solid-state properties associated with the interplay of optical, electronic, and vibrational degrees of freedom of the component materials. Light-matter interaction provides the elementary optoelectronic processes of photogeneration and radiative recombination, while coupling of charge carriers to lattice vibrations leads to energy conversion losses due to thermalization, dissipative transport, and nonradiative recombination.

In the presence of spatial inhomogeneities, the optoelectronic material properties change if the characteristic length scale of the spatial variation is reduced below a certain threshold. The critical length scale depends on the physical degree of freedom (optical, electronic, vibrational) and corresponds to the wavelength of the associated wave (electromagnetic [EM] wave, de Broglie wave, sound wave). In this *subwavelength* regime, the physics is governed by the corresponding wave equations (Schrödinger, Maxwell, harmonic lattice dynamics), with the peculiar features such as interference or nonlocality. Along with the states, the interaction of different degrees of freedom is strongly modified in the presence of confinement, mainly due to the effect of localization and symmetry breaking. This tunability of physical properties via variation of material composition, size and shape of structures below a characteristic length scale—termed *nanostructures* (NSs)[†]—provides the basis for the vast field of nanotechnological applications, and the relation between configurational parameters and target NS functionality is a major topic of past and current research. Almost since the advent of semiconductor NS fabrication technologies, such as molecular beam epitaxy (MBE) or metal-organic chemical vapor deposition (MOCVD), but also lithographic processes, the peculiar physical properties of the NS produced in this way have also been utilized for the design of novel SC devices with the potential of increased photovoltaic (PV) conversion efficiency. In later years, new NS fabrication techniques such as nanoimprinting, catalytic vapor deposition, or solution-based processes have enlarged the zoo of NSs and NS architectures proposed for SC with either enhanced efficiency or the potential for substantial cost reduction.

[†] While the term *nano* is used for all of the above degrees of freedom (nanophotonics, -electronics, -phononics) to distinguish the confinement regime from bulk behavior, the characteristic confinement length varies strongly (0.1 μm to 1 cm for photons, 0.5 to 10 nm for electrons, 5 nm to 1 cm for phonons).

A common way to characterize NSs is according to the dimensionality: The material without confinement, i.e., featuring extended states in all three spatial dimensions, is referred to as *bulk*. For electronic confinement in one, two, and three spatial dimensions, the resulting 2D, 1D, and 0D NSs are called *quantum well* (QW), *quantum wire* (QWR), and *quantum dot* (QD), respectively. At length scales where only optical confinement is present, the corresponding objects are usually called thin films, nanowires, and nanoparticles, respectively. On the other hand, the functionalities of nanostructure solar cell (NSSC) components can be categorized according to the physical degree of freedom that is to be affected: *Optical* functionalities are used for the purpose of light trapping for absorption enhancement via engineering of the EM field strength inside the absorber; the (opto-)*electronic* functionalities, which are related to the tuning of (opto-)electronic properties such as bandgaps, absorption coefficients, mobilities, and carrier relaxation rates via electronic structure engineering; and *vibrational* functionalities, which affect the dissipation of carrier energy due to coupling to lattice vibrations, for instance, via inhibition of suitable phonon modes. As explained earlier, all of the functionalities emerge as the consequence of the nanoscale dimension of the structures that leads to quantization and confinement of the photonic, electronic, and vibrational states and results in physical properties that can deviate considerably from the bulk behavior.

The main applications of NS in photovoltaics are in the so-called second and third generations of PV devices [1], where NS are used in the implementation of SC concepts aiming at either lower cost through strongly reduced material usage and cheap production technology or at increased energy conversion efficiencies through enhanced spectrum utilization and/or reduced thermalization losses. NS with optical functionality have been instrumental in the realization of thin-film SC (second generation) based on low absorption materials, such as thin-film silicon devices [2] or ultrathin absorbers [3]. Relevant representatives of such NS applications for light-trapping include plasmonic metal nanoparticles [4] and nanostructured metal films [5], dielectric scatterers [6] and diffraction gratings [7], engineering of the optical density of states (DOS) via photonic crystals [8], light focusing antenna effects in nanowires and nanorods [9], as well as spectral conversion via QDs [10]. This wide field of optical NS for PV applications is covered by a number of topical reviews [11–14].

The electronic functionality of NS concerns first of all the tuning of optical transitions and nonradiative charge carrier dynamics in high-efficiency SC architectures for concentrator and space applications (third generation). Since the early nineties, multi-QW (MQW—multiple decoupled QW) and QW superlattice (QWSL—periodic and coupled QW) structures have been implemented in different III-V semiconductor materials and have been investigated for application in single-junction SC [15] and multijunction (MJ) PV devices [16], and, more recently, also for hot-carrier SC (HCSC) [17]. The practical relevance of these structures lies mainly in the adjustability of the absorption edge to the wavelength required for efficiency optimization, in the case where suitable bulk materials are not available. Similar bandgap engineering features are provided by multi-QD (MQD) or QD superlattice (QDSL) structures implemented in III-V semiconductors [18–21] or silicon alloys [22], with the advantage of improved strain relaxation mechanisms. However, while QWSCs based on strain-balanced components have enabled efficiencies comparable to the bulk limit [23], severe recombination losses did not allow thus far a similar achievement in QDSC [24]. On the other hand, QD structures were considered to be particularly useful for the implementation of a number of specific third-generation PV concepts: for intermediate band SC (IBSC) due to the possibility to obtain an energetically separated intermediate state or band from the ground state of the QD [25]; for multiple exciton generation (MEG) via impact ionization (providing increased current) due to enhanced electron–electron interaction as a consequence of wave function localization [26]; and for the extraction of hot carriers in HCSC (providing increased voltage) due to inhibited electron–phonon scattering resulting from the low density and spectral sparsity of electronic and vibrational modes [27]. These effects have been demonstrated in dedicated experimental setups [28–33], but they remain to be shown in working SC devices.

Besides these highly ordered epitaxial QD structures, there is a whole class of QD materials made as colloids from solution, such as, e.g., PbSe and PbS [34], which are arranged into different SC architectures, ranging from QD-sensitized TiO_2 electrodes to hybrid organic–inorganic devices and closely packed QD solids with carrier-selective contacting schemes [35–38]. While present efficiencies are still low (11.3%), these approaches promise low-cost manufacturing, and there is still the prospect of MEG exploitation as observed in the constituent particles [39].

For the sake of completeness, a word is due here on the case of QWRs. Even though they offer a perfect combination of size tunability in transverse dimensions with extended states in transport direction, there are almost no experimental structures including 1D NS with electronic functionality. The reasons are twofold. On the one hand, it is very challenging to produce wires with diameters small enough such as to exhibit electronic confinement. On the other hand, it is not possible to achieve at the same time optical and electronic confinement: At the dimensions suitable for optical confinement, the absorber is electronically bulk-like, and no guided optical modes can exist in the tiny dimensions required for electronic confinement.

As compared to the optical and electronic features emerging at the nanoscale, the vibrational functionalities of NSs have received the least attention so far. Most potential applications consider exploitation of reduced carrier cooling by engineering the phonon modes in NSs such as QWSL [40] and QDSL [22,27]. The issue of confined phonons is discussed also in the context of fast relaxation of hot photogenerated excitons in isolated QD, one of the detrimental factors suppressing MEG, which was thought to be reduced in QD. Recent experiments and simulations emphasize the influence of softened surface modes enabling efficient multiphonon relaxation [41].

All of the above functionalities might be relevant in a single SC device, for instance a HCSC based on a III-V QWSL structure (similar to [40]) with absorption enhancement induced by plasmonic nanoparticles or a cavity effect. In such a case, deviations from bulk physics need to be considered for photons, electrons, and phonons at the same time. Furthermore, while the single functionality of an NS component (e.g., enhancement of interband absorption due to wave function localization) might be beneficial for a certain aspect of device operation (e.g., photogeneration), it may at the same time have a detrimental impact on another aspect (e.g., recombination and transport). Indeed, in experimental implementations of advanced PV concepts, NSs are often found to introduce electronic and optical losses, e.g., due to high defect density at NS interfaces and parasitic absorption into localized states with insufficient carrier extraction efficiency. In the idealized models considered for the determination of efficiency bounds such as the famous Shockley–Queisser (SQ) limit [42], such losses are usually neglected. This explains to some extent the large discrepancy between theoretical ideal case predictions and experimentally achieved results. However, for a realistic assessment of the technological potential of a given NSSC concept, the detrimental aspects associated with the NS need to be considered and quantified. This underlines the importance of advanced modeling approaches reaching beyond bulk semiconductor physics for this category of PV devices, which will be in the focus of the remainder of this chapter.

Space does not permit a detailed discussion of the modeling approaches for all of the NS used in photovoltaics as outlined above. Since nanophotonic aspects of optoelectronic device modeling are covered elsewhere in this book, NS with purely optical functionality—including the important case of the nanowire SC, where the absorber is at the same time its own light-trapping structure—will not be addressed here. The other important class of NSSC that will not be covered comprises the colloidal QD SC architectures. In these devices, the PV processes are strongly influenced by the surface chemistry, such as, e.g., the properties associated with different ligand molecules, but also by strong carrier localization, large exciton binding energies, and the considerable fluctuations in size and position. As a consequence, mobility is usually limited by thermally activated hopping processes [43], resulting in poor transport properties. The disorder-induced smearing of the band edge has also unfavorable impact on the achievable open-circuit voltages, even at the radiative limit and under the assumption of perfect transport [44]. Overall, the situation is

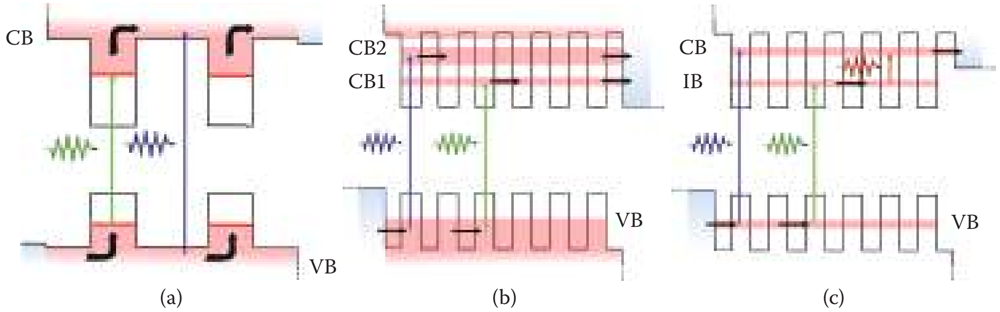


FIGURE 41.1 Common architectures of NSSC: (a) multi-QW or QD structures, where the confined states of distinct NS are not coupled, and photogenerated carriers therefore need first to escape to extended (quasi-)continuum states in order to contribute to current; (b) superlattices of QW or QD, where miniband formation or sequential tunneling enables photocarrier transport and extraction via partially localized NS states; (c) in contrast to (b), the IBSC configuration features noncontacted NS states, which, however, can still be delocalized, but, contrary to (a), should be only radiatively coupled to the contacted continuum states by which carriers are extracted. CB, conduction band; VB, valence band; IB, intermediate band.

closer to the case of organic and hybrid SC, for which a whole range of dedicated simulation approaches exist [45].

The main focus of this chapter, however, will be on the modeling of *regimented* QW and QD architectures as utilized for the implementation of high-efficiency SC concepts using *inorganic* semiconductor materials. Among these structures, three main categories can be distinguished according to the participation of the NS states in the transport of charge carriers. The first category includes devices where the NS states are localized in transport direction and need to be coupled to another set of extended (i.e., bulk-like or unconfined) states to enable charge carrier flow. Typical representatives are the MQW and MQD SC (Figure 41.1a). The second category consists of architectures where both generation-recombination and transport processes are mediated by NS states, which requires a finite degree of wave function delocalization in transport direction related to the coupling of NS, as in QDSL devices with large confinement barriers (Figure 41.1b). Finally, there are structures where coupling-induced delocalization exists between both confined and unconfined states, but only the continuum states are connected to contacts, as in some QD-IBSC implementations (Figure 41.1c).

In Section 41.2, the main physical mechanisms governing the PV device operation are discussed for the different types of NS architectures with regard to their theoretical description and modelization. In Section 41.3, existing methods are reviewed to obtain the device characteristics of NS-based SCs under consideration of the modeling requirements resulting from Section 41.2 and the associated implementation challenges, establishing a simulation hierarchy ranging from global detailed balance approaches for limiting efficiency calculations to microscopic quantum-kinetic theories for the local charge carrier dynamics in nonclassical device regions. In Section 41.4, the simulation of a prototypical NSSC in the form of a single QW photodiode is considered in the light of the approaches discussed in Section 41.3.

41.2 Physical Mechanisms of NSSC Device Operation

In the following, the different stages of the PV energy conversion process in NSSC—from the incident photon flux to the extracted charge current—are considered with focus on the modeling requirements imposed by the presence of NS components.

41.2.1 Photogeneration of Electron–Hole Pairs

The structures that are of interest in NS PV architectures are small compared to the coherence length of sunlight. The spectral *photon flux* $\Phi_\gamma(\mathbf{r},\hbar\omega)$ at energy $\hbar\omega$, with $\hbar$ the reduced Planck constant and ω the frequency of the light, and position $\mathbf{r}$ inside the absorber of the SC under consideration is therefore well described by the propagation of coherent EM radiation in a dispersive medium, as provided by the solution of (macroscopic) Maxwell's equations. In the most general case, the linear response[†] of the medium to the transverse electrical field $\boldsymbol{\mathcal{E}}^{\mathrm{t}}$ is characterized by the complex microscopic dielectric tensor $\overleftrightarrow{\varepsilon}(\mathbf{r},\mathbf{r}',\hbar\omega)$, which considers the electronic excitations in dependence of the energy and polarization of the incident radiation field and of the available electronic states. The corresponding local volume rate $\mathcal{G}$ for the generation of electron–hole pairs by photon absorption is obtained from the dissipated EM energy as follows [46]:

$$\mathcal{G}(\mathbf{r},\hbar\omega) = \eta_{\mathrm{gen}}(\mathbf{r},\hbar\omega)(2\hbar\mathcal{V})^{-1}\sum_{\mu,\nu}\int d^3r'\,\Im\left\{\varepsilon_0\varepsilon_{\mu\nu}(\mathbf{r},\mathbf{r}',\hbar\omega)\mathcal{E}^{t*}_{\mu}(\mathbf{r},\hbar\omega)\mathcal{E}^{t}_{\nu}(\mathbf{r}',\hbar\omega)\right\}, \tag{41.1}$$

where $\Im$ denotes the imaginary part, ε_0 is the vacuum permittivity, η_{gen} is the generation efficiency ($\eta_{\mathrm{gen}} = 1$ means that each photon that is absorbed generates exactly one electron-hole pair), and $\mathcal{V}$ is the absorbing volume. In the above expression, the time-harmonic monochromatic Ansatz $\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r},t) = \Re\{\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r},\hbar\omega)\exp(-i\omega t)\}$ was used. While the interaction with the light is inherently nonlocal from the electronic point of view, an averaged local version of the dielectric tensor can be used if the EM field varies only slowly over the length scale of electronic inhomogeneities. For an isotropic medium and unit generation efficiency, Equation 41.1 can then be rewritten as

$$\mathcal{G}(\mathbf{r},\hbar\omega) = (2\hbar)^{-1}\varepsilon_0\Im\varepsilon(\mathbf{r},\hbar\omega)|\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r},\hbar\omega)|^2 \approx \alpha(\mathbf{r},\hbar\omega)\Phi(\mathbf{r},\hbar\omega), \tag{41.2}$$

where

$$\alpha(\mathbf{r},\hbar\omega) \equiv \frac{\omega}{n_{\mathrm{r}}c_0}\Im\varepsilon(\mathbf{r},\hbar\omega) \tag{41.3}$$

denotes the (locally defined) *absorption coefficient* that reflects the decay of the optical intensity due to all optical excitations possible at a given photon energy, weighted by their respective strength (c_0: speed of light in vacuum, n_{r}: refractive index of the background).

41.2.1.1 Photon Flux

In the general time-harmonic case, the photon flux is given by the Poynting vector $\mathbf{S}$ of the transverse EM fields $(\boldsymbol{\mathcal{E}}^{\mathrm{t}},\boldsymbol{\mathcal{B}}^{\mathrm{t}})$:

$$\mathbf{S}(\mathbf{r},\hbar\omega) = (2\mu_0)^{-1}\left\{\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r},\hbar\omega)\times\boldsymbol{\mathcal{B}}^{t}(\mathbf{r},\hbar\omega)\right\} \equiv \Phi(\mathbf{r},\hbar\omega)\hbar\omega\,\hat{\mathbf{n}}(\mathbf{r},\hbar\omega), \tag{41.4}$$

where μ_0 is the vacuum permeability and $\hat{\mathbf{n}}$ is the unit vector in the direction of propagation. The evolution of the flux inside the absorber is related to the EM dissipation due to light-matter coupling via the optical conservation law:

$$\nabla\cdot\mathbf{S}(\mathbf{r},\hbar\omega) = -\frac{1}{2}\omega\varepsilon_0\,\Im\left\{\varepsilon(\mathbf{r},\hbar\omega)[\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r},\hbar\omega)]^*\cdot\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r},\hbar\omega)\right\}, \tag{41.5}$$

[†] This is appropriate for standard solar illumination conditions, i.e., the AM1.5g spectrum.

where the transverse electrical field is given by Maxwell's equations acquiring the form

$$\mu_0^{-1}\nabla\times\nabla\times\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r},\hbar\omega)=\omega^2\varepsilon_0\boldsymbol{\varepsilon}(\mathbf{r},\hbar\omega)\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r},\hbar\omega),\tag{41.6}$$

$$\nabla\cdot\varepsilon_0\boldsymbol{\varepsilon}(\mathbf{r},\hbar\omega)\boldsymbol{\mathcal{E}}^{\mathrm{t}}(\mathbf{r},\hbar\omega)=0.\tag{41.7}$$

The *absorptance* of the absorbing volume is then given by the ratio $a\equiv P_{\mathrm{abs}}/P_{\mathrm{in}}$ of absorbed and incident EM power, with $P_{\mathrm{abs}}=\int_{\mathcal{V}}d\mathcal{V}(-\nabla\cdot\mathbf{S})$ and $P_{\mathrm{in}}=\int_{\mathcal{S}}d\boldsymbol{S}\cdot\mathbf{S}$, where $\mathcal{S}$ is the surface of the absorbing volume. In the absence of optical confinement—for instance, in optically thick absorbers—the decay of the flux can be written in terms of the absorption coefficient via the *Lambert–Beer* (LB) law, e.g., for propagation along the z-direction:

$$\Phi_{\mathrm{LB}}(z,\hbar\omega)=\Phi_0^{\swarrow}(\hbar\omega)\exp\left(-\int_0^z dz'\,\alpha(z',\hbar\omega)\right),\tag{41.8}$$

with $\Phi_0^{\swarrow}$ denoting the incident photon flux at $z=0$. This defines the absorptance $a_{\mathrm{LB}}(z,\hbar\omega)\equiv 1-\Phi_{\mathrm{LB}}(z,\hbar\omega)/\Phi_0^{\swarrow}(\hbar\omega)$ after an optical path of length z. For planar systems featuring optical elements such as antireflection coatings (ARC) or a back reflector, as frequently used in thin-film PV devices due to the requirement of absorption enhancement via light trapping, the required solution of Maxwell's equations can be performed using a transfer-matrix method (TMM) [47], which relates the coefficients of incident, reflected, and transmitted plane wave components for piecewise constant dielectric function. In the case of 2-D/3-D nanophotonic or plasmonic light trapping structures, Maxwell's equations (Equation 41.7) have to be solved by advanced numerical approaches such as the finite-difference time-domain (FDTD) method or the finite element method (FEM).

41.2.1.2 Absorption Coefficient

Equations 41.2 and 41.3 can be used to determine the absorption coefficient from the complex dielectric function or the linear *susceptibility* $\chi(\omega)=\varepsilon(\omega)-\varepsilon_b$ ($\varepsilon_b\equiv n_r^2$: background dielectric constant) or, equivalently, from the net absorption rate $r_{\mathrm{abs}}=\mathcal{G}/\eta_{\mathrm{gen}}$ (net means after subtraction of the rate for stimulated emission). For weak perturbation, the net volume rate of uncorrelated direct electronic transitions between states v and c by absorption of photons of energy $E_\gamma=\hbar\omega$ is given by *Fermi's Golden Rule* (FGR):

$$r_{v\to c}^{\mathrm{abs}}(E_\gamma)=\frac{2\pi}{\hbar\mathcal{V}}|\langle c|\hat{H}_{e\gamma}|v\rangle|^2\delta(\varepsilon_c-\varepsilon_v-E_\gamma)(f_v-f_c),\tag{41.9}$$

with $\hat{H}_{e\gamma}$ the electron–photon interaction Hamiltonian, $\varepsilon_{c,v}$ the energies of single electron and hole states, and $f_{c,v}$ the occupation of these states. The electron–photon Hamiltonian has the general form

$$\hat{H}_{e\gamma}(\mathbf{r},t)=\frac{e}{2m_0}\left\{\hat{\mathbf{A}}(\mathbf{r},t)\cdot\hat{\mathbf{p}}+\hat{\mathbf{p}}\cdot\hat{\mathbf{A}}(\mathbf{r},t)\right\}+\frac{e^2[A(\mathbf{r},t)]^2}{2m_0}\approx\frac{e}{m_0}\hat{\mathbf{A}}(\mathbf{r},t)\cdot\hat{\mathbf{p}},\tag{41.10}$$

where $\hat{\mathbf{A}}$ is the EM vector potential, $\hat{\mathbf{p}}=-i\hbar\nabla$ is the momentum operator, and for the last expression (minimal coupling), the assumption of weak vector potential $\hat{\mathbf{A}}$ was made and Coulomb gauge ($\nabla\cdot\mathbf{A}=0$) was used. In the standard *semiclassical* picture of light-matter interaction in semiconductors, the light is treated as a classical wave. Using for the incident field, a monochromatic time-harmonic plane wave expansion:

$$\mathbf{A}(\mathbf{r},t)=\frac{A_0}{2}\boldsymbol{\epsilon}\left\{e^{i(\mathbf{q}\cdot\mathbf{r}-\omega t)}+c.c.\right\}\tag{41.11}$$

of the vector potential, with amplitude $A_0/2$, wave vector $\mathbf{q} = \frac{\omega}{c}\hat{\mathbf{n}}$, and polarization $\boldsymbol{\epsilon}$, together with the minimal coupling form of $H_{e\gamma}$ in the FGR rate (Equation 41.9) yields

$$r^{\text{abs}}_{v\to c}(E_\gamma) = \frac{2\pi}{\hbar V}\frac{e^2 A_0^2}{4m_0^2}\mathcal{M}^2_{\text{cv}}\delta(\varepsilon_{cv} - E_\gamma)(f_v - f_c), \tag{41.12}$$

$$\mathcal{M}_{\text{cv}} \equiv |\langle c|\mathrm{e}^{i\mathbf{q}\cdot\mathbf{r}}\boldsymbol{\epsilon}\cdot\hat{\mathbf{p}}|v\rangle|, \tag{41.13}$$

where $\varepsilon_{cv} \equiv \varepsilon_c - \varepsilon_v$. If $\mathbf{q}\cdot\mathbf{r} \ll 1$ (i.e., the wavelength of the light is much larger than the characteristic length scale of the absorbing structure), the phase factor $\mathrm{e}^{i\mathbf{q}\cdot\mathbf{r}}$ in the matrix element can be neglected, which corresponds to the *dipole approximation*, meaning that contributions of higher order electric and magnetic multipoles are not considered. In this approximation, the coupling via the momentum operator $\hat{\mathbf{p}}$ can be replaced by a coupling via the dipole operator $\hat{\mathbf{d}} = -e\hat{\mathbf{r}}$ ($\hat{\mathbf{r}}$: position operator) due to the relation $(e/m_0)\mathbf{A}\cdot\hat{\mathbf{p}} \approx -\hat{\mathbf{d}}\cdot\boldsymbol{\mathcal{E}}^{\text{t}}$ [48]. The photon flux associated with Equation 41.11 amounts to

$$\Phi_0^{\swarrow}(\hbar\omega) = \frac{n_r c_0 \varepsilon_0 \omega A_0^2}{2\hbar}, \tag{41.14}$$

which, together with Equation 41.12, provides the general FGR absorption coefficient:

$$\alpha^{\text{FGR}}(E_\gamma) = \frac{\pi\hbar}{n_r c_0 \varepsilon_0 E_\gamma \mathcal{V}}\left(\frac{e}{m_0}\right)^2 \sum_{v,c}\mathcal{M}^2_{cv}\delta(\varepsilon_{cv} - E_\gamma)(f_v - f_c), \tag{41.15}$$

where the sum is over all initial and final states.

In Equation 41.15, the occupation of the initial and final states is assumed to be known, and in the case of interband transitions, $f_v = 1$ and $f_c = 0$ are often assumed. However, this is no longer appropriate in the case of transitions between partially filled subbands, as for instance in the case ot the IBSC concept. Thus, in general, the occupation needs to be determined by a microscopic treatment of the charge carrier dynamics under illumination in the framework of the semiconductor Bloch equations [48–50] or nonequilibrium Green's functions [51].

The main quantities determining the generation of electron-hole pairs are, thus, the local value of the transverse EM field on the one hand, and the local value of the dipole or momentum matrix elements (MMEs) and of the density of occupied initial and empty final states of the optical transitions, respectively, on the other hand. While the spatial variation of the transverse EM fields is small over the extent of semiconductor NS in the range of few nanometers, the electronic quantities reflect the impact of symmetry breaking, wave function localization, and energy quantization associated with reduced dimensionality. A prominent example is silicon QDs, where the contribution of direct transitions—absent in bulk—and an increase in oscillator strength due to larger overlap of electron and hole wave functions (in this case equivalent to larger MME) result in a strongly increased absorption [52,53].

Considering Equation 41.12, evaluation of the absorption in semiconductor NSs starts with the computation of the wave functions $\psi_{c,v}$ and energies $\varepsilon_{c,v}$ of the single-particle states $\{|v\rangle, |c\rangle\}$ participating in the optical transitions. There is a vast amount of literature on that topic, which is also covered by a number of chapters in this volume, and will thus not be treated explicitly here. With the electronic structure information, idealized absorption spectra can be obtained from Equation 41.12. However, for a realistic assessment of the PV response, the experimental spectra should be reproduced as closely as possible, which requires consideration of the actual effects that have an impact on the line shape, such as finite fields, excitonic contributions, lifetime broadening due to coupling to phonons and hybridization with continuum states, compositional disorder, and finite-size distributions. Various approaches with increasing degree of sophistication can be found in the literature.

For **QW**, excitons and field effects (i.e., electroabsorption) were included in [54] using a variational ansatz for the exciton wave functions. The consistent consideration of the bound-to-bound state and the bound-to-quasibound state transitions including excitonic and field effects was achieved using an exciton Green's function approach [55]. To describe lineshape broadening beyond phenomenological approaches, microscopic quantum-kinetic theories such as the density matrix formalism, including the different scattering mechanisms (electron–phonon, electron-electron, etc.) can be used [48–50]. Recently, excitonic absorption enhancement for QWs under consideration of the electron–phonon interaction was formulated in a general nonequilibrium Green's function framework [56]. In practice, however, some degree of fitting is used to reproduce the experimental spectra. For QWSC, a semiempirical approach was introduced by Paxman et al. [57], based on earlier work on QW optical modulators [58,59], in which the field-dependent absorption coefficient is expressed as

$$\alpha(E_\gamma, \mathcal{E}^z) = \sum_{c,v} \alpha_{cv} M_{cv}(\mathcal{E}^z) \Big[\sum_{l} r^l_{cv} \mathcal{L}[E_\gamma, \varepsilon_{cv}(\mathcal{E}^z) - B^l_{cv}] + \int_{\varepsilon_{cv}(\mathcal{E}^z)}^{\infty} dE' \mathcal{L}(E_\gamma, E') \Big], \tag{41.16}$$

where $\mathcal{E}^z$ is the longitudinal electric field in growth direction, $M_{cv} = \int_L dz \psi_c^*(z, \mathcal{E}^z)\psi_v(z, \mathcal{E}^z)$ (L: normalization length) is the overlap matrix element of the field-dependent single-particle electron and hole wave functions, r^l_{cv} (in units of energy) and B^l_{cv} are the relative oscillator strength and binding energy of the lth exciton, respectively, and $\mathcal{L}(E_\gamma, E') = \Gamma/(2\pi[(E_\gamma - E')^2 + \Gamma^2/4])$ is a Lorenzian lineshape function to include homogeneous broadening characterized by Γ. The parameter α_{cv} corresponds to the absorption coefficient at the onset of the continuum absorption, and, in practice, is used together with B and r to fit the experimental spectra. For the absorption at photon energies above the bulk bandgap of the barrier material, the bulk absorption coefficient of the well material is used, neglecting any field effects. A similar semiempirical approach was used in [60], based on [61].

For **QD**, the single-particle absorption spectra can be obtained using similar density matrix formalisms for the linear susceptibility [62] or the equivalent FGR, with straightforward extension from single QD to QDSL [63,64]; in the case of significant excitonic effects, the latter can be treated by means of the configuration interaction (CI) method [65]. While lineshape broadening due to dephasing via electron–phonon interactions can in principle be included in quantum-kinetic approaches, the finite-size distribution of QD ensembles requires additional convolution with some distribution function representing inhomogeneous broadening, commonly of the Gaussian type [66]. In self-assembled QD formed on a wetting layer (WL), the WL states may also contribute to the subgap absorption [67]. For practical device simulation, a common approach is the use of an effective medium absorption coefficient [68]:

$$\bar{\alpha}_{\text{eff}} = w_{\text{QD}}\bar{\alpha}_{\text{QD}} + w_{\text{WL}}\bar{\alpha}_{\text{WL}} + w_{\text{bulk(QD)}}\bar{\alpha}_{\text{bulk(QD)}} + w_{\text{barrier}}\bar{\alpha}_{\text{barrier}}, \tag{41.17}$$

which contains the contribution of the confined states in the QD, the WL, the bulk absorption of the QD material for continuum states, and the barrier absorption, weighted according to the relative volume fractions and a given distribution function reflecting the variations in size and shape. The general expression for the absorption of an ensemble of QD configurations C with distribution f using these simple approaches reads:

$$\bar{\alpha}_{\text{QD}}(E_\gamma) = \int dCf(C)\alpha_C(E_\gamma), \tag{41.18}$$

$$\alpha_C(E_\gamma) = \frac{\pi\hbar}{n_r c_0 \varepsilon_0 E_\gamma \mathcal{V}_{QD}} \left(\frac{e}{m_0}\right)^2 \sum_{c,v} \bar{p}^2_{cv}(C) \frac{\Gamma}{\left(\varepsilon_{cv}(C) - E_\gamma\right)^2 + \Gamma^2/4} (f_v - f_c), \tag{41.19}$$

where $\bar{p}_{cv}$ is the polarization-averaged MME and Γ is the homogeneous broadening. The WL absorption is given by the expression for QW (usually neglecting excitonic absorption), where consideration of thickness fluctuations results in an additional absorption tail. Since the absorbing volume of a QD is not always well defined, especially in the case of quasibound states, it is common to write the QD contribution absorption in terms of an *absorption cross-section* σ_{QD} per QD and of the QD concentration N_{QD} via $\alpha_{QD} = N_{QD}\sigma_{QD}$.

41.2.2 Charge Separation and Extraction

In general, the extraction of photogenerated charge carriers represents one of the most critical aspects in NS-based PV devices. Indeed, while in some devices based on QW or QD, the tailoring of the absorption via band structure engineering could be successfully demonstrated [69,70], the efficiency of carrier collection has until now remained below the values found for the bulk counterparts, in some cases on a dramatically poor level, especially in devices where transport proceeds via states with increased degree of localization. There are several reasons for the observed performance issue, which are associated with different stages in the extraction process, such as exciton dissociation, carrier escape and capture between localized and extended states, mobility issues associated with scattering, and the virtual absence of true miniband formation in realistic situations, which shall be discussed below.

41.2.2.1 Exciton Dissociation

Since electrons and holes have to be extracted via separate contacts, the dissociation of the photogenerated excitons and the subsequent charge separation need to occur at some point in the device. In bipolar bulk SC, exciton binding energies are usually in the range of a few meV, resulting in thermal dissociation immediately after generation, and the generation can thus safely be described via (Equation 41.2) in terms of noninteracting electron-hole pairs. In isolated NS on the other hand, spatial confinement enhances the electron-hole interaction, and exciton binding energies can amount to multiples of the thermal energy $k_B T$ (k_B: Boltzmann constant, T: temperature), in which case rapid thermal dissociation is no longer possible, and the dissociation process needs to be considered explicitly [53]. Inefficient exciton dissociation can result in a transport behavior dominated by exciton diffusion, similar to the situation in organic SC devices, where a specially designed bulk–heterojunction interface is required for exciton dissociation. As a SC based on exciton diffusion is not likely to reach very high efficiencies, the description of excitons in the simulation of efficient NSSC devices is primarily focused on the excitonic enhancement of optical transitions close to the effective band edge, as described above in the treatment of the absorption coefficient.

41.2.2.2 Charge Carrier Escape and Capture

Charge carriers that are generated in localized states of electronically decoupled NS need to be transferred to extended states in order to contribute to current, which normally amounts to a sequence of scattering events involving confined (2D—QW/0D—QD) subband states followed by scattering from a bound or quasibound state to a (3D) continuum state. While this *escape* process constitutes a loss mechanism in light-emitting devices, which is sought to be suppressed in optimized designs, the opposite applies in NSSC architectures. Hence, understanding the escape process via identification of dominant and limiting mechanisms and design of corresponding models has been an important focus in NSSC research from the beginning. On the other hand, any escape channel enables its inverse capture process, which removes photocarriers from contacted states and increases dark current flow.

In general, while the microscopic processes underlying the carrier dynamics are always related to the fundamental coupling between the electronic, vibrational, and optical degrees of freedom, there are several mechanisms by which photogenerated charge carriers can escape from confined NS states, which are commonly characterized by respective time constants τ^i_{esc}, with total escape rate proportional to $\tau^{-1}_{\text{esc}} = \sum_i (\tau^i_{\text{esc}})^{-1}$. For **QW** at room temperature—the relevant temperature regime for terrestrial SC operation—and with shallow confinement, the dominant escape mechanism was identified as *thermionic emission* [71], which describes the current due to charge carriers that occupy the part of the energy distribution located above the barrier edge. In the presence of a rectifying built-in field, the escape process is affected by the latter through the modified height (with respect to the center of the QW) and triangular shape of the barrier, giving rise to escape enhancement due to *tunneling*. For the common assumption of a constant field, the effective barrier for escape in direction of and against the field is given by $V_{B\mp}(\mathcal{E}^z) = \Delta E_b \mp q\mathcal{E}^z L_w/2$, where ΔE_b is the offset of bulk band edges and L_w is the QW width. A general *phenomenological* expression for the corresponding escape time is then given by [71]

$$\tau_{\text{esc}}^{-1} = \frac{1}{L_w}\int_0^\infty \tilde{n}(E)\mathcal{T}(E)v(E)dE, \quad \tilde{n}(E) = \frac{n(E)}{\int_0^\infty n(E)dE}, \tag{41.20}$$

where $\mathcal{T}(E)$ is the probability function for transmission through the barriers, $v(E)$ is the carrier velocity, and $n(E) = g(E)f(E)$ is the spectral carrier density at an energy E measured from the center of the bottom and assuming a given distribution function $f(E)$ for charge carriers in the QW. The actual form of Equation 41.20 depends on the specific description of the spectral quantities $(\mathcal{T}, v, g, f)$. The simplest assumption for $\mathcal{T}$ is a step function with unit transmission above the effective barrier edge and zero below, which neglects tunneling contributions. Tunneling can be included on the level of the Wentzel–Kramers–Brillouin (WKB) theory via [72]

$$\mathcal{T}_{\text{WKB}}(E) = \exp\left(-\frac{2}{\hbar}\int_0^{L_B} dz\sqrt{2m_B^*[V_B(z) - E]}\right), \tag{41.21}$$

where m_B^* is the effective mass of carriers in the barrier material, L_B is the width of the barrier, and $V_B(z)$ is the spatial profile of the barrier potential. More accurate approaches for determination of the transmission function are based on analytical or numerical solution of the Schrödinger equation using the exact barrier potential, e.g., by means of an electronic transfer matrix formalism [71]. For free carriers, the velocity is related to the (kinetic) carrier energy via the momentum, e.g., $v_z = \hbar k_z/m_w^* = \sqrt{2E_z/m_w^*}$ for propagation in the z-direction. In basic approximation, the thermal velocity v_{th} is used, corresponding to $E_z = k_{\text{B}}T$. The exact form of the DOS g reflects the transition of dimensionality from 2D to 3D at energies close to the barrier edge and can be captured in advanced descriptions based on scattering states in a real space basis representation, which do not discriminate between bound, quasibound, and continuum states, such as the Green's function approaches [73,74]. In most implementations, however, a combination of 2D square well DOS and 3D bulk DOS is used [71]. Finally, the occupation function f needs to encode the nonequilibrium steady-state carrier distribution under illumination and bias voltage. In general, this is achieved by using Fermi or Boltzmann statistics together with a quasi-Fermi level (QFL) for the carriers in the QW. In the case of Boltzmann statistics, the QFL drops out of the expression for the escape time due to the normalization to the total carrier density [75]. The advantage of the above approach is that it covers any combination of thermionic, direct tunneling, and thermally assisted tunneling escape. The drawback is that it requires assumptions on the carrier distribution in the QW and it does not contain any information on the microscopic mechanisms that lead to the carrier population at a certain energy. Such mechanisms include, e.g., the very fast electron-electron scattering that establishes a hot-carrier quasi-equilibrium at a fs timescale. But even the slower scattering of electrons with longitudinal optical (LO) phonons was shown to produce a hot-carrier population via subband scattering that enables fast thermally assisted carrier escape [76–78]. This latter result, while based on simple application of FGR to electron-LO-phonon scattering between electronic subbands, was confirmed by the state-of-the-art quantum-kinetic simulation of the phonon-assisted tunneling escape [79].

In **QD** devices, in addition to continuum (3D) and confined (0D) QD states, the 2D states of the WL play an important role in the dynamics of carrier transitions. Experimentally, similar thermionic and tunneling escape processes were identified [80–82]. However, thermal escape is much less efficient than in QW, and larger fields are required to enable tunneling escape. In terms of microscopic mechanisms, evidence for both Auger scattering [83] and (multi)phonon absorption [84] was found. While the carrier-carrier scattering depends strongly on carrier density and decreases quickly with QD size due to shrinking Coulomb (Cb) matrix elements—and is, therefore, less prominent in QW—the phonon processes depend critically on the energetic spacing of the states involved in the transitions, which produces strong resonances in the size dependence of the scattering mechanism. However, the lifetime broadening of QD states due to dynamical processes and delocalization increasing with energy alleviate the impact of the discrete DOS. Finally, in addition to the nonradiative escape processes, QDs offer the possibility of photon-assisted or purely radiative escape via subband transitions as described in Section 41.2.1. In QW, the subband absorption is strongly suppressed for in-plane polarization; hence, special photon management is required to scatter light into large angles, preferentially with coupling to guided modes.

The inverse process to escape is carrier *capture*. In SCs, this process is relevant in MQW and MQD devices, where photogenerated carriers that escaped to the current-carrying continuum states are recaptured into confined states on their way to the contacts. Since on the other hand, this process is instrumental for light-emitting devices, its theoretical description and numerical simulation have received considerable attention and are extensively covered in the literature. For **QW**, traditional models consider the FGR rate for transitions from 3D bulk states to confined states via the emission of LO-phonons [85–95] or carrier-carrier scattering [96,97], in combination with population dynamics from rate equations [98,99] or Monte Carlo simulations [100–104], while more recent approaches are based on quantum-kinetic theory [105]. At the relatively low densities of injected carriers at the operating point of a SC (which is below the bias regime of LED operation), relaxation due to emission of phonons was found to dominate over the effects of carrier-carrier scattering. In the case of **QD** devices, capture from the WL and subband relaxation were evaluated in the FGR framework for Auger processes [106,107], including the effects of line broadening induced by the carrier dynamics [108], and for the emission of single phonons [109,110] and multiple phonons [111], with extension to a polaron picture with finite phonon lifetime [112] indicated due to the strong coupling of QD states to phonons. The mechanisms were also combined in unified theories on semiclassical [113] and quantum-kinetic [114] levels.

41.2.2.3 Transport to and Extraction at Carrier-Selective Contacts

The choice of the picture for charge transport is primarily dictated by the kind of NS device architecture under consideration. In the second and third types of NSSC devices displayed in Figure 41.1, the NS states are involved not only in the absorption, but also in the carrier transport. These states thus need to be to some degree extended, which amounts to the requirement of coupling between individual NS. Indeed, most of the concepts of this kind are based on the formation of minibands in superlattices of QWs or QDs. For that situation, transport may be described as band-like, with a low-field Bloch-type electron *mobility* $\mu = \sigma/(\rho e)$ defined in terms of the conductivity tensor σ, and in the diffusive limit is obtained from the linearized Boltzmann transport equation (BTE) [115]:

$$\mu_{\alpha\beta} = e\left[\sum_{n,\mathbf{K}} \tau_{n\mathbf{K}} v^{\alpha}_{n\mathbf{K}} v^{\beta}_{n\mathbf{K}} \left(-\frac{\partial f^{0}_{n\mathbf{K}}}{\varepsilon_{n\mathbf{K}}}\right)\right] \Big/ \sum_{n,\mathbf{K}} f^{0}_{n\mathbf{K}}. \tag{41.22}$$

In Equation 41.22, n is the band index, $\mathbf{K}$ the superlattice wave vector, $\tau_{n\mathbf{K}}$ is the relaxation time, $\varepsilon_{n\mathbf{K}}$ is the miniband dispersion, $f^{0}_{n\mathbf{K}}$ is the equilibrium Fermi-Dirac electron distribution function, and $v^{\alpha}_{n\mathbf{K}} = \hbar^{-1}\partial_{K_\alpha}\varepsilon_{n\mathbf{K}}$ is the band velocity. The relaxation time depends on the scattering processes that are present in the device and can be computed—in some approximation—from the corresponding FGR formalism [116]:

$$\tau^{-1}_{n\mathbf{K}} = \sum_{n'\mathbf{K}'} \left(\frac{1 - f^{0}_{n'\mathbf{K}'}}{1 - f^{0}_{n\mathbf{K}}}\right) \left\{1 - \cos(\theta_{\mathbf{K}\mathbf{K}'})\right\} P^{nn'}_{\mathbf{K}\mathbf{K}'} \tag{41.23}$$

where $\cos(\theta_{\mathbf{K}\mathbf{K}'}) = (\mathbf{v}_{n\mathbf{K}} \cdot \mathbf{v}_{n'\mathbf{K}'})/(|\mathbf{v}_{n\mathbf{K}}||\mathbf{v}_{n'\mathbf{K}'}|)$ and $P^{nn'}_{\mathbf{K}\mathbf{K}'}$ is the FGR rate for scattering from the occupied state $|n\mathbf{K}\rangle$ to the empty state $|n'\mathbf{K}'\rangle$. However, even in the ideal case of a perfect miniband, the associated Bloch mobility may be critically low, as in the case of silicon QDs in a dielectric matrix material [117]. In realistic situations, due to the presence of built-in fields and any kind of spatial, configurational and compositional disorder, the NS states usually show a high degree of localization such that tunneling between NS is restricted to nearest neighbors. In addition to inducing localization, these effects lead to a misalignment of energy levels in adjacent NS, with the formation of Wannier-Stark ladders in the extreme case of very strong fields [118], in which situation transport is only possible via an inelastic scattering process and is best described by hopping between localized states. In coupled QD systems, due to the sparsity of the DOS, the presence of charge on the NS can lead to pronounced shifts in the energy level structure [53], inhibiting transport (Coulomb blockade regime). In devices where charge transport acquires molecular character, such as in colloidal QD solids, the hopping process involves even a change in configuration and may be phenomenologically described by the Marcus theory [44]. The bulk picture of charge carrier transport in band states does thus in general not provide a proper description of the real situation, which in addition to the remaining coherence and nonlocality effects should include all the localization effects and scattering mechanisms required to overcome energetic misalignment. The modification of the electronic structure due to the presence of NS also affects the scattering processes responsible for the limitation of mobility and for carrier relaxation, such as electron–phonon interaction, which becomes especially relevant in devices where this form of energy dissipation is sought to be suppressed, as in the hot-carrier concept, and in the case where extraction distances are on the order of the mean free path and transport approaches the ballistic regime.

The final step of the PV energy conversion consists in the extraction of the charge carriers at carrier-selective contacts. Carrier selectivity—which is an essential requirement on any PV device architecture [119]—can be described on a phenomenological level via the surface recombination velocities for majority and minority carriers used in the boundary conditions for carrier density/current in macroscopic transport equations. Microscopically, transport across the boundary to the contact depends again on the electronic structure of the interface and on the prevailing scattering processes. Since those can be largely engineered using NS, the latter also play an important role in the implementation of energy-selective [22] and carrier-selective [120] contacts.

41.2.3 Recombination

On the way to or from the contacts, the interaction of optically or electronically injected charge carriers with the EM and vibrational fields can induce different types of *recombination* (the mutual annihilation of an electron and a hole), which are usually classified in *radiative*, *Auger*, and *defect-mediated* processes. Like for the generation process, a proper consideration of NS effects on the recombination processes requires a careful consideration of the wave functions and local DOS together with the occupation of these states in general nonequilibrium conditions.

41.2.3.1 Radiative

Even in ideal, defect-free absorber materials, the principle of *detailed balance* dictates the presence of radiative recombination by spontaneous and stimulated photon emission as the inverse process to carrier generation by photon absorption. Conventionally, the local volume rate of radiative recombination is related to the charge carrier densities and the local optical material constants (α, n_r) as follows [121]:

$$r^{\mathrm{em}}(\mathbf{r}) = \mathcal{B}(\mathbf{r})\left\{\rho_e(\mathbf{r})\rho_h(\mathbf{r}) - \rho_e^0(\mathbf{r})\rho_h^0(\mathbf{r})\right\}, \tag{41.24}$$

$$\mathcal{B}(\mathbf{r}) = \rho_i^{-2}\int dE_\gamma\, \alpha(\mathbf{r}, E_\gamma)\bar{\Phi}_{bb}(E_\gamma), \tag{41.25}$$

where ρ^0 and ρ_i denote equilibrium and intrinsic carrier densities, respectively, and $\bar{\Phi}_{bb} = 4\pi\Phi_{bb}$ is the angle-integrated blackbody radiation flux, with

$$\Phi_{bb}(E_\gamma) = \frac{E_\gamma^2 n_r^2}{4\pi^3\hbar^3 c_0^2}\left\{\exp\left(\frac{E_\gamma}{k_B T}\right) - 1\right\}^{-1}. \tag{41.26}$$

In Equation 41.25, the assumption of emission in an optically homogeneous medium is made, with isotropic photon DOS. In thin-film SCs with nanophotonic light trapping, this is often not appropriate, and the radiative recombination should be modeled starting from the FGR for spontaneous emission under consideration of the actual density of final photon states. For NS, the emission is enhanced as compared to the bulk, due to carrier localization [122] resulting in maximized density product in Equation 41.25 and to larger overlap of electron and hole wave functions in Equation 41.13. A further point to note in the case of NSSC is the dependence of the absorption coefficient on the operating conditions, mainly the bias voltage via the effect of the built-in field on the local electronic structure, which should be reflected in the quantities used in Equation 41.25.

41.2.3.2 Auger

In the regime of large optical injection, relevant for the operation under high optical concentration, the PV performance of a defect-free absorber is limited by the interband Auger recombination mechanism [123], which as an intrinsic effect is unavoidable and needs thus to be considered for a realistic estimate of the limiting energy conversion efficiency. In this process originating in carrier-carrier scattering, electrons recombine with holes by giving up the excess energy to excite either another electron or another hole. In the semiclassical bulk continuum formulation, this rate takes the form [124]

$$r^{\text{Aug}}(\mathbf{r}) = \left[C_e(\mathbf{r})\rho_e(\mathbf{r}) + C_h(\mathbf{r})\rho_h(\mathbf{r})\right]\left\{\rho_e(\mathbf{r})\rho_h(\mathbf{r}) - \rho_e^0(\mathbf{r})\rho_h^0(\mathbf{r})\right\}, \tag{41.27}$$

where $C_{e/h}$ are the Auger coefficients for electron/hole excitation. For NS, some or all of the band indices are replaced by the labels of subbands (QW) or discrete levels (QD). The Auger coefficients can be computed microscopically by application of perturbation theory for the carrier-carrier interaction under consideration of the appropriate dimensionality of the (screened) Cb potential. For QW, FGR-type approaches were used to assess direct [125–129] and phonon-assisted [130] Auger processes, and the latter was also investigated using a Green's function technique [131]. For QD, similar FGR models were developed for colloidal QD [132] and applied to self-assembled III-V single QD [133] and QD arrays [134]. Recently, the complete dynamics of excitons in colloidal QD was also simulated on an *ab initio* level using a combination of density functional theory and molecular dynamics [135]. While in QW, the magnitude of the Auger recombination rate was found to be similar to bulk, it is strongly enhanced in QD due to the effects of wave function localization and associated relaxation of momentum selection rules.

41.2.3.3 Defect Mediated

While radiative and Auger losses are unavoidable and thus present even in structures without imperfections, recombination in real-world devices is usually dominated by defect-mediated processes where the carrier energy is dissipated as heat and is lost for the conversion process. Indeed, there is an inherent increase in surface area associated with the presence of NS, which are thus likely to act as centers of nonradiative recombination. In some cases, it is possible to reduce the NS-related defects via specially engineered strain-balancing techniques [136] or via passivation [137], but in general, the insertion of NS leads to losses in the open-circuit voltage of the SC. With regard to the modeling, the role of lower dimensional NS in the context of defect-mediated recombination has received little attention so far, and models are usually based on the conventional bulk Shockley–Read–Hall (SRH) theory [138,139], with modification to relate the recombination to interface or surface defects rather than volume defects, e.g., for QW [60]

$$r^{\mathrm{SRH}}(\mathbf{r}) = \frac{\rho_e(\mathbf{r})\rho_h(\mathbf{r}) - \rho_e^0(\mathbf{r})\rho_h^0(\mathbf{r})}{(\rho_e(\mathbf{r}) + \rho_e^t(\mathbf{r}))\tau_e + (\rho_h(\mathbf{r}) + \rho_h^t(\mathbf{r}))\tau_h}, \tag{41.28}$$

where $\tau_s^{-1} = \sigma_s v_{th} N_S / L_w$ $(s = e, h)$ with σ_s the carrier capture cross-section, v_{th} the thermal velocity, N_S the sheet density of interface defects, L_w the QW width, and ρ^t the carrier density for a Fermi level at the trap energy. The capture cross-sections depend on the local electronic structure of the NS and defect states, as well as on the phonon modes coupling to those states in the multiphonon relaxation mechanism [124], and can be obtained based on microscopic theories on phenomenological [140] and fully *ab initio* [141] levels.

41.3 Modeling of NSSC Device Characteristics

The basic function of a SC consists in the conversion of the solar energy flux impinging on its surface into electrical power. A central requirement on any theory suitable for the modeling of SC device operation is thus the ability to predict the PV energy conversion efficiency as defined via the ratio $\eta_{\mathrm{PV}} \equiv P_{el}^{\nearrow}/P_{\gamma}^{\swarrow}$ of generated electrical power to the incident radiative power, with

$$P_{el}^{\nearrow} = \max_V \{I(V) \cdot V\}|_{T_C}, \tag{41.29}$$

$$P_{\gamma}^{\swarrow} = \int dE_{\gamma} \int_S \int_{\Omega} d\boldsymbol{\Omega} \cdot d\boldsymbol{S}\, \Phi_{\gamma}^{\swarrow}(\mathbf{r}_S, \Omega, E_{\gamma}) E_{\gamma}, \tag{41.30}$$

where V is the terminal voltage (i.e., related to the separation of chemical potentials at the contacts), $I(V)$ is the charge current extracted from the device at this voltage, T_C is the temperature of the SC, E_{γ} is the photon energy, and $\Phi_{\gamma}^{\swarrow}$ denotes the incident spectral photon flux impinging on a cell of (outer) surface $\mathcal{S}$ under a solid angle of incidence $\Omega = (\theta, \varphi)$. Hence, any PV model needs to provide I as a function of V and $\Phi_{\gamma}^{\swarrow}$ at given T_C.

As outlined in Sections 41.1 and 41.2, depending on the functionality of the NS within the SC device, some or all of the fundamental PV processes are affected by the peculiar physical properties of the NS component, with major implications for the requirements on a suitable simulation approach. The general approach reflects the multiscale and multiphysics scheme depicted in Figure 41.2: Starting from the NS states, the dynamics giving the occupation of these states is evaluated and the resulting generation and recombination rates are used to obtain the device characteristics. Three levels of sophistication that can be identified in the hierarchy of theoretical descriptions of SC device operation providing the current–voltage characteristics $I(V, \Phi_{\gamma}^{\swarrow})$, depending on the explicity of consideration of the various physical mechanisms, are: *detailed balance* theories providing efficiency limits of idealized systems, *semiclassical* transport models with detailed balance generation-recombination rates, and *quantum-kinetic* theories with consistent and fully microscopic treatment of charge carrier dynamics and transport [142]. These three approaches shall now be discussed in some more detail.

41.3.1 Global Detailed Balance Theories for Ideal Systems

On the most basic level, NS properties are considered only in generation and recombination at the radiative limit, while transport is assumed to be ideal. This means that from a device point of view, the system is treated as a black box, and there is no information on geometry or any spatial inhomogeneity in the electronic system that would depart from effective bulk properties. The most prominent examples of this

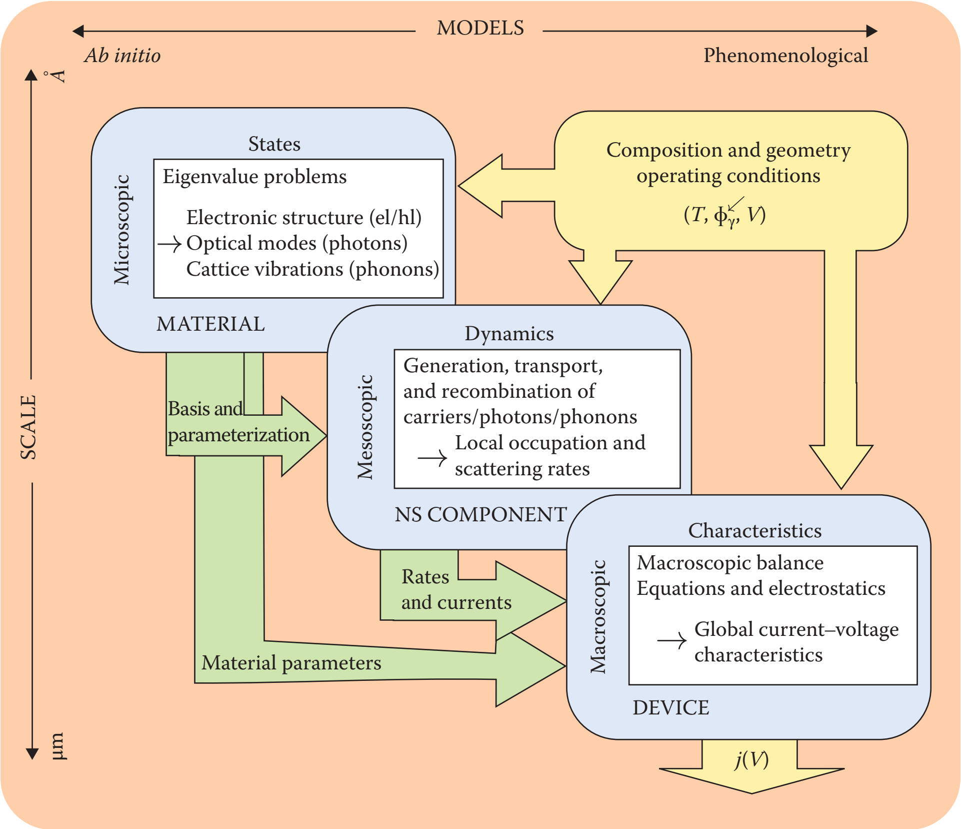

MODELS
Ab initio
Phenomenological
SCALE
Å
μm
States
Microscopic
Eigenvalue problems
Electronic structure (el/hl)
→ Optical modes (photons)
Cattice vibrations (phonons)
MATERIAL
Composition and geometry
operating conditions
(T, Φ↙γ, V)
Basis and
parameterization
Dynamics
Mesoscopic
Generation, transport,
and recombination of
carriers/photons/phonons
→ Local occupation and
scattering rates
NS COMPONENT
Rates
and currents
Material parameters
Characteristics
Macroscopic
Macroscopic balance
Equations and electrostatics
→ Global current–voltage
characteristics
DEVICE
j(V)

FIGURE 41.2 Schematic representation of the multiscale and multiphysics approach to the simulation of nanostructure-based solar cell devices.

category are the detailed balance models [143] used to obtain an upper bound to the energy conversion efficiency, such as the SQ limit. In this approach, the current–voltage characteristics are obtained by equating the terminal current with the difference of the number of photons per unit time entering and leaving the absorber volume at given terminal bias voltage [143,144]:

$$I(V,\Phi_\gamma^{\swarrow})|_{T_C} = -q\int \mathrm{d}E_\gamma \int_{\mathcal{S}}\int_{\Omega} \mathrm{d}\boldsymbol{\Omega}\cdot \mathrm{d}\boldsymbol{\mathcal{S}}\,\left\{\Phi_{\mathrm{abs}}(\mathbf{r}_{\mathcal{S}},\Omega,E_\gamma)-\Phi_{\mathrm{em}}(\mathbf{r}_{\mathcal{S}},\Omega,E_\gamma)\right\}|_{V,\Phi_\gamma^{\swarrow},T_C}. \tag{41.31}$$

At the radiative limit, where the recombination is determined by the emission, and assuming unit generation efficiency, the electronic equivalent to the above balance relation is obtained from the integration of the charge continuity equation over the absorber volume $\mathcal{V}$ (e/h: electrons/holes):

$$I(V,\Phi_\gamma^{\swarrow})|_{T_C} = \mp q\int_{\mathcal{V}} d^3\mathbf{r}\left\{\mathcal{G}_{e/h}(\mathbf{r})-\mathcal{R}_{e/h}(\mathbf{r})\right\}|_{V,\Phi_\gamma^{\swarrow},T_C}. \tag{41.32}$$

The absorbed photon flux is derived in terms of the absorptance—or, synonymously, the absorptivity—*a* of the device, and of the reflection coefficient *r* for incident light, via

$$\Phi_{\mathrm{abs}}(\mathbf{r}_{\mathcal{S}},\Omega,E_\gamma) = [1-r(\mathbf{r}_{\mathcal{S}},\Omega,E_\gamma)]a(\mathbf{r}_{\mathcal{S}},\Omega',E_\gamma)\left\{\Phi_\gamma^{\swarrow}(\mathbf{r}_{\mathcal{S}},\Omega,E_\gamma)+\Phi_{\gamma 0}(E_\gamma,\Omega,T_C)\right\}, \tag{41.33}$$

where Ω' is the solid angle after refraction, related to the former by Fresnel's law [145]. The second term in the curly brackets is the isotropic thermal equilibrium flux and is given by

$$\Phi_{\gamma 0}(E_\gamma, \Omega, T_C) = \Phi_{bb}(E_\gamma, T_C) \cos\theta, \tag{41.34}$$

where Φ_{bb} is the blackbody radiation flux (Equation 41.26) introduced in Section 41.2.3. The generation rate in terms of local photon flux and absorption coefficient is given in Equation 41.2. However, for equivalence with the above result, photon recycling needs to be considered in the evaluation of the local photon flux. At unit generation efficiency and in the absence of nonradiative recombination, the product in front of the incident flux in Equation 41.33 is identified as the *external quantum efficiency* (*EQE*). Under the assumption of perfect transport, i.e., infinite mobility, which amounts to flat QFLs split by the terminal voltage, i.e., $\Delta\mu = qV$, and of an optically isotropic medium, detailed balance then provides the emitted photon flux as a function of the absorptance [144,146]:

$$\Phi_{\mathrm{em}}(\mathbf{r}_S, \Omega, E_\gamma)|_V = [1 - r(\mathbf{r}_S, \Omega, E_\gamma)]a(\mathbf{r}_S, \Omega', E_\gamma)\tilde{\Phi}_{bb}(\Omega, E_\gamma, T_C, V) \tag{41.35}$$

$$\equiv EQE(\mathbf{r}_S, \Omega, E_\gamma)\tilde{\Phi}_{bb}(\Omega, E_\gamma, V), \tag{41.36}$$

where:

$$\tilde{\Phi}_{bb}(\Omega, E_\gamma, T_C, V) = \Phi_{bb}(E_\gamma - qV, T_C) \cos\theta \tag{41.37}$$

$$\approx \Phi_{\gamma 0}(E_\gamma, \Omega, T_C) \mathrm{e}^{qV/(k_B T_C)} \qquad (E_\gamma \gg qV). \tag{41.38}$$

The approximation in Equation 41.38 permits a compact formulation of the current–voltage characteristics:

$$\begin{aligned} I(V, \Phi_\gamma^{\swarrow})|_{T_C} \approx & -q \int dE_\gamma \int_S \int_\Omega d\mathbf{\Omega} \cdot d\boldsymbol{S}\, EQE(\mathbf{r}_S, \Omega, E_\gamma) \\ & \times \left\{ \Phi_\gamma^{\swarrow}(\mathbf{r}_S, \Omega, E_\gamma) - \Phi_{\gamma 0}(E_\gamma, \Omega, T_C)\left[\mathrm{e}^{qV/(k_B T_C)} - 1\right] \right\}. \end{aligned} \tag{41.39}$$

Finally, the local recombination rate for the determination of the radiative dark current via Equation 41.32 is given by Equation 41.25.

In the SQ limit, the absorptance is a unit step function with threshold energy at the value of the bandgap E_g, which is then the only electronic parameter entering the calculation of the current–voltage characteristics. This simplifies the expressions for the absorbed and emitted photon fluxes considerably. In a first step beyond the SQ analysis, the effect of incomplete absorption with non–abrupt onset as well as the impact of light-trapping structures can be considered by using in Equations 41.33 and 41.35 the proper absorptance computed by optical models for the light propagation as outlined in Section 41.2. It is common to neglect the bias dependence of the photocurrent also in situations where a realistic absorptance is considered, which results in current–voltage characteristics given by the *superposition* of photocurrent I_γ and dark current I_{d}:

$$I(V, \Phi_\gamma^{\swarrow}) = I_\gamma(\Phi_\gamma^{\swarrow}) + I_{\mathrm{d}}(V). \tag{41.40}$$

However, this tends to be inappropriate in nanoscale absorbers due to the strong impact of built-in fields and finite state filling on the local absorption coefficient [147,148].

Due to its formal simplicity, the detailed balance analysis of the radiative limit is applied to most novel NS PV devices in a first assessment of the performance potential. On the other hand, the restrictive assumptions made are seldom met in realistic devices, and the upper efficiency bound obtained in this way may therefore be far from a realistic estimate.

Within the above setting of idealized conditions, an alternative starting point for the device characteristics is a modification of the ideal diode equation [149]:

$$I(V, \Phi_{\gamma}^{\swarrow}) = I_0\left(e^{qV/(k_B T_C)} - 1\right) - I_G(\Phi_{\gamma}^{\swarrow}) + I_R(V), \tag{41.41}$$

which is formally equivalent to Equation 41.39, but where the generation and recombination currents I_G and I_R are composed of the contributions of the bulk "baseline" cell and of the NS components. The NS contributions are related to those of the baseline using correction factors for quantities such as geometry (fraction of NS material), oscillator strength, and DOS (for radiative transitions). It needs to be noted, however, that such modifications still have to comply with the detailed balance relations between absorption and emission as given above. Furthermore, even in its original formulation devised for the radiative limit, the theory goes beyond that limit in considering in the reverse saturation current I_0 the standard nonradiative minority carrier diffusion contribution of an ideal p-(i-)n diode.

41.3.1.1 Application to QWSC

Soon after the proposal of the QWSC concept, the detailed balance approach was used to show that under the idealized conditions of that theory, the QWSC could not provide an efficiency in excess of that of an ideal gap bulk cell, since both photocurrent generation and radiative recombination are based on the same modified absorptivity [150,151]. An SQ-type detailed balance model was then used to demonstrate that QWSC efficiency could still exceed the bulk limit if variations of the quasi-Fermi level splitting (QFLS) between bulk and QW material were allowed for [152]. This was shown to be equivalent to the detailed balance picture of the intermediate band concept, with the difficulty of the QW featuring suppressed subband absorption and fast population equilibration due the continuous electronic DOS [153]. Indeed, the issue of QFLS variation in QWSC is still a matter of debate and ongoing research. In more recent QWSC developments, the detailed balance approach was used to model tandem configurations and the impact of directional emission in strained QW [23]. Following the original work on the ideal diode QWSC model [149], a number of variations and refinements of this semianalytical approach were devised for MQWSC [154–161] and QWSLSC [159,162–164].

41.3.1.2 Application to QDSC

For QD intermediate band SC (QDIBSC), the original detailed balance IBSC model [165] was adapted to account for the specific absorption characteristics of QD structures [166,167] and was implemented for specific material systems using a Kronig-Penney model for supracrystals of InAsN/GaAsSb QD [168], InGaN/GaN QD [169], and InGaN/InN [170], as well as a $\mathbf{k} \cdot \mathbf{p}$ model for isolated InAs/GaAs QD [171]. The ideal diode theory was applied to theoretically investigate an InAs/GaAs MQD concept with AlGaAs fences to prevent capture of photogenerated carriers [172].

41.3.2 Hybrid Models: Semiclassical Transport with Detailed Balance Rates

The main shortcoming of global detailed balance approaches and ideal diode theories is the lack of a realistic consideration of photocarrier transport, i.e., the assumption of unit collection efficiency, which is often inappropriate in NSSC devices. In order to account for the effects of finite mobility, actual transport equations have to be solved for the charge carriers. The main difficulty in that undertaking resides in the consideration of the contribution of NS states with a higher degree of localization. To obtain the self-consistent occupation of confined NS states (c) and bulk host states (b), separate but coupled rate equations need to be formulated for the densities of carriers occupying different types of states. Under the standard assumption of continuum (i.e., no atomistic information) and complete thermalization (i.e., occupation is characterized by a local QFL), the current–voltage characteristics are obtained from the *drift-diffusion* (DD) charge current for carrier species $s \in \{e, h\}$:

$$\mathbf{J}_{\mathrm{s,i}}(\mathbf{r}) = -q\big\{\mu_{\mathrm{s,i}}(\mathbf{r})\rho_{\mathrm{s,i}}(\mathbf{r})\nabla\phi(\mathbf{r}) \mp D_{\mathrm{s,i}}(\mathbf{r})\nabla\rho_{\mathrm{s,i}}(\mathbf{r})\big\}, \qquad i = b, c \tag{41.42}$$

where μ denotes the mobility, $\rho_{e/h}$ denotes the charge carrier densities, D is the diffusion coefficient, ϕ is the electrostatic potential, q is the elementary charge, and the upper (lower) sign is for electrons (holes). In this general formulation, no restriction is made on the charge transport in NS states, which allows for application to all three NS architectures introduced in Section 41.1. The charge carrier densities and the electrostatic potential that determine the charge currents in Equation 41.42 are obtained from the coupled solution of the steady-state continuity equations:

$$\mp\frac{1}{q}\nabla\cdot\mathbf{J}_{s,i}(\mathbf{r}) = \sum_{j\neq i}\big\{\mathcal{G}_{s,j\to i}(\mathbf{r}) - \mathcal{R}_{s,i\to j}(\mathbf{r})\big\}, \tag{41.43}$$

where $\mathcal{G}/\mathcal{R}$ is the carrier (volume) generation/recombination rate due to transitions between (sub)bands and the upper (lower) sign applies again to electrons (holes), and Poisson's equation for the total charge density:

$$\epsilon_0\nabla\cdot\big\{\varepsilon(\mathbf{r})\nabla\phi(\mathbf{r})\big\} = -\,q\Big\{\sum_i \rho_{\mathrm{h,i}}(\mathbf{r}) - \sum_j \rho_{\mathrm{e,j}}(\mathbf{r}) + N_{\mathrm{dop}}(\mathbf{r})\Big\}, \tag{41.44}$$

where ε is the static dielectric function and $q\cdot N_{\mathrm{dop}}$ is the net charge density due to ionized dopants. The boundary conditions used in the solution of the coupled equations are the same as in the bulk case, reflecting the nature of the physical contact in terms of (Schottky-)barriers and surface recombination. In Equation 41.43, the generation rate $\mathcal{G}_{s,j\to i} = \mathcal{G}_{s,j\to i}[\Phi_\gamma(\mathbf{r}), \alpha_{j\to i}(\mathbf{r}, E_\gamma)]$ $(i = b, c)$ is a functional of the local photon flux and the local absorption coefficient as given by Equation 41.2 and the recombination rate $\mathcal{R}_{s,i\to j} = \mathcal{R}_{s,i\to j}[\rho_{e,i}, \rho_{h,j}, \ldots]$ depends on the carrier densities and on recombination mechanism (radiative, Auger, defect-mediated)–specific parameters. In the case where, due to strong localization as for isolated QW or QD absorbers, direct transport between NS states can be neglected, the current term in Equation 41.43 is absent for these states, and exchange of carriers between localized states always proceeds via scattering to and from extended states. The coupling of the equations for localized and extended states is then provided on the level of a pure rate equation, where the rates are subject to the detailed balance condition and are commonly expressed via the density in the initial state and an associated lifetime depending on the scattering process, e.g.,

$$\mathcal{R}^{x}_{e,i\to j} = \frac{\rho_{e,i}}{\tau^{x}_{e,i\to j}}, \qquad \text{x: scattering mechanism.} \tag{41.45}$$

The lifetimes τ (which are local quantities) are either used as fitting parameters to reproduce experimental characteristics, or derived from an appropriate description for the microscopic mechanisms of the scattering process responsible for the coupling as reviewed in Section 41.2, e.g., via FGR, based on the solution of the Schrödinger-Poisson problem for the NS states and energies. For consistency, the macroscopic material parameters based on microscopic information, such as absorption coefficients and mobilities used in the generation term and the current expressions, should be computed on the basis of the same solutions of the microscopic equations for the electronic structure. It needs to be emphasized at this point that the lifetimes themselves still depend on the occupation of the states, e.g., for the radiative recombination, $\tau^{\mathrm{em}}_{e}(\mathbf{r}) \approx \big[\mathcal{B}(\mathbf{r})\rho_h(\mathbf{r})\big]^{-1}$, where the very small equilibrium term in (Equation 41.24) is neglected.

In the above formulation of the charge current, the local electronic structure does not appear as a parameter that could be determined microscopically. If the Einstein relation between mobility and diffusion coefficient, i.e., $D = \mu k_B T/q$, and the Boltzmann approximation for the carrier distribution function are valid, an equivalent formulation of the charge carrier densities and currents in terms of effective DOS $\mathcal{N}$ and QFL E_F can be used:

$$\rho_{\mathrm{s,i}}(\mathbf{r}) = \int dE\,\mathcal{D}_{s,i}(\mathbf{r},E) f_{s,i}(E) \approx \mathcal{N}_{s,i} \exp\left(\frac{\pm E_{F_{s,i}}(\mathbf{r}) \mp \varepsilon_{s,i}(\mathbf{r})}{k_B T}\right), \tag{41.46}$$

$$\mathcal{N}_{e,i} \equiv \int_{\varepsilon_{e,i}}^{\infty} dE\,\mathcal{D}_{e,i}(\mathbf{r},E) \exp\left(\frac{\varepsilon_{e,i}(\mathbf{r}) - E}{k_B T}\right), \tag{41.47}$$

$$\mathcal{N}_{h,i} \equiv \int_{-\infty}^{\varepsilon_{h,i}} dE\,\mathcal{D}_{h,i}(\mathbf{r},E) \exp\left(\frac{E - \varepsilon_{h,i}(\mathbf{r})}{k_B T}\right), \tag{41.48}$$

$$\mathbf{J}_{\mathrm{s,i}}(\mathbf{r}) = \pm\, \mu_{\mathrm{s,i}}(\mathbf{r}) \rho_{\mathrm{s,i}}(\mathbf{r}) \nabla E_{F_{s,i}}(\mathbf{r}), \tag{41.49}$$

with upper (lower) sign applying to electrons (holes) and $\mathcal{D}$ denoting the DOS for carrier species s and band i, with $\varepsilon_{s,i}$ the minimum (maximum) energy associated ("band edge"). This approach, in addition to the implicit consideration of the microscopic DOS for NS states, provides the possibility to study the emergence of distinct QFL for the NS populations, which is instrumental in PV concepts such as the IBSC.

The *hybrid* (macroscopic-microscopic) approach outlined above can be used to reproduce experimental device characteristics with remarkable accuracy. However, being a macroscopic and local model, any situation requiring energy resolution, nonlocality, or coherence cannot be described properly, e.g., nonthermalized carrier distributions or resonant tunneling. To include such processes in a consistent description, transition to a truly microscopic picture of carrier transport is indicated.

41.3.2.1 Applications to QWSC

Semiclassical DD Poisson models with detailed balance rates for generation and recombination were applied to III-V MQWSC [60,173,174] and to Si-SiO$_x$ QWSLSC [175]. Simplifications include rate equations with semianalytical solutions for current contributions [176,177] and the analytical solution of diffusion equation for excess minority carriers in neutral layers [57,154,178].

41.3.2.2 Applications to QDSC

The above mentioned DD Poisson approach with local rates was implemented in a comprehensive fashion for MQDSC [68,179,180] and QD-IBSC [181–183]. Again, there are also approximations in the form of analytical solutions of the diffusion equation [184–186], of simple rate equations for the population of bulk and NS states [187], or a combination of the two [188].

41.3.3 Fully Microscopic Quantum-Kinetic Models

In situations where the NSs are also involved in transport and the latter is not entirely band-like or purely diffusive, a consistent microscopic picture of the PV processes can be formulated by means of nonequilibrium quantum statistical mechanics using the nonequilibrium Green's function formalism (NEGF) [189]. In the NEGF picture, the device current density $\mathbf{J} = \mathbf{J}_n + \mathbf{J}_p$ is given in terms of the charge carrier Green's function components as follows:

$$\mathbf{J}_{n/p}(\mathbf{r}) = \pm \lim_{\mathbf{r}'\to\mathbf{r}} \frac{e\hbar}{m_0} \left(\nabla_{\mathbf{r}} - \nabla_{\mathbf{r}'}\right) \int \frac{dE}{2\pi} G^{\lessgtr}(\mathbf{r},\mathbf{r}';E). \tag{41.50}$$

The Green's functions in Equation 41.50 are determined by the steady-state Keldysh equation:

$$G^{\lessgtr}(\mathbf{r},\mathbf{r}';E)=\int d\mathbf{r}_1\int d\mathbf{r}_2 G^R(\mathbf{r},\mathbf{r}_1;E)\Sigma^{\lessgtr}(\mathbf{r}_1,\mathbf{r}_2;E)G^A(\mathbf{r}_2,\mathbf{r}';E), \tag{41.51}$$

with retarded and advanced components resulting from the Dyson equation:

$$\int d\mathbf{r}_1\left[\left\{G_0^{R/A}\right\}^{-1}(\mathbf{r},\mathbf{r}_1;E)-\Sigma^{R/A}(\mathbf{r},\mathbf{r}_1;E)\right]G^{R/A}(\mathbf{r}_1,\mathbf{r}';E)=\delta(\mathbf{r}-\mathbf{r}'). \tag{41.52}$$

While the retarded and advanced Green's functions $G^{R/A}$ are related to the charge carrier DOS, the correlation functions $G^{\lessgtr}$ additionally contain information on the (nonequilibrium) occupation of these states. The noninteracting system is described by G_0. The self-energies Σ, on the other hand, are scattering functions describing the renormalization of the Green's functions due to coupling to the environment in the form of interactions with photons (mediating photogeneration and radiative recombination), phonons (mediating relaxation and indirect transitions), and other carriers (relevant for excitonic and Auger processes). An additional self-energy term describes injection and extraction of carriers as contacts with arbitrary chemical potential, enabling the treatment of an open nonequilibrium system. The computation of the NEGF is self-consistently coupled to the Poisson equation (Equation 41.44) for the mean-field electrostatic potential ϕ—entering the equation for G_0—via the charge carrier density:

$$\rho_{e/h}(\mathbf{r})=\mp i\int\frac{dE}{2\pi}G^{\lessgtr}(\mathbf{r},\mathbf{r};E). \tag{41.53}$$

For an in-depth review of the general NEGF formalism and its application to NS PV, the reader is referred to [189] and references therein.

In contrast to the semiclassical models presented in Section 41.3.2, the microscopic approach provides insight into the physical mechanisms underlying the PV device operation, as those are properly captured by the formalism. On the downside, the method is extremely demanding from a computational point of view and applications are therefore limited to functional NSSC components of mesoscopic extension. An important direction of future research will thus consist in the combination of quantum-kinetic models for local carrier dynamics in NS regions with semiclassical models for the global characteristics of the extended device.

41.3.3.1 Applications to Nanowire SCs

As discussed in Section 41.1, QWRs are not in the focus of efficient solar energy harvesting systems. However, the low dimensionality allows for an efficient numerical treatment of the system. Indeed, the earliest application of NEGF for PV energy conversion considered quasi-1D carbon nanotube systems with tight-binding (TB) band structure [190–192]. More recently, a $\mathbf{k}\cdot\mathbf{p}$-NEGF approach was formulated to study photocurrents in GaAs nanowires [193].

41.3.3.2 Applications to QWSC

Single and double QW structures in III-V material systems were studied in [194,195] using a simple TB band structure and in [79,196] within an effective mass approach, with focus on the modification of absorption and emission spectra by the complex potential profile and on the escape of carriers from confined to extended, current-carrying states. Photocarrier transport regimes from ballistic extraction to phonon-mediated sequential tunneling in SL absorbers were investigated for Si-SiO$_x$ [197] (including phonon-mediated indirect transitions) and InAlGaAs-InGaAs [198].

41.3.3.3 Applications to QDSC

The main challenge for the implementation of NEGF models for QDSC is the complexity of the electronic structure with respect to 3D spatial variation, which prevents the use of the slab models conveniently applied to planar NSs. A simple approach is to use a coarse-grained model with a basis of localized QD orbitals—e.g., the Wannier functions of a regular QDSL—in the spirit of the TB approximation in atomic systems. This method was used to formulate a NEGF model for Si QDSL absorbers embedded in the SiO_x-SiC matrix material [199], and to study the impact of inter-QD and QD-contact coupling on the PV performance of generic QD arrays [200]. Within such an approach, however, it is not possible to study on a microscopic level, similar to that achieved in MQWSC, the carrier escape and capture processes that are instrumental in the operation of MQDSC.

41.4 Case Study: Single Quantum Well p-i-n SC

In this section, the different elements and levels of sophistication in the simulation of NSSC shall be illustrated by consideration of a prototypical device architecture, for which a single quantum well (SQW) p-i-n photodiode is chosen. The device as displayed in Figure 41.3 consists of a 100-nm thin bulk GaAs baseline SC with heavily doped emitter and base components ($N_A = N_D = 10^{18}$ cm^{-3}) of 20-nm extension. In the center of the intrinsic region, a slab of 10-nm InGaAs is inserted, which forms a type-I QW. For the band offsets, $\Delta E_{C/V} = 0.15/0.10$ eV is chosen. The contacts are designed to be perfectly carrier selective, which corresponds to the assumption of ohmic boundary conditions for majorities and infinite barriers for minorities. The light is assumed to be incident from the p-side, whereas a gold reflector is applied to the *n*-side, which, however, is not considered in the electronic simulation.

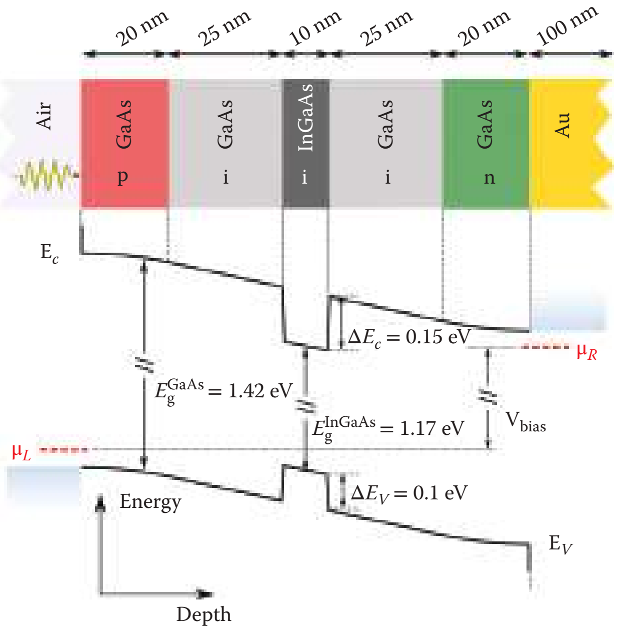


FIGURE 41.3 Single GaAs-InGaAs quantum well p-i-n solar cell architecture used as a test case for the simulation. The contacts are designed to be perfectly selective, i.e., with ohmic boundary conditions for majorities and infinite barriers for minorities. The gold reflector is used only in the optical simulation.

41.4.1 States

In the first step of the simulation, the electronic and optical device states need to be determined. For the description of the electronic structure of the component semiconductor bulk materials, a basic combination of two decoupled single-band effective mass approximation (EMA) models is used, with $m_{e/h}^{*\mathrm{GaAs}} = 0.067/0.1\ m_0$ and $m_{e/h}^{*\mathrm{InGaAs}} = 0.055/0.22\ m_0$. Under the standard thin-film assumption of a transverse continuum due to very large spatial extension perpendicular to the growth direction, the wave functions (WF) can be written as the decomposition

$$\Psi_{n\alpha\mathbf{k}_{\parallel}}(\mathbf{r}) = \frac{e^{i\mathbf{k}_{\parallel}\mathbf{r}_{\parallel}}}{\sqrt{\mathcal{A}}}\psi_{n\alpha\mathbf{k}_{\parallel}}(z)u_{n0}(\mathbf{r}), \tag{41.54}$$

where n is the bulk Bloch band index, $||$ denotes transverse quantities, $\mathcal{A}$ is the transverse cross-section area, ψ is the envelope function in the growth direction, and u_{n0} is the periodic part of the Bloch function at the Γ point. The 1D EMA Schrödinger equation for the envelope functions of the QW states is solved numerically for closed-system (Dirichlet) boundary conditions and at vanishing field. The wave functions $\psi_{c/v}(z)$ (z: depth) and eigenenergies $\varepsilon_{c/v}$ of the resulting NS states at $\mathbf{k}_{\parallel} = 0$ are displayed in Figure 41.4a. There are two electron and three hole states with energies below the band edge of the host material, exhibiting the expected symmetry properties and increasing barrier penetration as a function of energy. For comparison, the local density of states (LDOS) at $\mathbf{k}_{\parallel} = 0$ is computed using the NEGF formalism for identical EMA Hamiltonian and at flat potential (Figure 41.4b), but with open boundary conditions, which, however, does not affect either shape or energies of the confined states. While the solutions of the square well potential are readily available, the situation in the device at operating conditions is quite different, as demonstrated by Figure 41.4c, where the LDOS is shown for a realistic band profile at an operating forward bias voltage

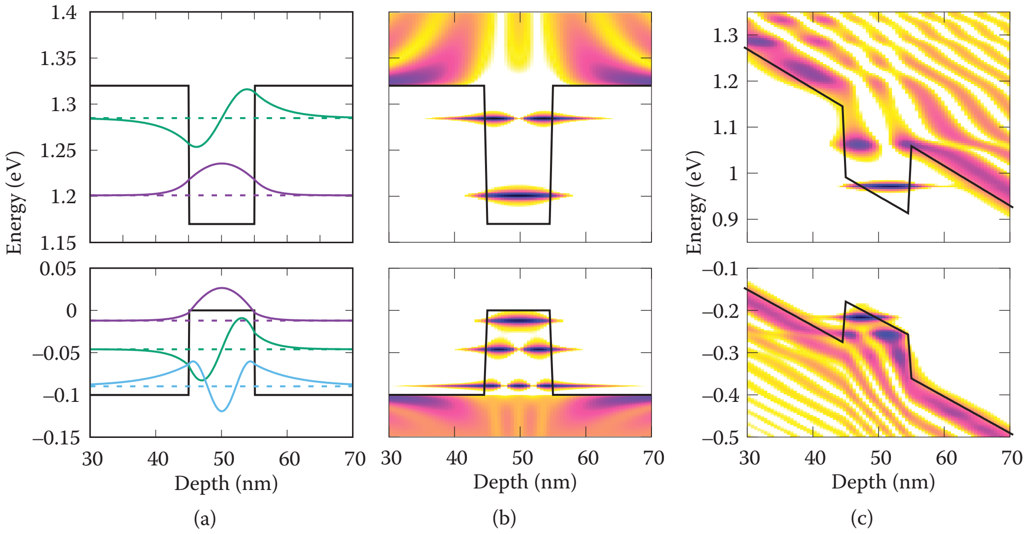


FIGURE 41.4 (a) Wave functions and energies of the eigenstates in the 10-nm InGaAs QW embedded in a GaAs host as obtained by the solution of the 1D single band effective mass Schrödinger equation along the growth direction. (b) One dimensional 1D LDOS ($\mathbf{k}_{\parallel} = 0$) obtained using the NEGF formalism with the same effective mass Hamiltonian and at flat band conditions. (c) 1D LDOS of the same SQW, but in the p-i-n device configuration close to operating conditions at an applied forward bias voltage $V = 0.8$ V. The most remarkable feature is the unbinding of the higher lying QW states.

of $V = 0.8$ V. Most remarkably, the higher lying QW states lose their confinement in the large built-in field and turn into quasibound resonances.

The optical modes of the device are computed under consideration of the absorbing components and of the reflector using an in-house developed wave propagation solver based on the TMM. As will be shown, this is more appropriate than simple consideration of LB law, since both slab and reflector form an optical cavity, which can lead to nontrivial spatial variation of the optical intensity inside the absorber. Since the states of the quasicontinuum in the QW material are not well described by the chosen approach with closed boundaries, the optical response will be computed only for energies below the host gap. The determination of the optical absorption strength based on the computed electronic states thus forms the next step in the simulation procedure.

41.4.2 Dynamics

The absorption coefficient is evaluated following the FGR procedure outlined in Section 41.2, using the states computed in Section 41.4.1. For energies below the host gap, assuming fully occupied valence bands and empty conduction bands, and neglecting the variation of states and energies with transverse momentum, the absorption coefficient of a square well (SW) potential of width L_w reads

$$\alpha_{\mathrm{SW}}(E_\gamma) = f \cdot E_\gamma^{-1} \sum_{i,j} |M_{c_i v_j}|^2 \Theta\left(E_\gamma - \varepsilon_{c_i v_j}\right), \tag{41.55}$$

with Θ the unit step function and

$$f = \frac{e^2 m_r^* P_{cv}^2}{m_0^2 \hbar n_r c_0 \varepsilon_0 L_w}, \tag{41.56}$$

where m_r^* is the reduced effective mass, $P_{cv} = \sqrt{E_P m_0 / 6}$ is the bulk MME for Kane energy $E_P = 26.9$ eV, $n_r = 3.7$ is the refractive index of the bulk material, and c_0 and ε_0 are speed of light and permittivity in vacuum, respectively. The overlap matrix elements are

$$M_{c_i v_j} = \int_L dz\, \psi_{c_i}^*(z) \psi_{v_j}(z), \tag{41.57}$$

where L ($> L_w$) is the normalization length of the wave functions, introduce the optical selection rules according to the WF symmetry. As can be verified in Figure 41.5, displaying the above FGR absorption coefficient (full line), only two of the six transitions between the three hole and two electron levels are not suppressed by symmetry. The validity of the selection rules in the absorption coefficient computed via NEGF (dotted line), while evident, is a nontrivial result, which requires careful consideration of the nonlocality in the Green's function picture of electron-photon interaction. In the absence of a symmetric potential, which is not found in the real operating device, the selection rules are strongly relaxed, and the absorption coefficient exhibits a shape that interpolates between 2D and 3D DOS (dashed line).

Based on this absorption coefficient, the emission coefficient $\mathcal{B}$ of the radiative recombination rate introduced in Section 41.2 is determined using the Van Roosbroeck–Shockley (VRS) relation (Equation 41.25). These two quantities are the main ingredients for the assessment of the device characteristics at the radiative limit.

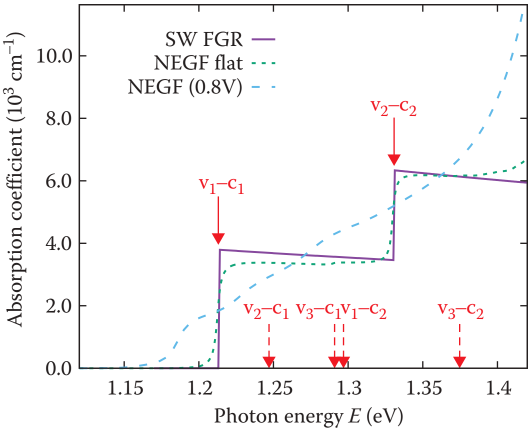


FIGURE 41.5 Absorption coefficient as computed within the Fermi's golden rule approach based on the solution of the single-band effective mass Schrödinger equation for the square well potential (SW FGR, full line), together with the NEGF absorption coefficients (spatially averaged over the QW region) for flat band (dotted) and the real potential at $V = 0.8$ V (dashed). While the selection rules are properly reflected in the NEGF approach, the real spectra feature subband tails and deviate from purely 2-D character due to transitions between quasibound states as a consequence of the large built-in fields present in the intrinsic region of the p-i-n diode.

41.4.3 Characteristics

First, the current–voltage characteristics are evaluated on the level of a detailed balance analysis, i.e., photocurrent and radiative dark current are computed from the absorptance as explained in Section 41.3. To this end, the absorptance is determined both from the LB law (a_{LB}) and from TMM (a_{TMM}) using the previously computed absorption coefficient α_{SW}. The photocurrent density can be evaluated at normal incidence. Dark current, on the other hand, should in principle be evaluated under consideration of the finite opening angle of the loss cone, as determined by the condition of total internal reflection. In the case of emission from high to low refractive index medium, which applies in the situation under consideration here ($n_r^{\mathrm{GaAs}} = 3.6$ to $n_r^{\mathrm{air}} = 1$), the angular dependence may be dropped, i.e., $a(\Omega, E_\gamma) \equiv a(\theta, E_\gamma) \approx a(E_\gamma)$, where the first equality is due to the planar geometry. The angular integration in Equation 41.39 thus provides just a factor of π, and the current–voltage characteristics can be written as

$$J(V) = J_{SC} - J_{0r}\left(e^{qV/(k_B T)} - 1\right), \tag{41.58}$$

with

$$J_{SC} = -\,q \int \mathrm{d}E_\gamma a(E_\gamma)\Phi_\gamma^{\swarrow}(E_\gamma), \tag{41.59}$$

$$J_{0r} = -\,q \int \mathrm{d}E_\gamma a(E_\gamma)\pi\Phi_{bb}(E_\gamma). \tag{41.60}$$

For simplicity, a monochromatic incident photon flux at $E_\gamma = 1.3$ eV and intensity of $I_\gamma \equiv \Phi_\gamma^{\swarrow}(E_\gamma)E_\gamma = 0.1$ kW/cm^2 is used for the evaluation of the characteristics. Figure 41.6 displays the JV curves obtained in this way using the absorptances a_{LB} and a_{TMM} based on α_{SW}, which shows that in this unoptimized situation, the reflector does not increase the absorptance of the QW significantly. Comparison with a 1D numerical DD-Poisson-TMM solver (*ASA*-TU Delft) using the same absorption coefficient α_{SW} and corresponding VRS emission coefficient, but otherwise the transport properties of bulk GaAs ($\mu_{n/p} = 400/40$ cm^2/Vs, $\mathcal{N}_{C/V} = 4.35/7.93 \times 10^{17}$ cm^{-3}, $\varepsilon_r = 13.18$, $S_{\mathrm{min/maj}} = 0/10^8$ m/s), reveals—as to be expected—coincidence of the photocurrent with that from a_{TMM}, but also a significant underestimation

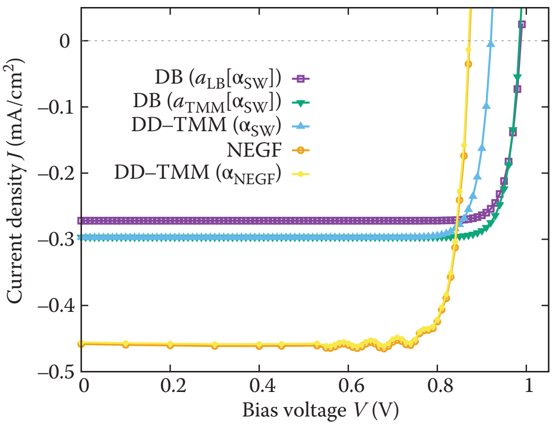


FIGURE 41.6 Current–voltage characteristics as provided by different simulation approaches: detailed balance (DB) currents based on the absorptance from Lambert–Beer's (LB) law or the TMM method, both using the square well potential (SW) absorption coefficient; the semiclassical drift-diffusion-Poisson-TMM model (DD-TMM) for the SW absorption coefficient; the fully microscopic NEGF formalism, revealing a bias-dependent structure in the level of photogeneration. The NEGF characteristics can be reproduced by using in the DD-TMM model the optical data as computed by NEGF, which hints at close to ideal transport in terms of carrier escape from the QW.

of the recombination due to the neglectance of the proper loss cone. The insensitivity to transport is a consequence of the high mobility of GaAs in combination with the short extraction distance and the absence of fast recombination, which results in the flat Fermi levels assumed in the global detailed balance approach. In the full NEGF simulation, on the other hand, both photocurrent and dark current are increased as compared to the characteristics based on the SW absorptance. Remarkably, the NEGF characteristics can be recovered almost completely in the DD-TMM model if local absorption and emission coefficients are extracted from NEGF. This hints at almost perfect transport in terms of unit carrier escape probability (implicitly assumed in the semiclassical model), in consistency with experimental observation for QW structures of similar depth and operated at room temperature. Deviations between the results of the two models are expected in the case of deep wells with presence of fast recombination processes competing with the slow carrier escape: While the NEGF picture considers that a fraction of the photogenerated carriers recombines in the QW prior to escape to the continuum, the semiclassical bulk model assumes the whole generation in bulk states, in which the carriers diffuse quickly from the highly absorbing and—due to detailed balance—emitting QW locations to regions with lower optical coupling strength. That regime, however, is currently not accessible due to the lack of a suitable NEGF picture of nonradiative recombination.

41.5 Summary

The use of NSs in photovoltaics is manifold, and many applications require dedicated models to assess their impact on the PV device performance. There is, however, a common rationale in the consideration of NS in PV device models, according to the device architecture and the NS functionality, which in general amounts to the application of a multiscale and multiphysics scheme. Hence, NS device simulation generally involves many different levels of modeling and simulation, from the determination of NS states to their consideration in the dynamics of carrier generation, transport, and recombination and the final synthesis in the propagation of NS properties to features in the global device characteristics. Reliable consideration of NS effects hence demands a derivation of the key PV material parameters (α, μ, τ) based on a consistent microscopic picture. In the case of ideal transport, an assessment in terms of the optical properties and detailed balance relations between absorption and emission gives already a good estimate of the device performance. For less ideal systems, the individual loss mechanisms need to be considered in more detail, as provided by semiclassical models on device level, down to the microscopic picture of the fundamental physical mechanisms as delivered by quantum-kinetic approaches in the case where nonclassical effects are relevant, such as transport away from diffusive and band-like regimes.

References

1. M. A. Green. Third generation photovoltaics: Ultra-high conversion efficiency at low cost. *Prog. Photovolt. Res. Appl.*, 9:123, 2001.
2. J. Müller, B. Rech, J. Springer, and M. Vanecek. TCO and light trapping in silicon thin film solar cells. *Sol. Energy*, 77(6):917–930, 2004.
3. S. Liu, D. Ding, S. R. Johnson, and Y.-H. Zhang. Approaching single-junction theoretical limit using ultra-thin GaAs solar cells with optimal optical designs. In *38th IEEE Photovoltaic Specialists Conference (PVSC)*, pages 002082–002087, 2012.
4. S. Pillai, K. R. Catchpole, T. Trupke, and M. A. Green. Surface plasmon enhanced silicon solar cells. *J. Appl. Phys.*, 101(9):093105, 2007.
5. I. Massiot, C. Colin, N. Péré-Laperne, P. Roca i Cabarrocas, C. Sauvan, P. Lalanne, J.-L. Pelouard, and S. Collin. Nanopatterned front contact for broadband absorption in ultra-thin amorphous silicon solar cells. *Appl. Phys. Lett.*, 101(16):163901, 2012.
6. J. R. Nagel and M. A. Scarpulla. Enhanced absorption in optically thin solar cells by scattering from embedded dielectric nanoparticles. *Opt. Express*, 18(S2):A139–A146, 2010.
7. C. Heine and R. H. Morf. Submicrometer gratings for solar energy applications. *Appl. Opt.*, 34(14):2476–2482, 1995.
8. P. Bermel, C. Luo, L. Zeng, L. C. Kimerling, and J. D. Joannopoulos. Improving thin-film crystalline silicon solar cell efficiencies with photonic crystals. *Opt. Express*, 15(25):16986–17000, 2007.
9. L. Cao, P. Fan, A. P. Vasudev, J. S. White, Z. Yu, W. Cai, J. A. Schuller, S. Fan, and M. L. Brongersma. Semiconductor nanowire optical antenna solar absorbers. *Nano Lett.*, 10(2):439–445, 2010.
10. K. W. J. Barnham, I. I. Ballard, J. G. Connolly, B. G. Kluftinger N. Ekins-Daukes, J. Nelson, C. Rohr, and M. Mazzer. Recent results on quantum well solar cells. *J. Mater. Sci. Mater. Electron.*, 11:531, 2000.
11. K. R. Catchpole and A. Polman. Plasmonic solar cells. *Opt. Express*, 16(26):21793–21800, 2008.
12. H. A. Atwater and A. Polman. Plasmonics for improved photovoltaic devices. *Nat. Mater.*, 9(3):205–213, 2010.
13. K. R. Catchpole, S. Mokkapati, F. Beck, E.-C. Wang, A. McKinley, A. Basch, and J. Lee. Plasmonics and nanophotonics for photovoltaics. *MRS Bull.*, 36:461–467, 2011.
14. S. Mokkapati and K. R. Catchpole. Nanophotonic light trapping in solar cells. *J. Appl. Phys.*, 112(10):101101, 2012.
15. K. W. J. Barnham and G. Duggan. A new approach to high-efficiency multi-band-gap solar cells. *J. Appl. Phys.*, 67:3490–3493, 1990.
16. D. B. Bushnell, K. W. J. Barnham, J. P. Connolly, M. Mazzer, N. J. Ekins-Daukes, J. S. Roberts, G. Hill, R. Airey, and L. Nasi. Effect of barrier composition and well number on the dark current of quantum well solar cells. In *Proceeding of 3rd World Conference on Photovoltaic Energy Conversion*, Osaka, Japan, page 2709, 2003.
17. L. C. Hirst, M. Fuhrer, D. J. Farrell, A. Le Bris, J.-F. Guillemoles, M. J. Y. Tayebjee, R. Clady, T. W. Schmidt, M. Sugiyama, Y. Wang, H. Fujii, and N. J. Ekins-Daukes. InGaAs/GaAsP quantum wells for hot carrier solar cells. *Proc. SPIE.*, 8256:82560X, 2012.
18. R. Oshima, A. Takata, and Y. Okada. Strain-compensated InAs/GaNAs quantum dots for use in high-efficiency solar cells. *Appl. Phys. Lett.*, 93(8):083111, 2008.
19. S. M. Hubbard, C. D. Cress, C. G. Bailey, R. P. Raffaelle, S. G. Bailey, and D. M. Wilt. Effect of strain compensation on quantum dot enhanced GaAs solar cells. *Appl. Phys. Lett.*, 92(12):123512, 2008.

20. Y. Okada, R. Oshima, and A. Takata. Characteristics of InAs/GaNAs strain-compensated quantum dot solar cell. *J. Appl. Phys.*, 106(2):024306, 2009.
21. C. G. Bailey, D. V. Forbes, R. P. Raffaelle, and S. M. Hubbard. Near 1 V open circuit voltage InAs/GaAs quantum dot solar cells. *Appl. Phys. Lett.*, 98(16):163105, 2011.
22. G. Conibeer, M. Green, E.-C. Cho, D. König, Y.-H Cho, T. Fangsuwannarak, G. Scardera, E. Pink, Y. Huang, T. Puzzer, S. Huang, D. Song, C. Flynn, S. Park, X. Hao, and D. Mansfield. Silicon quantum dot nanostructures for tandem photovoltaic cells. *Thin Solid Films*, 516(20):6748–6756, 2008.
23. J. G. J. Adams, B. C. Browne, I. M. Ballard, J. P. Connolly, N. L. A. Chan, A. Ioannides, W. Elder, P. N. Stavrinou, K. W. J. Barnham, and N. J. Ekins-Daukes. Recent results for single-junction and tandem quantum well solar cells. *Prog. Photovolt. Res. Appl.*, 19(7):865–877, 2011.
24. D. Alonso-Álvarez, A. G. Taboada, J. M. Ripalda, B. Alen, Y. González, L. González, J. M. García, F. Briones, A. Martí, A. Luque, A. M. Sánchez, and S. I. Molina. Carrier recombination effects in strain compensated quantum dot stacks embedded in solar cells. *Appl. Phys. Lett.*, 93(12):123114, 2008.
25. A. Martí, L. Cuadra, and A. Luque. Partial filling of a quantum dot intermediate band for solar cells. *IEEE Trans. Electron Dev.*, 48(10):2394–2399, 2001.
26. A. J. Nozik. Quantum dot solar cells. *Physica E*, 14(1–2):PII S1386–9477(02)00374–0, 2002.
27. R. Patterson, M. Kirkengen, B. Puthen Veettil, D. König, M. A. Green, and G. Conibeer. Phonon lifetimes in model quantum dot superlattice systems with applications to the hot carrier solar cell. *Sol. Energy Mater. Sol. Cells*, 94(11):1931–1935, 2010.
28. V. I. Klimov. Detailed-balance power conversion limits of nanocrystal-quantum-dot solar cells in the presence of carrier multiplication. *Appl. Phys. Lett.*, 89(12):123118, 2006.
29. O. E. Semonin, J. M. Luther, S. Choi, H.-Y. Chen, J. Gao, A. J. Nozik, and M. C. Beard. Peak external photocurrent quantum efficiency exceeding 100% via MEG in a quantum dot solar cell. *Science*, 334(6062):1530–1533, 2011.
30. A. Martí, E. Antolín, C. R. Stanley, C. D. Farmer, N. López, P. Diaz, E. Cánovas, P. G. Linares, and A. Luque. Production of photocurrent due to intermediate-to-conduction-band transitions: A demonstration of a key operating principle of the intermediate-band solar cell. *Phys. Rev. Lett.*, 97(24):247701, 2006.
31. P. G. Linares, A. Martí, E. Antolín, and A. Luque. III-V compound semiconductor screening for implementing quantum dot intermediate band solar cells. *J. Appl. Phys.*, 109(1):014313, 2011.
32. L. C. Hirst, R. J. Walters, M. F. Fuehrer, and N. J. Ekins-Daukes. Experimental demonstration of hot-carrier photo-current in an InGaAs quantum well solar cell. *Appl. Phys. Lett.*, 104(23):231115, 2014.
33. J. A. R. Dimmock, S. Day, M. Kauer, K. Smith, and J. Heffernan. Demonstration of a hot-carrier photovoltaic cell. *Prog. Photovolt. Res. Appl.*, 22(2):151–160, 2014.
34. F. W. Wise. Lead salt quantum dots: The limit of strong quantum confinement. *Acc. Chem. Res.*, 33(11):773–780, 2000.
35. A. J. Nozik, M. C. Beard, J. M. Luther, M. Law, R. J. Ellingson, and J. C. Johnson. Semiconductor quantum dots and quantum dot arrays and applications of multiple exciton generation to third-generation photovoltaic solar cells. *Chem. Rev.*, 110(11):6873–6890, 2010.
36. Y. Kanai, Z. Wu, and J. C. Grossman. Charge separation in nanoscale photovoltaic materials: Recent insights from first-principles electronic structure theory. *J. Mater. Chem.*, 20(6):1053, 2010.
37. A. Ruland, C. Schulz-Drost, V. Sgobba, and D. M. Guldi. Enhancing photocurrent efficiencies by resonance energy transfer in CdTe quantum dot multilayers: Towards rainbow solar cells. *Adv. Mater.*, 23(39):4573–7, 2011.
38. I. J. Kramer and E. H. Sargent. The architecture of colloidal quantum dot solar cells: Materials to devices. *Chem. Rev.*, 114(1):863–882, 2014.
39. R. J. Ellingson, M. C. Beard, J. C. Johnson, P. Yu, O. I. Micic, A. J. Nozik, A. Shabaev, and A. L. Efros. Highly efficient multiple exciton generation in colloidal PbSe and PbS quantum dots. *Nano Lett.*, 5(5):865–871, 2005.

40. D. König, Y. Takeda, and B. Puthen-Veettil. Technology-compatible hot carrier solar cell with energy selective hot carrier absorber and carrier-selective contacts. *Appl. Phys. Lett.*, 101(15), 2012.
41. D. Bozyigit, N. Yazdani, M. Yarema, O. Yarema, W. M. M. Lin, S. Volk, K. Vuttivorakulchai, M. Luisier, F. Juranyi, and V. Wood. Soft surfaces of nanomaterials enable strong phonon interactions. *Nature*, 531(7596):618–622, 2016.
42. W. Shockley and H. J. Queisser. Detailed balance limit of efficiency of p-n junction solar cells. *J. Appl. Phys.*, 32(3):510–519, 1961.
43. R. E. Chandler, A. J. Houtepen, J. Nelson, and D. Vanmaekelbergh. Electron transport in quantum dot solids: Monte Carlo simulations of the effects of shell filling, Coulomb repulsions, and site disorder. *Phys. Rev. B*, 75(8):085325, 2007.
44. P. Guyot-Sionnest. Electrical transport in colloidal quantum dot films. *J. Phys. Chem. Lett.*, 3(9):1169–1175, 2012.
45. D. Beljonne and J. Cornil, editors. *Multiscale Modelling of Organic and Hybrid Photovoltaics*. Berlin: Springer, Berlin, Heidelberg, 2014.
46. J. D Jackson. *Classical Electrodynamics*. Hoboken, NJ: John Wiley & Sons, 1998.
47. P. Yeh, A. Yariv, and C. S. Hong. Electromagnetic propagation in periodic stratified media.1. General theory. *J. Opt. Soc. Am.*, 67(4):423–438, 1977.
48. W. Schäfer and M. Wegener. *Semiconductor Optics and Transport Phenomena*. Berlin: Springer, 2002.
49. H. Haug and A. P. Jauho. *Quantum Kinetics in Transport and Optics of Semiconductors*. Berlin: Springer, 1996.
50. H. Haug and S. W. Koch. *Quantum Theory of the Optical and Electronic Properties of Semiconductors*. Singapore: World Scientific, 2004.
51. U. Aeberhard. Quantum-kinetic theory of photocurrent generation via direct and phonon-mediated optical transitions. *Phys. Rev. B*, 84(3):035454, 2011.
52. G. Allan and C. Delerue. Efficient intraband optical transitions in Si nanocrystals. *Phys. Rev. B*, 66(23):233303, 2002.
53. J.-W. Luo, P. Stradins, and A. Zunger. Matrix-embedded silicon quantum dots for photovoltaic applications: A theoretical study of critical factors. *Energy Environ. Sci.*, 4(7):2546, 2011.
54. D. A. B. Miller, D. S. Chemla, T. C. Damen, A. C. Gossard, W. Wiegmann, T. H. Wood, and C. A. Burrus. Electric field dependence of optical absorption near the band gap of quantum-well structures. *Phys. Rev. B*, 32:1043–1060, 1985.
55. S.-L. Chuang, S. Schmitt-Rink, D. A. B. Miller, and D. S. Chemla. Exciton Green's-function approach to optical absorption in a quantum well with an applied electric field. *Phys. Rev. B*, 43:1500–1509, 1991.
56. U. Aeberhard. Nonequilibrium Green's function theory of coherent excitonic effects in the photocurrent response of semiconductor nanostructures. *Phys. Rev. B*, 86:115317, 2012.
57. M. Paxman, J. Nelson, B. Braun, J. Connolly, K. W. J. Barnham, C. T. Foxon, and J. S. Roberts. Modeling the spectral response of the quantum well solar cell. *J. Appl. Phys.*, 74:614, 1993.
58. P. J. Stevens, M. Whitehead, G. Parry, and K. Woodbridge. Computer modeling of the electric field dependent absorption spectrum of multiple quantum well material. *IEEE J. Quant. Electron.*, 24:2007, 1988.
59. D. Atkinson, G. Parry, and E. J. Austin. Modelling of electroabsorption in coupled quantum wells with applications to voltage optical modulation. *Semicond. Sci. Technol.*, 5:516, 1990.
60. S. M. Ramey and R. Khoie. Modeling of multiple-quantum-well solar cells including capture, escape, and recombination of photoexcited carriers in quantum wells. *IEEE Trans. Electron Dev.*, 50:1179, 2003.
61. G. Lengyel, K. W. Jelley, and R. W. H. Engelmann. A semi-empirical model for electroabsorption in GaAs/AlGaAs multiple quantum well modulator structures. *IEEE J. Quant. Electron.*, 26(2):296–304, 1990.

62. S. Schmitt-Rink, D. A. B. Miller, and D. S. Chemla. Theory of the linear and nonlinear optical properties of semiconductor microcrystallites. *Phys. Rev. B*, 35:8113–8125, 1987.
63. D. L. Nika, E. P. Pokatilov, Q. Shao, and A. A. Balandin. Charge-carrier states and light absorption in ordered quantum dot superlattices. *Phys. Rev. B*, 76(12):125417, 2007.
64. S. Tomić, T. S. Jones, and N. M. Harrison. Absorption characteristics of a quantum dot array induced intermediate band: Implications for solar cell design. *Appl. Phys. Lett.*, 93(26):263105, 2008.
65. Y. Z. Hu, M. Lindberg, and S. W. Koch. Theory of optically excited intrinsic semiconductor quantum dots. *Phys. Rev. B*, 42(3):1713–1723, 1990.
66. W.-Y. Wu, J. N. Schulman, T. Y. Hsu, and U. Efron. Effect of size nonuniformity on the absorption spectrum of a semiconductor quantum dot system. *Appl. Phys. Lett.*, 51(10):710–712, 1987.
67. A. Vasanelli, R. Ferreira, and G. Bastard. Continuous absorption background and decoherence in quantum dots. *Phys. Rev. Lett.*, 89:216804, 2002.
68. A. W. Walker, O. Thériault, J. F. Wheeldon, and K. Hinzer. The effects of absorption and recombination on quantum dot multijunction solar cell efficiency. *IEEE J. Photovolt.*, 3(3):1118–1124, 2013.
69. R. Riölver, B. Berghoff, D. Bätzner, B. Spangenberg, H. Kurz, M. Schmidt, and B. Stegemann. Si/SiO2 multiple quantum wells for all silicon tandem cells: Conductivity and photocurrent measurements. *Thin Solid Films*, 516(20):6763–6766, 2008.
70. M. Zacharias, J. Heitmann, R. Scholz, U. Kahler, M. Schmidt, and J. Bläsing. Size-controlled highly luminescent silicon nanocrystals: A SiO/SiO2 superlattice approach. *Appl. Phys. Lett.*, 80(4):661–663, 2002.
71. J. Nelson, M. Paxman, K. W. J. Barnham, J. S. Roberts, and C. Button. Steady state carrier escape rates from single quantum wells. *IEEE J. Quant. Electron.*, 29:1460–1468, 1993.
72. O. Y. Raisky, W. B. Wang, R. R. Alfano, C. L. Reynolds Jr., and V. Swaminathan. Investigation of photoluminescence and photocurrent in InGaAsP/InP strained multiple quantum well heterostructures. *J. Appl. Phys.*, 81:339, 1997.
73. K. R. Lefebvre and A. F. M. Anwar. Electron and hole escape times in single quantum wells. *J. Appl. Phys.*, 80(6):3595–3597, 1996.
74. K. R. Lefebvre and A. F. M. Anwar. Electron escape time from single quantum wells. *IEEE J. Quant. Electron.*, 33(2):187–191, 1997.
75. H. Schneider and K. V. Klitzing. Thermionic emission and Gaussian transport of holes in a GaAs/AlxGa1$-x$As multiple-quantum-well structure. *Phys. Rev. B*, 38(9):6160, 1988.
76. A. Larsson, P. A. Andrekson, S. T. Eng, and A. Yariv. Tunable superlattice p-i-n photodetectors: characteristics, theory, and application. *IEEE J. Quant. Electron.*, 24(5):787, 1988.
77. D. J. Moss, T. Ido, and H. Sano. Calculation of photogenerated carrier escape rates from GaAs/AlxGa1$-x$As quantum wells. *IEEE J. Quant. Electron.*, 30(4):1015, 1994.
78. C.-Y. Tsai, L. F. Eastman, Y.-H. Lo, and C.-Y. Tsai. Breakdown of thermionic emission theory for quantum wells. *Appl. Phys. Lett.*, 65(4):469–471, 1994.
79. U. Aeberhard. Microscopic theory and numerical simulation of quantum well solar cells. *Proc. SPIE*, 7597:759702, 2010.
80. P. N. Brounkov, A. A. Suvorova, M. V. Maximov, A. F. Tsatsul'nikov, A. E. Zhukov, A. Yu Egorov, A. R. Kovsh, S. G. Konnikov, T. Ihn, S. T. Stoddart, L. Eaves, and P. C. Main. Electron escape from self-assembled InAs/GaAs quantum dot stacks. *Physica B*, 249–251:267–270, 1998.
81. C. M. A. Kapteyn, F. Heinrichsdorff, O. Stier, R. Heitz, M. Grundmann, N. D. Zakharov, D. Bimberg, and P. Werner. Electron escape from InAs quantum dots. *Phys. Rev. B*, 60:14265–14268, 1999.
82. P. W. Fry, J. J. Finley, L. R. Wilson, A. Lemaitre, D. J. Mowbray, M. S. Skolnick, M. Hopkinson, G. Hill, and J. C. Clark. Electric-field-dependent carrier capture and escape in self-assembled InAs/GaAs quantum dots. *Appl. Phys. Lett.*, 77(26):4344–4346, 2000.
83. S. Marcinkevius and R. Leon. Carrier capture and escape in InxGa1$-x$As/GaAs quantum dots: Effects of intermixing. *Phys. Rev. B*, 59:4630–4633, 1999.

84. M. De Giorgi, C. Lingk, G. von Plessen, J. Feldmann, S. De Rinaldis, A. Passaseo, M. De Vittorio, R. Cingolani, and M. Lomascolo. Capture and thermal re-emission of carriers in long-wavelength InGaAs/GaAs quantum dots. *Appl. Phys. Lett.*, 79(24):3968–3970, 2001.
85. J. A. Brum and G. Bastard. Resonant carrier capture by semiconductor quantum wells. *Phys. Rev. B*, 33:1420–1423, 1986.
86. P. W. M. Blom, J. E. M. Haverkort, and J. H. Wolter. Optimization of barrier thickness for efficient carrier capture in graded-index and separate-confinement multiple quantum well lasers. *Appl. Phys. Lett.*, 58(24):2767–2769, 1991.
87. S. C. Kan, D. Vassilovski, T. C. Wu, and K. Y. Lau. Quantum capture and escape in quantum-well lasers—Implications on direct modulation bandwidth limitations. *IEEE Photonic. Technol. Lett.*, 4(5):428–431, 1992.
88. P. Sotirelis, P. von Allmen, and K. Hess. Electron intersubband relaxation in doped quantum wells. *Phys. Rev. B*, 47:12744–12753, 1993.
89. P. Sotirelis and K. Hess. Electron capture in GaAs quantum wells. *Phys. Rev. B*, 49:7543–7547, 1994.
90. B. K. Ridley. Space-charge-mediated capture of electrons and holes in a quantum well. *Phys. Rev. B*, 50:1717–1724, 1994.
91. C.-Y. Tsai, L. F. Eastman, Y.-H. Lo, and C.-Y. Tsai. Carrier capture and escape in multisubband quantum well lasers. *IEEE Photonic. Technol. Lett.*, 6(9):1088–1090, 1994.
92. L. Thibaudeau and B. Vinter. Phonon-assisted carrier capture into a quantum well in an electric field. *Appl. Phys. Lett.*, 65(16):2039–2041, 1994.
93. D. Bradt, Y. M. Sirenko, and V. Mitin. Inelastic and elastic mechanisms of electron capture to a quantum well. *Semicond. Sci. Technol.*, 10(3):260, 1995.
94. K. Kalna and M. Mosko. Electron capture in quantum wells via scattering by electrons, holes, and optical phonons. *Phys. Rev. B*, 54:17730–17737, 1996.
95. M. Mosko and K. Kalna. Carrier capture into a GaAs quantum well with a separate confinement region: Comment on quantum and classical aspects. *Semicond. Sci. Technol.*, 14(9):790, 1999.
96. J. A. Brum and G. Bastard. Direct and indirect carrier capture by semiconductor quantum wells. *Superlattice. Microst.*, 3(1):51–55, 1987.
97. P. W. M. Blom, C. Smit, J. E. M. Haverkort, and J. H. Wolter. Carrier capture into a semiconductor quantum well. *Phys. Rev. B*, 47(4):2072–2081, 1993.
98. C.-Y. Tsai, C.-Y. Tsai, Y.-H. Lo, and L. F. Eastman. Carrier dc and ac capture and escape times in quantum-well lasers. *IEEE Photonic. Technol. Lett.*, 7(6):599–601, 1995.
99. C.-Y. Tsai, C.-Y. Tsai, Y.-H. Lo, R. M. Spencer, and L. F. Eastman. Nonlinear gain coefficients in semiconductor quantum-well lasers: Effects of carrier diffusion, capture, and escape. *IEEE J. Sel. Top. Quant. Electron.*, 1(2):316–330, 1995.
100. Y. Lam and J. Singh. Monte Carlo studies on the well-width dependence of carrier capture time in graded-index separate confinement heterostructure quantum well laser structures. *Appl. Phys. Lett.*, 63(14):1874–1876, 1993.
101. L. Davis, Y. L. Lam, Y. C. Chen, J. Singh, and P. K. Bhattacharya. Carrier capture and relaxation in narrow quantum wells. *IEEE J. Quant. Electron.*, 30(11):2560–2564, 1994.
102. L. F. Register and K. Hess. Simulation of carrier capture in quantum well lasers due to strong inelastic scattering. *Superlattice. Microst.*, 18(3):223–228, 1995.
103. G. C. Crow and R. A. Abram. Monte Carlo simulations of capture into quantum well structures. *Semicond. Sci. Technol.*, 14(1):1, 1999.
104. M. Ryzhii and V. Ryzhii. Monte Carlo modeling of electron transport and capture processes in AlGaAs/GaAs multiple quantum well infrared photodetectors. *Jpn. J. Appl. Phys.*, 38(10R):5922, 1999.
105. J. Hader, J. V. Moloney, and S. W. Koch. Structural dependence of carrier capture time in semiconductor quantum-well lasers. *Appl. Phys. Lett.*, 85(3):369–371, 2004.

106. A. V. Uskov, J. McInerney, F. Adler, H. Schweizer, and M. H. Pilkuhn. Auger carrier capture kinetics in self-assembled quantum dot structures. *Appl. Phys. Lett.*, 72(1):58–60, 1998.
107. I. Magnusdottir, S. Bischoff, A. V. Uskov, and J. Mørk. Geometry dependence of Auger carrier capture rates into cone-shaped self-assembled quantum dots. *Phys. Rev. B*, 67:205326, 2003.
108. A. V. Uskov, I. Magnusdottir, B. Tromborg, J. Mørk, and R. Lang. Line broadening caused by Coulomb carrier-carrier correlations and dynamics of carrier capture and emission in quantum dots. *Appl. Phys. Lett.*, 79(11):1679–1681, 2001.
109. R. Ferreira and G. Bastard. Phonon-assisted capture and intradot Auger relaxation in quantum dots. *Appl. Phys. Lett.*, 74(19):2818–2820, 1999.
110. J. M. Miloszewski, M. S. Wartak, S. G. Wallace, and S. Fafard. Theoretical investigation of carrier capture and escape processes in cylindrical quantum dots. *J. Appl. Phys.*, 114(15):154311, 2013.
111. J.-Z. Zhang and I. Galbraith. Rapid hot-electron capture in self-assembled quantum dots via phonon processes. *Appl. Phys. Lett.*, 89(15):153119, 2006.
112. I. Magnusdottir, A. V. Uskov, R. Ferreira, G. Bastard, J. Mørk, and B. Tromborg. Influence of quasibound states on the carrier capture in quantum dots. *Appl. Phys. Lett.*, 81(23):4318–4320, 2002.
113. T. R. Nielsen, P. Gartner, and F. Jahnke. Many-body theory of carrier capture and relaxation in semiconductor quantum-dot lasers. *Phys. Rev. B*, 69:235314, 2004.
114. J. Seebeck, T. R. Nielsen, P. Gartner, and F. Jahnke. Polarons in semiconductor quantum dots and their role in the quantum kinetics of carrier relaxation. *Phys. Rev. B*, 71:125327, 2005.
115. G. D. Mahan. *Many-Particle Physics*, 2nd ed. New York, NY: Plenum, 1990.
116. T. Gunst, T. Markussen, K. Stokbro, and M. Brandbyge. First-principles method for electron-phonon coupling and electron mobility: Applications to two-dimensional materials. *Phys. Rev. B*, 93:035414, 2016.
117. C.-W. Jiang and M. A. Green. Silicon quantum dot superlattices: Modeling of energy bands, densities of states, and mobilities for silicon tandem solar cell applications. *J. Appl. Phys.*, 99(11):114902, 2006.
118. A. Wacker. Semiconductor superlattices: A model system for nonlinear transport. *Phys. Rep.*, 357:1–111, 2002.
119. P. Würfel. *Physics of Solar Cells: From Basic Principles to Advanced Concepts*. Weinheim: Wiley-VCH, 2009.
120. I. Mora-Sero, L. Bertoluzzi, V. Gonzalez-Pedro, S. Gimenez, F. Fabregat-Santiago, K. W. Kemp, E. H. Sargent, and J. Bisquert. Selective contacts drive charge extraction in quantum dot solids via asymmetry in carrier transfer kinetics. *Nat. Commun.*, 4:2272, 2013.
121. W. Van Roosbroeck and W. Shockley. Photon-radiative recombination of electrons and holes in germanium. *Phys. Rev.*, 94:1558, 1954.
122. E. O. Göbel, H. Jung, J. Kuhl, and K. Ploog. Recombination enhancement due to carrier localization in quantum well structures. *Phys. Rev. Lett.*, 51(17):1588–1591, 1983.
123. A. R. Beattie and P. T. Landsberg. Auger effect in semiconductors. *Proc. R. Soc. London Ser. A*, 249:16, 1959.
124. P. T. Landsberg. *Recombination in Semiconductors.* Cambridge: Cambridge University Press, 1991.
125. L. Chiu and A. Yariv. Auger recombination in quantum-well InGaAsP heterostructure lasers. *IEEE J. Quant. Electron.*, 18(10):1406–1409, 1982.
126. C. Smith, R. A. Abram, and M. G. Burt. Auger recombination in a quantum well heterostructure. *J. Phys. C Solid State Phys.*, 16(5):L171, 1983.
127. N. K. Dutta. Calculation of Auger rates in a quantum well structure and its application to InGaAsP quantum well lasers. *J. Appl. Phys.*, 54(3):1236–1245, 1983.
128. A. Sugimura. Auger recombination effect on threshold current of InGaAsP quantum well lasers. *IEEE J. Quant. Electron.*, 19(6):932–941, 1983.
129. P. K. Basu. Auger recombination rate in quantum well lasers: Modification by electronelectron interaction in quasi two dimensions. *J. Appl. Phys.*, 56(11):3344–3346, 1984.

130. A. Haug. Phonon-assisted auger recombination in quantum well semiconductors. *Appl. Phys. A*, 51:354, 1990.
131. M. Takeshima. Phonon-assisted Auger recombination in a quasi-two-dimensional structure semiconductor. *Phys. Rev. B*, 30(6):3302–3308, 1984.
132. L.-W. Wang, M. Califano, A. Zunger, and A. Franceschetti. Pseudopotential theory of auger processes in CdSe quantum dots. *Phys. Rev. Lett.*, 91:056404, 2003.
133. G. A. Narvaez, G. Bester, and A. Zunger. Carrier relaxation mechanisms in self-assembled (In,Ga)As/GaAs quantum dots: Efficient $P \rightarrow S$ Auger relaxation of electrons. *Phys. Rev. B*, 74:075403, 2006.
134. S. Tomić. Intermediate-band solar cells: Influence of band formation on dynamical processes in InAs/GaAs quantum dot arrays. *Phys. Rev. B*, 82(19):195321, 2010.
135. K. Hyeon-Deuk and O. V. Prezhdo. Time-domain ab initio study of Auger and phonon-assisted auger processes in a semiconductor quantum dot. *Nano Lett.*, 11(4):1845–1850, 2011.
136. N. J. Ekins-Daukes, K. W. J. Barnham, J. P. Connolly, J. S. Roberts, J. C. Clark, G. Hill, and M. Mazzer. Strain-balanced GaAsP/InGaAs quantum well solar cells. *Appl. Phys. Lett.*, 75(26):4195–4197, 1999.
137. K. Ding, U. Aeberhard, O. Astakhov, U. Breuer, M. Beigmohamadi, S. Suckow, B. Berghoff, W. Beyer, F. Finger, R. Carius, and U. Rau. Defect passivation by hydrogen reincorporation for silicon quantum dots in SiC/SiO$x$hetero-superlattice. *J. Non-Cryst. Solids*, 358(17):2145–2149, 2012.
138. W. Shockley and W. T. Read. Statistics of the recombinations of holes and electrons. *Phys. Rev.*, 87:835–842, 1952.
139. R. N. Hall. Electron-hole recombination in germanium. *Phys. Rev.*, 87:387, 1952.
140. A. Schenk. An improved approach to the Shockley-Read-Hall recombination in inhomogeneous fields of space-charge regions. *J. Appl. Phys.*, 71(7):3339–3349, 1992.
141. A. Alkauskas, Q. Yan, and C. G. Van de Walle. First-principles theory of nonradiative carrier capture via multiphonon emission. *Phys. Rev. B*, 90:075202, 2014.
142. U. Aeberhard. Simulation of nanostructure-based high-efficiency solar cells: Challenges, existing approaches, and future directions. *IEEE J. Sel. Top. Quant. Electron*, 19(5):4000411, 2013.
143. G. L. Araújo and A. Martí. Absolute limiting efficiencies for photovoltaic energy conversion. *Sol. Energy Mater. Sol. Cells*, 33:213, 1994.
144. U. Rau. Reciprocity relation between photovoltaic quantum efficiency and electroluminescent emission of solar cells. *Phys. Rev. B*, 76(8):085303, 2007.
145. M. Born and E. Wolf. *Principles of Optics*, 7th ed. Cambridge: Cambridge University Press, 1999.
146. P. Würfel. The chemical potential of radiation. *J. Phys. C Solid State Phys.*, 15:3967, 1982.
147. U. Aeberhard. Simulation of ultra-thin solar cells beyond the limits of the semi-classical bulk picture. *IEEE J. Photovolt.*, 6(3):654–660, 2016.
148. U. Aeberhard. Impact of built-in fields and contact configuration on the characteristics of ultra-thin GaAs solar cells. *Appl. Phys. Lett.*, 109(3), 2016.
149. N. G. Anderson. Ideal theory of quantum well solar cells. *J. Appl. Phys.*, 78:1861, 1995.
150. G. L. Araújo, A. Martí, F. W. Ragay, and J. H. Wolter. Efficiency of multiple quantum well solar cells. In *Proceeding of 12th European Photovoltaic Solar Energy Conference*, Amsterdam, The Netherlands, page 1481, 1994.
151. G. L. Araújo and A. Martí. Electroluminescence coupling in multiple quantum well diodes and solar cells. *Appl. Phys. Lett.*, 66(7):894–895, 1995.
152. S. P. Bremner, R. Corkish, and C. B. Honsberg. Detailed balance efficiency limits with quasi-Fermi level variations. *IEEE Trans. Electron Dev.*, 46(10):1932–1939, 1999.
153. A. Luque, A. Martí, and L Cuadra. Thermodynamic consistency of sub-bandgap absorbing solar cell proposals. *IEEE Trans. Electron Dev.*, 48:2118, 2001.
154. J. P. Connolly, J. Nelson, K. W. J. Barnham, I. Ballard, C. Roberts, J. S. Roberts, and C. T. Foxon. Simulating multiple quantum well solar cells. In *Proceeding of 28th IEEE Photovoltaic Specialists Conference*, Anchorage, Alaska, USA, page 1304, 2000.

155. J. C. Rimada and L. Hernández. Modelling of ideal AlGaAs quantum well solar cells. *Microelectron. J.*, 32(9):719–723, 2001.
156. S. J. Lade and A. Zahedi. A revised ideal model for AlGaAs/GaAs quantum well solar cells. *Microelectron. J.*, 35(5):401–410, 2004.
157. J. C. Rimada, L. Hernandez, J. P. Connolly, and K. W. J. Barnham. Quantum and conversion efficiency calculation of AlGaAs/GaAs multiple quantum well solar cells. *Phys. Status Solidi B*, 242(9):1842–1845, 2005.
158. J. C. Rimada, L. Hernández, J. P. Connolly, and K. W J Barnham. Conversion efficiency enhancement of AlGaAs quantum well solar cells. *Microelectron. J.*, 38(4–5):513–518, 2007.
159. M. Courel, J. C. Rimada, and L. Hernández. GaAs/GaInNAs quantum well and superlattice solar cell. *Appl. Phys. Lett.*, 100(7):073508, 2012.
160. C. I. Cabrera, J. C. Rimada, M. Courel, L. Hernandez, J. P. Connolly, A. Enciso, and D. A. Contreras-Solorio. Modeling multiple quantum well and superlattice solar cells. *Nat. Resour.*, 4(03):235–245, 2013.
161. A. Chatterjee, A. Kumar Biswas, and A. Sinha. An analytical study of the various current components of an AlGaAs/GaAs multiple quantum well solar cell. *Physica E*, 72:128–133, 2015.
162. M. Courel, J. C. Rimada, and L. Hernandez. An approach to high efficiencies using GaAs/GaInNAs multiple quantum well and superlattice solar cell. *J. Appl. Phys.*, 112(5):054511, 2012.
163. A. Alemu and A. Freundlich. Resonant thermotunneling design for high-performance single-junction quantum-well solar cells. *IEEE J. Photovolt.*, 2(3):256–260, 2012.
164. M. Courel, J. C. Rimada, and L. Hernández. AlGaAs/GaAs superlattice solar cells. *Prog. Photovolt. Res. Appl.*, 21(3):276–282, 2013.
165. A. Luque and A. Martí. Increasing the efficiency of ideal solar cells by photon induced transitions at intermediate levels. *Phys. Rev. Lett.*, 78(26):5014–5017, 1997.
166. Z. Wang, T. P. White, and K. R. Catchpole. Plasmonic near-field enhancement for planar ultra-thin photovoltaics. *IEEE Photon. J.*, 5(5):8400608, 2013.
167. W. Hu, M. Maksudur Rahman, M.-Y. Lee, Y. Li, and S. Samukawa. Simulation study of type-II Ge/Si quantum dot for solar cell applications. *J. Appl. Phys.*, 114(12):124509, 2013.
168. Q. Shao, A. A. Balandin, A. I. Fedoseyev, and M. Turowski. Intermediate-band solar cells based on quantum dot supracrystals. *Appl. Phys. Lett.*, 91(16):163503, 2007.
169. Q. Deng, X. Wang, H. Xiao, C. Wang, H. Yin, H. Chen, Q. Hou, D. Lin, J. Li, Z. Wang, and X. Hou. An investigation on $In_xGa_{1\text{-}x}N$/GaN multiple quantum well solar cells. *J. Phys. D Appl. Phys.*, 44(26):265103, 2011.
170. W. Wei, Q. Zhang, S. Zhao, and Y. Zhang. Two intermediate bands solar cells of InGaN/InN quantum dot supracrystals. *Appl. Phys. A*, 116(3):1009–1016, 2014.
171. A. Mellor, A. Luque, I. Tobias, and A. Marti. Realistic detailed balance study of the quantum efficiency of quantum dot solar cells. *Adv. Funct. Mater.*, 24(3):339–345, 2014.
172. G. Wei, K.-T. Shiu, N. C. Giebink, and S. R. Forrest. Thermodynamic limits of quantum photovoltaic cell efficiency. *Appl. Phys. Lett.*, 91(22), 2007.
173. J. Nelson, I. Ballard, K. Barnham, J. P. Connolly, J. S. Roberts, and M. Pate. Effect of quantum well location on single quantum well p-i-n photodiode dark currents. *J. Appl. Phys.*, 86:5898, 1999.
174. P. Kailuweit, R. Kellenbenz, S. P. Philipps, W. Guter, A. W. Bett, and F. Dimroth. Numerical simulation and modeling of GaAs quantum-well solar cells. *J. Appl. Phys.*, 107(6):064317, 2010.
175. T. Kirchartz, K. Seino, J.-M. Wagner, U. Rau, and F. Bechstedt. Efficiency limits of Si/SiO2 quantum well solar cells from first-principles calculations. *J. Appl. Phys.*, 105(10):104511, 2009.
176. C.-Y. Tsai and C.-Y. Tsai. Effects of carrier escape and capture processes on quantum well solar cells: A theoretical investigation. *IET Optoelectron.*, 3(6):300–304, 2009.
177. O. Kengradomying, S. Jiang, Q. Wang, N. Vogiatzis, and J. M. Rorison. Modelling escape and capture processes in GaInNAs quantum well solar cells. *Phys. Status Solidi C Curr. Top. Solid State Phys.*, 10(4):585–588, 2013.

178. J. Nelson, K. W. J Barnham, J. P Connolly, G. Haarpaintner, C. Button, and J. Roberts. Quantum well solar cell dark currents. In *Proceeding of 12th EU PVSEC*, Amsterdam, The Netherlands, 1994.
179. M. Gioannini, A. P. Cedola, N. Di Santo, F. Bertazzi, and F. Cappelluti. Simulation of quantum dot solar cells including carrier intersubband dynamics and transport. *IEEE J. Photovolt.*, 3(4):1271–1278, 2013.
180. K. Driscoll, M. F. Bennett, S. J. Polly, D. V. Forbes, and S. M. Hubbard. Effect of quantum dot position and background doping on the performance of quantum dot enhanced GaAs solar cells. *Appl. Phys. Lett.*, 104(2):023119, 2014.
181. K. Yoshida, Y. Okada, and N. Sano. Self-consistent simulation of intermediate band solar cells: Effect of occupation rates on device characteristics. *Appl. Phys. Lett.*, 97(13):133503, 2010.
182. Y.-X. Gu, X.-G. Yang, H.-M. Ji, P.-F. Xu, and T. Yang. Theoretical study of the effects of InAs/GaAs quantum dot layer's position in i-region on current-voltage characteristic in intermediate band solar cells. *Appl. Phys. Lett.*, 101(8):081118, 2012.
183. T. Sogabe, T. Kaizu, Y. Okada, and S. Tomić. Theoretical analysis of GaAs/AlGaAs quantum dots in quantum wire array for intermediate band solar cell. *J. Renew. Sustain. Ener.*, 6(1):011206, 2014.
184. V. Aroutiounian, S. Petrosyan, A. Khachatryan, and K. Touryan. Quantum dot solar cells. *J. Appl. Phys.*, 89(4):2268, 2001.
185. A. Martí, L. Cuadra, and A. Luque. Quasi-drift diffusion model for the quantum dot intermediate band solar cell. *IEEE Trans. Electron Dev.*, 49(9):1632–1639, 2002.
186. C.-C. Lin, M.-H. Tan, C.-P. Tsai, K.-Y. Chuang, and T. S. Lay. Numerical study of quantum-dot-embedded solar cells. *IEEE J. Sel. Top. Quant. Electron.*, 19(5), 2013.
187. V. Aroutiounian, S. Petrosyan, and A. Khachatryan. Studies of the photocurrent in quantum dot solar cells by the application of a new theoretical model. *Sol. Energy Mater. Sol. Cells*, 89(2–3):165–173, 2005.
188. S. M. Willis, J. A. R. Dimmock, F. Tutu, H. Y. Liu, M. G. Peinado, H. E. Assender, A. A. R. Watt, and I. R. Sellers. Defect mediated extraction in InAs/GaAs quantum dot solar cells. *Sol. Energy Mater. Sol. Cells*, 102:142–147, 2012.
189. U. Aeberhard. Theory and simulation of quantum photovoltaic devices based on the non-equilibrium Green's function formalism. *J. Comput. Electron.*, 10:394–413, 2011.
190. D. A. Stewart and F. Leonard. Photocurrents in nanotube junctions. *Phys. Rev. Lett.*, 93:107401, 2004.
191. D. A. Stewart and F. Leonard. Energy conversion efficiency in nanotube optoelectronics. *Nano Lett.*, 5:219, 2005.
192. J. Guo, M. A. Alam, and Y. Yoon. Theoretical investigation on photoconductivity of single intrinsic carbon nanotubes. *Appl. Phys. Lett.*, 88(13):133111, 2006.
193. A. Buin, A. Verma, and S. Saini. Optoelectronic response calculations in the framework of k.p coupled to non-equilibrium Green's functions for 1D systems in the ballistic limit. *J. Appl. Phys.*, 114(3):033111, 2013.
194. U. Aeberhard and R. H. Morf. Microscopic nonequilibrium theory of quantum well solar cells. *Phys. Rev. B*, 77:125343, 2008.
195. U. Aeberhard. Spectral properties of photogenerated carriers in quantum well solar cells. *Sol. Energy Mater. Sol. Cells*, 94(11):1897–1902, 2010.
196. N. Cavassilas, F. Michelini, and M. Bescond. Theoretical comparison of multiple quantum wells and thick-layer designs in InGaN/GaN solar cells. *Appl. Phys. Lett.*, 105(6), 2014.
197. U. Aeberhard. Theory and simulation of photogeneration and transport in Si-SiOx superlattice absorbers. *Nanoscale Res. Lett.*, 6:242, 2011.
198. U. Aeberhard. Simulation of nanostructure-based and ultra-thin film solar cell devices beyond the classical picture. *J. Photon. Energy*, 4(1):042099, 2014.
199. U. Aeberhard. Effective microscopic theory of quantum dot superlattice solar cells. *Opt. Quant. Electron.*, 44:133–140, 2012.
200. A. Berbezier and U. Aeberhard. Impact of nanostructure configuration on the photovoltaic performance of quantum-dot arrays. *Phys. Rev. Applied*, 4:044008, 2015

42

Nanowire Solar Cells: Electro-Optical Performance

42.1 Introduction 475
42.2 Efficiency of Nanostructured Solar Cells 476
Solar Cell Current Density • Thermal Equilibrium • Nonequilibrium • Efficiency Evaluation
42.3 Computational Models 481
Optical Models • Electronic Models
42.4 Light Capture and Emission 483
Absorptivity • Modal Analysis • Shockley–Queisser Efficiency of a Nanowire Array • Purcell Effect
42.5 Electronic Properties 489
Radial versus Axial Junction Arrangement • Surface Recombination
42.6 Summary 492

Bernd Witzigmann

42.1 Introduction

Nanowire arrays show unique optical and electronic features for photovoltaic applications. It has been demonstrated that they exhibit intrinsic optical antireflection properties [1,2]. Moreover, they also can act as optical microconcentrators [3] that focus light to an active region. Nanowires have a large aspect ratio, improving the electronic properties by combining a long optical absorption path along the wire axis and a short lateral carrier extraction [4]. Short carrier extraction paths are beneficial for the carrier collection efficiency. These mechanisms are illustrated in Figure 42.1. On the left, a conventional bulk solar cell collects light in the vertical direction, and the photogenerated carriers move to the respective contacts causing the photo current. A nanostructured solar cell (Figure 42.1a) can act as an optical concentrator in order to focus the incoming light to a small area. Moreover, the carrier extraction path length to the contacts is reduced, which minimizes the probability of carrier loss by unwanted recombination.

Finally, multijunction cells can be realized with strain-relaxed axial heterointerfaces. This allows optimum bandgap combinations [5] and low-defect densities at the same time. As a potential drawback, the large surface to volume ratio can lead to enhanced surface recombination. Epitaxial passivation techniques have been developed in order to suppress surface recombination [6].

In this chapter, a detailed simulation study of the electro-optical characteristics of nanowire array solar cells is presented. An efficiency analysis is derived that treats solar cells with nanostructures on a broader basis using the Shockley–Queisser framework [7]. As addition, the modification of optical density of states (DOS) and wave optics effects are included. Using this framework, the numerical simulation of nanostructured solar cells is discussed. The simulation model consists of the three-dimensional solution of the vectorial Helmholtz equation for solar illumination and a drift-diffusion/Poisson equation system for

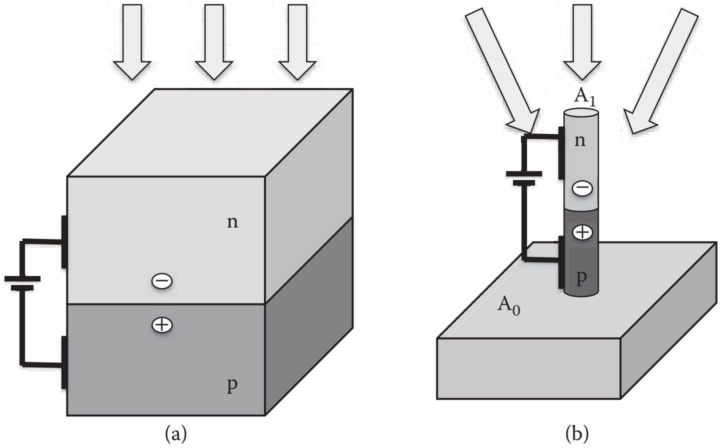


FIGURE 42.1 (a) Bulk solar cell with p-n junction. (b) Solar cell structured as a nanowire.

carrier extraction. In the results section, it is first shown how a nanowire array can act as an electromagnetic broadband absorber. Second, the choice of material composition and dopant placement is discussed with respect to efficiency in single-junction designs. Surface recombination is discussed for both radial and axial p-n junction arrangement.

42.2 Efficiency of Nanostructured Solar Cells

In this section, the efficiency of solar cells that contain nanostructures is discussed. Nanostructures can modify the optical or the electronic properties. As a basic theory, the established derivation by Shockley and Queisser [7] is used, which assumes detailed balance in the radiative limit, i.e., nonradiative recombination processes are not included. The main extension of Shockley and Queisser's classical model is to include the fact that light is viewed as electromagnetic waves and their modes in resonant structures. This leads to a modal dispersion and a strong local variation of the electromagnetic power. Moreover, resonant structures modify the optical DOS, and those enter the absorption as well as the emission properties of a cell, which is based on the Purcell effect [8].

The focus will be put on semiconductor materials, although the derivation can be transferred to other materials in a straightforward manner.

42.2.1 Solar Cell Current Density

The electronic current density of the solar cell can be easily obtained from current continuity in the steady state. It is derived from Maxwell's equations and builds on the mass conservation principle of charged particles:

$$\nabla \vec{J} = q(G - R), \tag{42.1}$$

with $\vec{J}$ being the local current density, G and R representing generation and recombination rates, respectively, and q being the elementary charge. In the integral form, the continuity equation reads

$$\int \vec{J} \mathrm{d}\vec{A} = q \int (G - R)\, \mathrm{d}V. \tag{42.2}$$

The surface area integration on the left hand side is carried out over the contacts (where carriers can leave or enter the solar cell), and the right hand side integration includes the solar cell volume.

42.2.2 Thermal Equilibrium

In thermal equilibrium, all of the emission and absorption processes of the solar cell can be equated, as there is no net current at the contacts. The left hand side of Equation 42.2 is zero, therefore

$$\int G\mathrm{d}V = \int R\mathrm{d}V = \int G^0\mathrm{d}V = \int R^0\mathrm{d}V. \tag{42.3}$$

The local generation rate per volume is

$$G^0 = \int \alpha \mathrm{d}j_{\mathrm{phot}}, \tag{42.4}$$

with α being the absorption coefficient (a material property) and j_{phot} is the photon current density (in units $1/\mathrm{m}^2\mathrm{s}$), which in thermal equilibrium is basically a blackbody radiation at the given reservoir temperature T_0. In order to relate the photon current density to the blackbody radiation (which is an energy density per spectral width), the following relation is used:

$$\mathrm{d}j_{\mathrm{phot}} = \mathrm{d}E_{\mathrm{phot}}\frac{c}{n}\frac{1}{\hbar\omega} \tag{42.5}$$

and

$$dE_{\mathrm{phot}} = D_{\mathrm{opt}}f_{\mathrm{phot}}\hbar\omega\mathrm{d}\hbar\omega. \tag{42.6}$$

The optical DOS D_{opt} depends on the dielectric surrounding of the solar cell and f_{phot} is the photon distribution function. Within an electromagnetic theory, D_{opt} can be calculated as the electromagnetic response of a local dipole excitation [9]:

$$D_{\mathrm{opt}}(\omega, \vec{x}) = \frac{1}{2\pi}\sum_{m}\int_{\vec{k}} \mathrm{d}\vec{k}\delta(\omega - \omega_m)|\vec{e}_{\mathrm{d}}\vec{E}_{m,\vec{k}}(\vec{r})|^2, \tag{42.7}$$

where $\vec{E}$ is the electric field as obtained from the solution of Maxwell's equations, $\vec{e}_{\mathrm{d}}$ is the dipole unit source, and ω_m the eigenfrequencies of the system.

The photon distribution function in equilibrium is

$$f_{\mathrm{phot}} = \frac{1}{\exp\left(\frac{\hbar\omega}{k_B T_0}\right) - 1}. \tag{42.8}$$

The generation rate, therefore, can be expressed as

$$G^0(\vec{x}) = R^0(\vec{x}) = \int \alpha(\vec{x}, \hbar\omega)\frac{c}{n_r}D^{\mathrm{sc}}_{\mathrm{opt}}(\vec{x}, \hbar\omega)f_{\mathrm{phot}}(T_0)d\hbar\omega, \tag{42.9}$$

which is equal to the recombination in equilibrium.

42.2.3 Nonequilibrium

In the nonequilibrium case, the solar cell is illuminated by solar radiation power, which in our case is represented by an external blackbody source at the temperature T_{s}. This leads to an energy density outside of the solar cell of

$$\mathrm{d}E^{\mathrm{s}}_{\mathrm{phot}} = D^{\mathrm{vac}}_{\mathrm{opt}}f_{\mathrm{phot}}(T_{\mathrm{s}})\hbar\omega d\hbar\omega. \tag{42.10}$$

The optical DOS here is the vacuum DOS, and the solar temperature T_s enters the Bose–Einstein photon distribution function. In principle, other illumination sources can be used, which then enter the calculation, e.g., as tabulated energy density. The external photon source generates additional electron–hole pairs ΔG inside the solar cell, and therefore

$$G^\mathrm{s} = \int \alpha a \frac{c}{n_r^\mathrm{vac}} D_\mathrm{opt}^\mathrm{vac} f_\mathrm{phot}(T_\mathrm{s}) \mathrm{d}\hbar\omega. \tag{42.11}$$

Here, a is the relative absorptivity, the amount of electromagnetic solar power entering the solar cell. It can be calculated using electromagnetic theory by

$$a(\vec{x}) = \frac{\langle \int Re\left\{\nabla\vec{S}\right\}\mathrm{d}V\rangle}{\int Re\left\{\vec{S}\right\}\mathrm{d}\vec{A}}, \tag{42.12}$$

where $\vec{S}$ is the Poynting vector, and the integration in the denominator is performed over a closed surface around the solar cell. This is especially important for nonplanar devices where the solar energy is captured not only from the surface, but from an angle of incidence larger than the half space π. The $\langle\rangle$ symbol in the numerator denotes an averaging over a volume that contains several optical wavelengths so that an inclusion into a drift-diffusion carrier transport framework is possible. The absorptivity $a(\vec{x})$ is the local relative electromagnetic energy entering in the solar cell, and in the ray optics approximation often is expressed as $(1 - r)\exp(-\alpha x)$, where the first factor is the back-reflected power from the air–top surface interface, and the second factor is the diminishing optical intensity inside the device due to internal absorption (Beer–Lambert law).

In the nonequilibrium case, the generation in the cell is increased, which in turn leads to an increased radiative recombination rate. Assuming that the radiative recombination is proportional to the product of electron and hole density, one gets

$$R^\mathrm{s} = R^0\frac{np}{n_\mathrm{i}^2} = \frac{N_cN_v}{n_\mathrm{i}^2}\exp\left(\frac{\eta_c\eta_v}{k_\mathrm{B}T_0}\right)G^0 = G^0\exp\left(\frac{\Delta E_\mathrm{F}}{k_\mathrm{B}T_0}\right), \tag{42.13}$$

with the following nomenclature: The intrinsic density is n_i, the effective state densities are N_c or N_v, the reduced Fermi energies are $\eta_c = E_\mathrm{F}^n - E^c$ and $\eta_v = -E_\mathrm{F}^p + E^v$, and ΔE_F is the quasi-Fermi level separation.

With these quantities, the current densities of the solar cell can be calculated. Using the current continuity equation, one gets

$$\int \vec{j}\mathrm{d}\vec{A} = q\int(G^s - R^s)\mathrm{d}V = q\int\left(G^s - G^0\exp(\frac{\Delta E_\mathrm{F}}{k_\mathrm{B}T_0})\right)\mathrm{d}V = I_\mathrm{gen} - I_\mathrm{rec}. \tag{42.14}$$

The quasi-Fermi level separation can be approximated by the voltage at the contacts, and after integration of the left hand side, the current of the solar cell becomes

$$I = I_\mathrm{gen} - I_\mathrm{sat}\exp\left(\frac{qV}{k_\mathrm{B}T_0}\right). \tag{42.15}$$

The currents introduced here are the total generation and recombination currents I_gen and I_rec, and the reverse saturation current I_sat. They can be evaluated from

$$I_\mathrm{sat} = q\int\int \alpha\frac{c}{n_r^\mathrm{sc}} D_\mathrm{opt}^\mathrm{sc} f_\mathrm{phot}(T_0)\,\mathrm{d}\hbar\omega\mathrm{d}V \tag{42.16}$$

and

$$I_{\mathrm{gen}} = q \int\int \alpha a \frac{c}{n_r^{\mathrm{vac}}} D_{\mathrm{opt}}^{\mathrm{vac}} f_{\mathrm{phot}}(T_{\mathrm{s}}) \mathrm{d}\hbar\omega \mathrm{d}V. \tag{42.17}$$

The short-circuit current of a solar cell is the current under illumination with zero voltage at the contacts, and therefore accounts to

$$I_{\mathrm{sc}} = I_{\mathrm{gen}} - I_{\mathrm{sat}}. \tag{42.18}$$

With this definition, the current–voltage relationship can be written in the more conventional Shockley form by replacing I_{gen} in Equation 42.15:

$$I = I_{\mathrm{sat}} \left(1 - \exp(\frac{qV}{k_{\mathrm{B}} T_0})\right) + I_{\mathrm{sc}}. \tag{42.19}$$

The open-circuit voltage is the voltage where the net current flow at the contacts is zero, and it can be evaluated from the relations obtained here as

$$V_{\mathrm{oc}} = \frac{k_{\mathrm{B}} T_0}{q} \log\left(\frac{I_{\mathrm{gen}}}{I_{\mathrm{sat}}}\right) = \frac{k_{\mathrm{B}} T_0}{q} \log\left(1 + \frac{I_{\mathrm{sc}}}{I_{\mathrm{sat}}}\right). \tag{42.20}$$

42.2.4 Efficiency Evaluation

From the current versus voltage characteristics, the electro-optical efficiency of a solar cell can be obtained from

$$\eta_{\mathrm{sc}} = \max\left(\frac{IV}{P_{\mathrm{inc}}}\right), \tag{42.21}$$

where the denominator is the total, spectrally integrated incident solar power. Obtaining a high efficiency translates into maximizing the open-circuit voltage V_{oc} and the short-circuit current I_{sc} as defined in Section 42.2.3, which maximizes the product IV. A maximum short-circuit current is achieved by capturing the incident broadband solar radiation and absorbing the photons in the solar cell. According to Equation 42.17, this requires a large absorption coefficient α, which is a material property, and a good absorptivity a, which is the ability to capture light into the solar cell, an electromagnetic property. In bulk solar cells, the absorptivity a can be maximized by broadband antireflection coatings. If the solar cell thickness is in the order of a few optical wavelengths, waveguiding or even evanescent wave effects can help maximize a. It will be shown in Section 42.2.4.2 how this is applied to nanowire solar cells.

The evaluation of the open-circuit voltage involves a maximum optical generation current I_{gen} and a minimum reverse saturation current I_{sat}, which is a challenge to fulfill. As can be seen from Equation 42.16, the major ingredient to I_{sat} is the optical DOS of the solar cell.

42.2.4.1 Bulk Solar Cell

First, the efficiency evaluation is shown for a bulk solar cell. In principle, all current commercial technologies fall into this category, also the thin-film designs. For a homogeneous medium, the optical DOS is

$$D_{\mathrm{opt}}^{\mathrm{hom}} = \left(\frac{n_r}{c}\right)^3 \frac{1}{4\pi\hbar^3} (\hbar\omega)^2 \Omega, \tag{42.22}$$

where Ω is the solid angle of the radiation emission and absorption, respectively.

For the bulk solar cell, the solid angle of illumination is $\Omega = \pi$, and therefore

$$D_{\text{opt}}^{\text{vac}} = \left(\frac{1}{c}\right)^3 \frac{1}{4\hbar^3}(\hbar\omega)^2. \tag{42.23}$$

The solid angle of emission at any location inside the solar cell is $\Omega = 4\pi$, leading to

$$D_{\text{opt}}^{\text{sc}} = \left(\frac{n_r}{c}\right)^3 \frac{1}{\hbar^3}(\hbar\omega)^2. \tag{42.24}$$

For the absorptivity a, a simple ray optics model can be applied, which leads to

$$a = (1 - r)(1 - \exp(-\alpha d)) \approx 1, \tag{42.25}$$

with r being the reflectivity at the air-semiconductor interface, and d the cell thickness. By design, one can realize a good antireflective coating and sufficient cell thickness, leading to the relation $a \approx 1$. The short-circuit current can therefore be maximized by engineering the cell.

The open-circuit voltage contains the ratio of $\frac{I_{\text{gen}}}{I_{\text{sat}}}$, which is approximately proportional to the ratio $\frac{D_{\text{opt}}^{\text{vac}} n_r^{\text{sc}}}{D_{\text{opt}}^{\text{sc}} n_r^{\text{vac}}}$. This results in a factor $\frac{1}{4n_r^2}$ in the short-circuit current calculation. The physics behind this factor is the light intensity concentration. The factor of 4 in the denominator can be reduced by efficient light management, as it comes from the photon collection from the angle π but photon emission into an angle of 4π [10]. Introducing a dielectric environment that allows emission into an angle π only (by, e.g., photonic crystals or conventional mirrors) increases the open-circuit voltage. Therefore, light management is an important physical mechanism for advanced solar cell design [10,11].

42.2.4.2 Optical Microconcentrators

Waveguides, resonators, or optical antennas can influence the optical absorption behavior of a solar cell. As first case, an optical concentrator that acts as absorber is taken that focuses the incoming solar radiation to the absorber area A_1, whereas the illuminated (large) area is A_0 (for illustration, see Figure 42.1). This leads to an absorptivity $a \propto \frac{A_0}{A_1} = f$, as the absorber concentrates the energy flux density to the area A_1. As a consequence, the short-circuit current is obtained with a much smaller volume of absorption compared to a nonconcentrated case (see integral in Equation 42.18). At the same time, the reverse saturation current (Equation 42.16) is also integrated over a smaller volume compared to a bulk solar cell. The ratio of $\frac{I_{\text{gen}}}{I_{\text{sat}}}$, and therefore the open circuit voltage, increases. However, two assumptions need to be made: First, the reverse saturation current density needs to be independent of the local carrier density, which is, within the Shockley theory, a valid assumption. Second, the optical concentrator modifies the optical DOS of the absorber. This needs to be taken into account in the open-circuit voltage calculation. Therefore, the open-circuit voltage is only increased if the following relation holds:

$$\frac{D_{\text{opt}}^{\text{vac}}}{D_{\text{opt}}^{\text{mc}}} f > \frac{D_{\text{opt}}^{\text{vac}}}{D_{\text{opt}}^{\text{bulk}}}, \tag{42.26}$$

or

$$\frac{D_{\text{opt}}^{\text{bulk}}}{D_{\text{opt}}^{\text{mc}}} f > 1, \tag{42.27}$$

where $D_{\text{opt}}^{\text{bulk}}$ is the optical DOS for the bulk cell and $D_{\text{opt}}^{\text{mc}}$ is the optical DOS for the micro concentrated cell, respectively; their ratio can be interpreted as the Purcell factor. This result shows an interesting design opportunity for high-efficiency nanostructured solar cells: If optical elements are introduced that focus the solar electromagnetic energy to a small volume, the open-circuit voltage, and therefore the efficiency, can be increased compared to a bulk or thin-film solar cell. However, the Purcell factor, or the ratio of optical DOS, should not be increased at the same rate, as it compensates the concentration effect.

42.2.4.3 Electronic Nanostructures

In the field of lasers and light-emitting diodes, electronic nanostructures such as quantum wells and quantum dots have increased the performance considerably [12]. In solar cells, the application of electronic nanostructures has not been used as a standard feature to improve the cell efficiency. In general, the introduction of nanostructures alters the electronic band structure. It leads to the following consequences:

- The effective bandgap, i.e., the absorption edge, is shifted to the lowest subband energy difference.
- The absorption coefficient increases, as the electronic DOS is higher at the band edge compared to a bulk material, and in addition, the overlap of the envelope wave functions enters the absorption coefficient.
- The absorption coefficient becomes anisotropic due to the anisotropy of the underlying carrier wave functions and DOS.

As a drawback, electronic nanostructures typically have dimensions around 10 nm, which is much smaller than the optical wavelength or a typical absorption path length. Therefore, in order to obtain efficient absorption, multiple nanometer-sized regions need to be stacked in the direction of the light path or efficient focusing of the light onto the absorbing nanostructure needs to be done. Fundamentally, the anisotropy of the absorption and emission can be used to restrict the absorptivity α to the solid angles of the solar radiation, and to reduce the reverse saturation current by a maximum theoretical factor of 4 (see explanation in Section 42.2.4.1) [13]. In addition, the increase of the absorption coefficient leads to smaller active regions, which increases the ratio of short-circuit current to reverse saturation current as outlined in Section 42.2.4.2.

Finally, some advanced concepts have been pursued with electronic nanostructures such as intermediate band solar cells [14] or hot-carrier contacts [15]. The basic idea is to introduce additional energy levels with nanostructures in order to overcome the single bandgap limitation of the semiconductor. Such structures require nonequilibrium carrier transport models and are not discussed here.

42.3 Computational Models

In the previous sections, some fundamental properties of nanostructured solar cells have been described within the semianalytical framework of the Shockley–Queisser model. In general, the incoming light generates electron–hole pairs by absorption in the semiconductor. In Section 42.2, it has been shown that the efficiency of a solar cell depends on the solar optical generation $G^s(\vec{x})$, the generation in equilibrium $G^0(\vec{x})$, and the recombination under illumination $R^s(\vec{x})$ (see Equation 42.14). In principle, those are all local quantities. In nanostructured solar cells, both light intensity and carrier density can be controlled on the scale of tens of nanometers by optical waveguides, plasmonic subwavelength structures, or quantum wells or dots. This introduces a strong local variation of the physical quantities contributing to the cell efficiency. Computational models can capture the local variations, and a correct determination of the efficiency is feasible.

42.3.1 Optical Models

The governing optical equations for solar cells with structures similar or smaller than the optical wavelength are Maxwell's equations. The solar illumination is modeled as a broadband current source, and the resulting electromagnetic field contributes to the local carrier generation rate. There is a multitude of models to solve Maxwell's equations either analytically, semianalytically, or numerically.

A popular analytical solution is the plane wave ansatz, which after some simplifications leads to the transfer matrix method (TMM). Using TMM, light propagation in multilayer structures can be evaluated or the even simpler Beer–Lambert law can be obtained:

$$I(x) = I_0 e^{-\alpha x}, \tag{42.28}$$

with the intensity I.

Most semianalytical methods expand the solution of Maxwell's equations in modes. For nanowire solar cells, which are essentially cylinders, the modes of a cylindrical optical fiber can be used, and the solution is then represented as superposition of the fiber modes [16].

Numerical methods essentially discretize the simulation domain and solve the respective equations only on parts of the domain, e.g., vertices or edges. The finite-difference time-domain (FDTD) method [17] is a popular method that discretizes space and time using explicit finite differences. With sources placed in the simulation domain, the propagation of electromagnetic waves can be computed iteratively in space and time. The FDTD method is straightforward to implement and can be adapted for parallel computing in a natural manner. The original FDTD kernels [17] require rectangular tensor product meshes, which can be restrictive for complex geometries. Moreover, the solutions are only second-order accurate in space and time, which requires many grid points per wavelength for accurate solutions, in particular for optical frequencies. A rather elegant approach to overcome these limitations is the discontinuous Galerkin time-domain method [18]. It combines explicit time stepping with a local finite-element space discretization, and therefore powerful high-order accuracy and flexible meshes can be used. The local elements are connected with constraints for the numerical flux in order to ensure consistency and convergence to the exact solution. For solar cell simulations, which use the broadband solar spectrum as source, the time-domain methods allow the use of spectral broadband sources, with only one simulation run for the entire spectrum. Dispersive materials need special treatment in the time domain, and therefore additional computational resources.

A second class of numerical methods solve Maxwell's equations in the frequency domain. Here, monochromatic harmonic waves replace the time derivatives with an imaginary frequency factor, and any multifrequency solution can be built as superposition in the linear case. Maxwell's equations can be reduced to the Helmholtz equation, which for the electric field $\vec{E}$ reads

$$\left(\Delta - \omega^2 \mu \epsilon\right) \vec{E} = -j\mu\vec{J}, \tag{42.29}$$

where μ and ϵ are the respective permeability and permittivity, ω is the frequency, and $\vec{J}$ is a current source. This equation is valid for homogeneous isotropic materials, and the permittivity is in general a dispersive complex quantity in optics. This equation can be solved by a finite-element method in three dimensions, with sophisticated meshes, and an approximation to the exact solution. In its homogeneous form (without current sources), it yields an eigenvalue problem that give the eigenfrequencies and eigenvectors of a given geometry. For complex permittivities, the eigenvalue problem becomes complex as well, and the imaginary part of the eigenfrequency gives the quality factor of the eigenmode [19,20].

In order to evaluate the efficiency of a solar cell beyond the analytical Shockley–Queisser limit, the framework presented in Section 42.2 as efficiency analysis for nanostructured solar cells requires evaluation of the classical electromagnetic field for the solar cell geometry in the presence of a complex refractive index. This enters both the absorptivity a (see Equation 42.12) and the optical DOS D_{opt} (Equation 42.7).

42.3.2 Electronic Models

Any electronic model describes the dynamics of the charge carriers inside the solar cell, i.e., the photogenerated carriers and the intrinsic carriers under applied bias. In a generalized description, solar cell operation means that the photogenerated carriers thermally relax to the band edges and are subsequently driven to the contacts creating the photocurrent. In the radiative (i.e., Shockley–Queisser) limit, the carriers can recombine radiatively on their way to the contacts, and of course, the resulting photons can be reabsorbed. This gives an upper limit to the photocurrent. However, carriers can be lost by nonradiative recombination, which decreases the efficiency further. An excellent description of such an efficiency analysis can be found in [21]. In order to understand local features of the carrier transport, such as surface recombination effects, or defect recombination related to specific doping species or materials, a space-resolved carrier transport model is needed. In nanostructures, additional physical properties are potentially relevant, e.g., the thermal relaxation process is different from bulk processes, which can lead to ballistic or semiballistic carrier transport. If these heated carriers can be detected by a contact without thermal relaxation, a higher efficiency is achieved as the thermalization energy loss can be avoided [22].

Therefore, electronic properties of nanostructures require simulations by quantum transport models in order to include effects such as carrier tunneling, phase coherence (and carrier localization), and ballistic transport. These models have been successfully applied to nanowire devices by several groups [23] and established methods are the nonequilibrium Green's functions (NEGF) method or the Wigner function formalism [24]. The NEGF method has been applied to quantum solar cells [25,26]; however, these methods in general are computationally demanding, as they discretize both real and phase space. Therefore, simulation domains of only a few nanometers can be accounted for. In the future, the so called multiscale approaches might be feasible that combine computationally intense quantum transport methods with efficient semiclassical methods; several approaches have been made, with still many open questions [27].

In the following sections, the carrier transport in nanowire solar cells is analyzed. It will turn out that these solar cells have dimensions with diameters larger than 50 nm and lengths more than 1000 nm. In this regime, semiclassical models still have a decent validity, therefore the established and efficient drift-diffusion model will be used. It comprises the self-consistent solution of the current continuity equations for electrons and holes with the Poisson equation for the electrostatic potential.

42.4 Light Capture and Emission

Figure 42.2 shows the arrangement of the nanowire array for a solar cell. The wire diameter is d, the lattice constant is a, and the wire height is L. The first question is how such an array can incouple and absorb light. Solar radiation spans from the ultraviolet to the infrared wavelength range, and the solar cell therefore needs to be a broadband absorber. It has been reported in the literature that an array of nanowires can act as a good broadband absorber [16,28]. For highest efficiency with a single bandgap material, the classical Shockley–Queisser analysis predicts a band gap of the semiconductor material between $E_g \approx 1.3$ eV and 1.4 eV [11].This bandgap optimum can be explained by the balance of having minimum carrier thermalization losses from high-energy photons on the one hand and minimizing the losses due

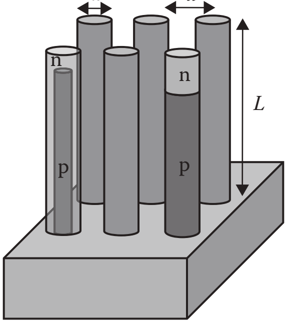


FIGURE 42.2 Illustration of an ordered nanowire array solar cell. The p-n junction can be arranged in a radial (left) or axial (right) manner. The periodicity can be hexagonal or cubic.

to the optical transparency of the semiconductor for low-energy photons. Indium phosphide (InP) with $E_g = 1.34eV$ matches this range very well; therefore, the following analysis will use this material. The findings are applicable to any other material as well.

42.4.1 Absorptivity

Using a three-dimensional finite-element solution of the complex vectorial Maxwell equations, the wire arrangement as shown in Figure 42.2 is illuminated from the top side with a linearly polarized sheet source. The source matches the power spectrum of AM1.5D solar power and is placed at a distance above the wires, which is beyond the evanescent mode decay length. The space between the wires is vacuum, and both the wire and the substrate material is InP, with the complex refractive indices following the dispersion relation given in [12]. The simulation is carried out for a quarter of the wire only, with periodic boundary conditions on the cut planes in order to represent an infinite periodicity of the array . Perfectly matched layer absorbing boundary conditions are placed at the top and the bottom of the simulation domain.

From the electromagnetic fields, the relative local absorptivity $a(\vec{x}, \lambda)$ is calculated according to Equation 42.12, which becomes, after integration of the wire volume, the spectral absorptivity of the entire wire $\tilde{a}(\lambda)$. The optical absorption in the substrate is not included in the integration and does not contribute to the efficiency of the cell. Figure 42.3 shows the spectral absorptivity for different wire periodicity and diameters; the wire height is constant at $L = 2$ μm. A geometric fill factor is defined that describes the volume of active material in a nanostructured cell relative to a bulk material volume. This fill factor can be expressed as

$$f_{nw} = \frac{\pi d^2}{4a^2}. \tag{42.30}$$

The highest fill factor can be achieved with $d = a$ and a fill factor of $f_{nw} \approx 0.79$, i.e., when the nanowires touch. For the simulations shown in Figure 42.3, combinations of periodicity and diameter have been

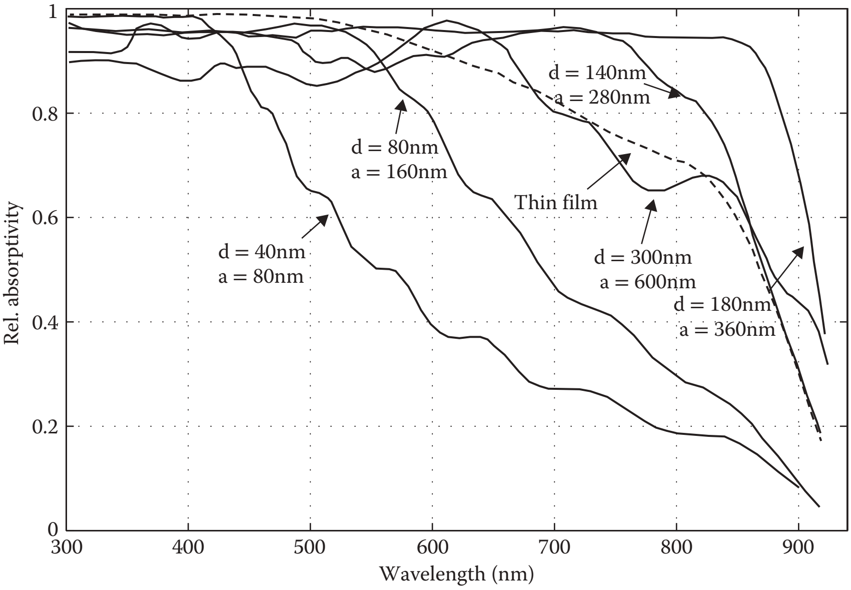


FIGURE 42.3 Spectral relative absorptivity of an InP nanowire array. The wire diameter is labeled *d*, and the wire distance is *a*.

chosen so that the material volume fill factor is identical at $f_{nw} = 0.196$. The arrangement with the smallest wire diameter of $d = 40$ nm and distance of $a = 80$ nm shows a substantial reduction of absorptivity above a wavelength of 500 nm. Nanowires with larger diameter and larger array pitch show a better long wavelength absorptivity, and the case of $d = 180$ nm and $a = 360$ nm has an absorptivity of more than 90% up to the bandgap wavelength of InP. The dashed curve shown in Figure 42.3 is the absorptivity of a homogeneous InP film with a thickness of $2\mu m \times 0.196 = 392$ nm, which is the same material volume as the nanowire array. This thin film, however, is covered by an ideal antireflective coating in order to eliminate reflections (which is not added to the wire array case). The decrease of absorptivity at wavelengths larger than 600 nm comes from the insufficient thickness of the InP layer.

42.4.2 Modal Analysis

In order to understand the findings of Section 42.4.1, the spectral behavior of the absorptivity of the nanowire array can be analyzed by a modal description of the electromagnetic field distribution. In [16], it has been shown that the absorptivity obtained from the three-dimensional calculation can be constructed from a separation of the power absorptivity parallel to the nanowire axis according to the Beer–Lambert law. The effective attenuation coefficient is obtained from the two-dimensional mode calculation of a cross-section of the wire array perpendicular to the wire axis. This assumes an infinite extension of the wires along their axes, and the calculation gives the eigenmodes with their complex eigenvalues $n_m^{\text{eff}} = n'_m + in''_m$, with m as the mode index. The real part of the eigenvalue describes the coupling efficiency of the electromagnetic incident TEM wave to the wire array and is inserted in the separation ansatz for the calculation of the reflected power. The imaginary part is the effective absorptivity that enters the Beer–Lambert expression. For an illustration, see Figure 42.4. The optical power at the interface between vacuum and the nanowire array top can be expressed as

$$P_m(z=0) = P_0 \frac{4n_0 n'_m}{(n_0 + n'_m)^2} \frac{\Gamma_m}{\sum_i \Gamma_i}, \tag{42.31}$$

with n_0 being the vacuum refractive index, P_0 the source power, and Γ_m is the overlap integral of the electric field of the incoming TEM mode with the mth mode of the wire array cross-section. The absorbed power in the nanowire material is then

$$P_{\text{abs}}(\lambda) = \sum_i P_i(z=0)\left(1 - \exp(\frac{4\pi}{\lambda} n''_i L)\right). \tag{42.32}$$

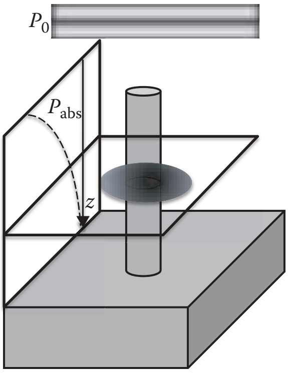


FIGURE 42.4 Illustration of the electromagnetic separation approach for a nanowire array. The incoming plane wave power P_0 is separated into two-dimensional cross-sectional modes and an exponential power decay in z-direction following a Beer–Lambert law.

Using this separation ansatz, the spectral absorptivity curves of the three-dimensional calculation can be reproduced with excellent accuracy, as shown in [28].

This procedure identifies the type of modes that couple the linearly polarized free-space electromagnetic field to the nanowire array. The unit cell of the array consists of the wire itself (which resembles a cylindrical fiber) and the periodic boundary to the neighboring wires that can be viewed as the Neumann boundary condition at the border of the unit cell.

The broadband absorptivity for the nanowire array can be described by only very few modes. For the case of $d = 180$ nm and $a = 360$ nm, only two modes are necessary to create the absorptivity characteristic as shown in Figure 42.3. These modes are plotted in Figure 42.5 with their electric field mode profile and their relative incoupling efficiency. The latter describes the ability of the wire array to act as an antireflective structure.The effect of relative incoupling in Figure 42.5 (bottom figure) is a combination of the confinement factor Γ and the real part of the effective index of the in-plane mode in the nanowire array n' (see Equation 42.31). For incoupling of 90%, the effective index therefore is close to 1 (i.e., the surrounding air), which for a fiber mode, is the regime of weak confinement. Still, most of the electromagnetic power couples into the wire array, and the field distribution as shown in in Figure 42.5 (e.g., mode A at 500 nm) shows a strong field between the wires, which is also absorbed according to the Beer–Lambert law. Between 300 nm and 700 nm, mode A couples close to 90% into the wire, and the electric field analysis shows that the electric field distribution resembles an fiber mode. Finally, the electric field distribution between the wires resembles a fundamental TEM mode, meaning parallel arrangement of the electric field vector. At 750 nm, mode A reaches its cutoff wavelength, and most of the light couples into mode B, which is of type HE_{11}. For further detail, the reader is referred to the optical fiber standard dispersion relations and eigenmodes, which are similar to the case of the sparse wire array.

It is surprising that only two modes are responsible for the broadband absorptivity of the electromagnetic power. This modal description helps understand the characteristics: If the wire spacing is increased from the ideal distance, the overlap integral of the wire modes decreases, and therefore the absorptivity decreases. If the wire spacing is decreased, the effective index of the wire array n' increases, and therefore the relative incoupling decreases. If, at the same distance, the wire diameter is modified, similar effects occur; an increase of the wire diameter moves the fiber mode dispersion relation to the multimode regime, where the effective indices are larger (a standard optical fiber dispersion relation illustrates this). A decrease of the wire diameter will decrease the mode number to a point where only the fundamental mode can be used.

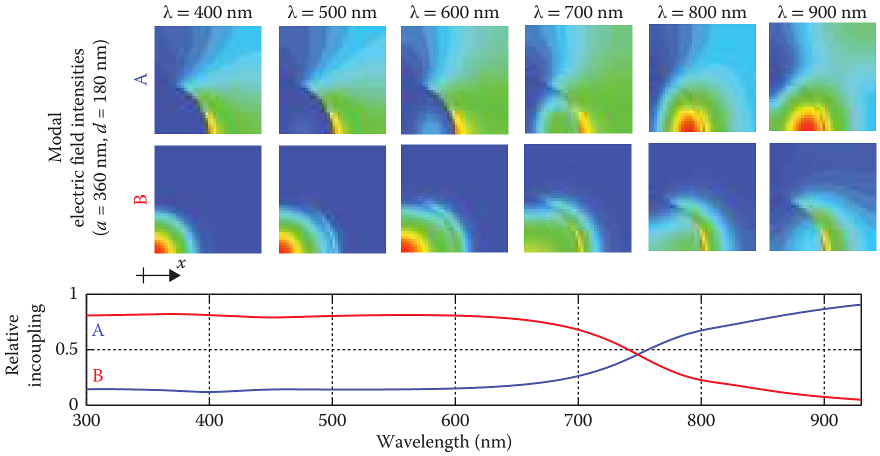


FIGURE 42.5 Top: Electric field distribution in the nanowire array for mode A and mode B, for different wavelengths. Bottom: Relative incoupling P_A/P_0 and P_B/P_0 (see Equation 42.31). (Reprinted from J. Kupec et al., *Optics Express*, 18, 27589–27605, 2010. With permission.)

As a rule of thumb, therefore, the wire distance and diameters should be chosen so that in the dispersion relation of the cross-sectional eigenvalue problem, the bandwidth of operation can span the fundamental and a few higher order modes. Moreover, the wire length should be chosen so that the imaginary part of the eigenvalue ensures full absorption of the optical power according to the Beer–Lambert law.

The modal analysis of the wire array can also be applied for solar radiation under an incident angle different from vertical illumination. For moderate incident angles (below 15°), the results do not change significantly.

42.4.3 Shockley–Queisser Efficiency of a Nanowire Array

Having calculated the relative absorptivity from electromagnetics, the Shockley–Queisser efficiency can subsequently be evaluated from Equation 42.21. Carrier transport effects are not included at this level of description and will be described later. Also, as approximation, the optical DOS in the semiconductor $D_{\mathrm{opt}}^{\mathrm{sc}}$ is taken to be a bulk DOS, for simplicity. This neglects the Purcell effect for the radiative emission, and therefore we get an upper limit for the solar cell efficiency. The impact of the modified optical DOS will be discussed in Section 42.4.4.

Figure 42.6 shows the calculated efficiency of the InP nanowire array with a square wire arrangement and a wire length of $L = 2\ \mu\mathrm{m}$. The efficiency is plotted versus the volumetric fill factor. For small wire diameters of below 100 nm, the efficiency is highest for high fill factors and decreases for larger wire distances. Here, mostly the fundamental HE_{11} mode is responsible for the absorptivity; however, the overlap with the wire is too small for efficient absorption. For larger wire distance, the electromagnetic energy therefore is lost in the substrate. For thicker wires with diameters larger than 100 nm, the most dense array with highest fill factor shows the same efficiency as the thinner wires; however, with increasing wire distance, the efficiency increases up to values of 30% for the diameter of $d = 180$ nm. This increase can be explained by a reduced effective index of the wire array with lower fill factor, and therefore less back reflection from the vacuum–wire array interface. The electromagnetic energy is concentrated to the wires by the modal light capture, and the decrease of efficiency below fill factors of 10% is due to the decreasing coupling factor. As a result, a maximum efficiency of 30% can be reached with a fill factor of approximately 20%, a wire diameter of $d = 180$ nm, and a wire separation of $a = 360$ nm, without an additional anti-reflective coating. Hence, the wire array acts as a microconcentrator element as well as an active absorber at the same time. An InP bulk cell with ideal broadband antireflective coating gives a detailed balance efficiency of $\eta_{\mathrm{ideal}} = 31\%$ for comparison.

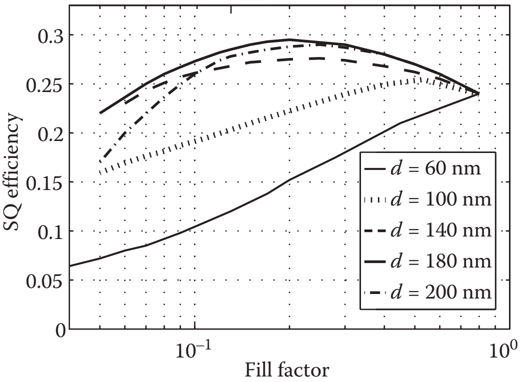


FIGURE 42.6 Efficiency of an InP nanowire array based on the Shockley–Queisser analysis in the radiative limit for different wire diameters d. An infinitely thick, perfect antireflective InP bulk cell has an efficiency of approximately 32%.

42.4.4 Purcell Effect

The nanowire array acts as a microconcentrator, and therefore, as discussed in Section 42.2.4.2, the open-circuit voltage can be increased compared to a bulk solar cell, if the concentration factor outweighs the increase in optical DOS. The reciprocity theorem states that any lens or concentrator not only increases the collection angle of the solar cell, but also the emission angle. In addition, the spontaneous emission rate (which in the bulk case is a material property) can be altered if the dielectric surrounding of the emission volume is modified, which changes the recombination rates. For the InP nanowire model array, the question is now how large the Purcell factor is and if it substantially alters the efficiency considerations described up to now.

In order to calculate the optical DOS, dipole sources were placed in the nanowire array, with vertical (z-) and horizontal (xy-) polarization. As position, the cross-section of the nanowire was sampled at a height of $z_0 = 1.9$ μm, where the absorptivity is large. A wavelength sample of $\lambda = 700$, 800, and 900 nm was used, which is close to the spontaneous emission maximum in InP. A description similar to Equation 42.7 has been used in order to calculate the local optical density of states (LDOS) (for a detailed procedure, please see [19]). In Figure 42.7, the LDOS is plotted, normalized to the free-spaced DOS. It basically describes the enhancement or suppression factor for the spontaneous emission rate; a number larger than 1 would mean that spontaneous emission is enhanced. From the calculation, the z-polarized normalized LDOS reaches values around 2–3, close to the wire surface, where the maximum absorption takes place (see Figure 42.5 for comparison). This is a photonic crystal effect, as the z-polarized dipole emits predominantly in the horizontal direction. The xy-LDOS is 0.5–0.8, close to unity. As a consequence, the averaged LDOS over all directions leads to a normalized LDOS very close to unity, which according to Equation 42.27 allows the full benefit of the optical microconcentration effect to enter the open-circuit voltage calculation.

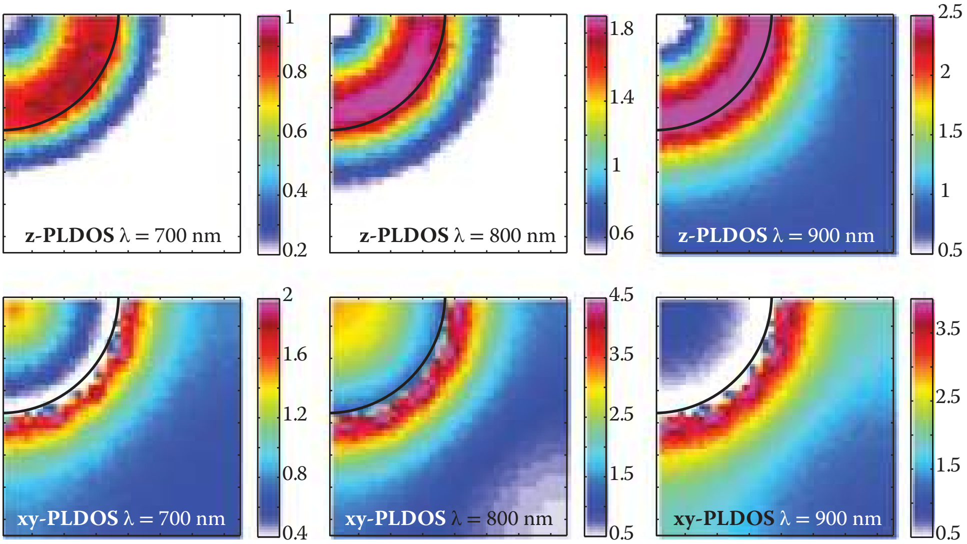


FIGURE 42.7 Local optical density of states (LDOS) relative to the bulk density of states in a nanowire array with diameter of $d = 180$ nm and pitch $a = 360$ nm. The top row is from an emitter polarized parallel to the wire axis, the bottom row for a dipole emitter with horizontal polarization. (Reprinted from J. Kupec et al., *Optics Express*, 18, 27589–27605, 2010. With permission)

42.5 Electronic Properties

42.5.1 Radial versus Axial Junction Arrangement

The nanowire geometry by nature has a high aspect ratio of length to diameter, and the p-n junction can be arranged in radial (i.e., vertical) or axial (i.e., horizontal) manner generating large or small quasi-neutral regions. Both implementations are shown in Figure 42.2 schematically and also have been demonstrated experimentally by epitaxial growth techniques. This p-n junction placement introduces another important aspect to nanostructured solar cell design. Assuming an abrupt junction, and the depletion approximation, the p-n junction can be separated into three distinct regions: the depleted region, and the quasi-neutral p- and n-regions. Photogenerated carriers in the depleted region are separated by the electric field and constitute a local photo current. In the quasi-neutral regions, the carriers diffuse, ideally to the contacts, if a local gradient is present. The cell current at the contacts is reduced if photogenerated carriers recombine on their way to the contact. This can be internal radiative or nonradiative recombination, consisting of spontaneous emission, Auger recombination, and Shockley–Read–Hall recombination. As a general guideline for minimizing the influence of nonradiative mechanisms, the recombination limited carrier lifetime needs to be much larger than the transit time of the carriers to the contacts. This principle is used in the thin-film solar cell technology.

In the analysis in Section 42.2.3, only the radiative recombination was included in the detailed balance limit. The analysis here is also done in the detailed balance limit as well, as it gives an upper boundary for the efficiency.

As example, the InP nanowire geometry is used, with a diameter of $d = 180$ nm and a wire height of 2 μm. The top n-region has a thickness of 100 nm, and the core diameter of the core-shell implementation is $d = 120$ nm. As p-doping, an active concentration of $p = 1\times10^{18} cm^{-3}$ and n-doping of $n = 5\times10^{18} cm^{-3}$ is used. The junction arrangement of the respective axial and core-shell structure is shown in Figure 42.2. The space and frequency resolved absorptivity profile has been calculated solving the vectorial Helmholtz equation under AM1.5D illumination; an illustration can be seen in Figure 42.8 after integration over frequency.

The wire array fill factor is 20%, which gives a maximum Shockley–Queisser efficiency. After spectral integration of the absorptivity, a local optical generation rate can be evaluated. The electronic simulation has been carried out only for a single wire by solving the continuity equations for holes and electrons coupled to the Poisson equation for the electrostatic potential. This means that any interaction between the

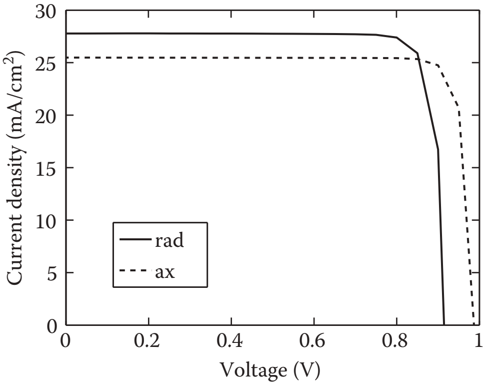


FIGURE 42.8 Current density versus voltage characteristics of axial and radial junction nanowire array solar cells made of InP, illuminated with the AM15.D solar spectrum. (ax, axial; rad, radial.)

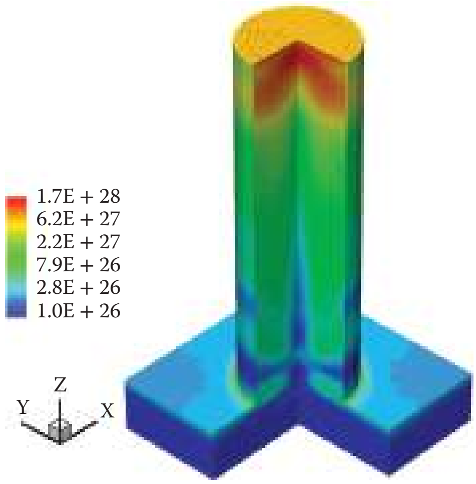


FIGURE 42.9 Optical generation rate profile in a nanowire array (only single nanowire shown) with a solar AM1.5D illumination spectrum (in units 1/s).

wires has been neglected for the carrier transport. Figure 42.9 shows the current versus voltage characteristics for the two designs. The axial p-n junction has a lower short-circuit current than the radial junction; however, it also shows a larger open-circuit voltage. The resulting maximum efficiencies are therefore close: 22% for the axial cell and 21% for the radial cell. In the axial wire, the photocurrent is generated from two effects. For one, the electric field in the depletion zone leads to drift current of the photogenerated carriers. This drift current removes carriers from the depletion zone, and in turn results in a diffusion current of the photogenerated carriers as excess minority carriers toward the junction. For the doping values in this wire, the depletion region is less than 15-nm thick at zero voltage, which would yield almost no photocurrent from direct drift. However, the simulations show that 600 nm of the quasi-neutral p-region contributes to the electron photocurrent by diffusion, and the entire quasi-neutral top n-region of 100 nm contributes to the hole photocurrent. The thickness of these diffusion regions depends critically on the minority carrier lifetime, and therefore impurity densities in the quasi-neutral regions need to be minimized. A more detailed analysis of this effect has been shown in [4]. In our simulation of the axial structure (with only radiative recombination present), a total of 700 nm of the wire height therefore can be viewed as the active region for photocurrent generation [29], which is 35% of the total length of the wire. In contrast, the radial junction contains a depletion region along the entire wire, and the charge separation due to drift and diffusion currents takes place in the entire wire volume. This leads to a higher photocurrent compared to the axial wire. The difference is not as large as the active volume ratio though, as the optical generation rate decreases exponentially from top to bottom, and the high generation top region is an active region for both structures. In [4], it has been found that the efficiency of a radial nanowire solar cell is much less prone to degradation if the minority carrier lifetime decreases, which can be directly explained by the increased depleted volume in the radial cell. It should be noted that the n-doped shell thickness of 30 nm for the radial wire has been chosen since it is larger than the depletion region so that a vertical photocurrent to the n-contact can be established.

The active region analysis explains the increased photocurrent for the radial structure, but it also can explain the reduction in open-circuit voltage compared to the axial device. The open-circuit voltage is given by Equation 42.20, with the fraction $\frac{I_{\text{gen}}}{I_{\text{sat}}}$ as contribution. Including the above described carrier transport effects, the volume integration for the calculation of the optical generation current I_{gen} and the reverse saturation current I_{sat} should only be done for the active volume. In the case of the axial junction, it is the depletion volume plus the diffusion zones in the quasi-neutral regions, which, in our specific case, account to the top third of the wire. The radial junction wire collects photocarriers in the entire wire volume, and therefore the integration volume is the entire wire. As stated, this maximizes the optical generation current, and at the same time, also increases the reverse saturation current. Toward the bottom of the radial junction wire, the integration adds the same amount of reverse saturation current as in the upper region; however, the optical generation contribution decreases to a negligible amount. This is due to the exponential decay of the relative absorptivity *a* (see Equation 42.12). As result, the open-circuit voltage decreases if the wire is too long. This is plotted in Figure 42.10, where the wire array efficiency of a radial junction cell has been calculated for different wire lengths. For wire lengths below 2 μm, the solar spectrum is not fully absorbed within the wire (especially the long wavelengths), which limits the short-circuit current. For lengths above 2 μm, the reverse saturation current increases more strongly than the photocurrent, which decreases the open-circuit voltage. It should be noted that the values for the wire length and resulting efficiencies are specific to the material, dimensions, and doping concentrations that have been chosen for this example; however, the physical principles are valid for the general case of nanowire solar cells. For silicon nanowires, the optical absorption coefficient is smaller, and therefore the optimum wire length would be larger.

42.5.2 Surface Recombination

Up to now, all the calculations have been presented with radiative recombination only, excluding any non-radiative processes. In this section, the impact of surface recombination is analyzed. Nanostructures show a large surface to volume ratio, and any surface effect has a pronounced impact on the characteristic behavior of a device. For solar cells, surface recombination contributes to the nonradiative recombination mechanisms and decreases the cell efficiency, as discussed in the previous section. Here, the goal is to study the impact of surface recombination in the cell efficiency within the framework of the carrier continuity equation and the Poisson equation. The surface recombination rate is calculated via a Shockley–Read–Hall recombination with 2D trap densities [30]:

$$R_s = \frac{n_s p_s - n_i^2}{1/v_p(n_s + n_1) + 1/v_n(p_s + p_1)}, \tag{42.33}$$

where n_i is the intrinsic carrier concentration, v_n and v_p represent the surface recombination velocities, and n_1 and p_1 represent the surface trap densities. This model is used in the simulation procedure. The surface carrier densities are n_s and p_s, respectively.

The radial and axial structures have been simulated with constant surface recombination velocity of $v_s = 10^4$ cm/s, which is a realistic value for InP [31], and surface trap densities n_1 and p_1 set to a value for midgap states . Figure 42.11 shows the current–voltage characteristics with (dashed lines) and without surface recombination (solid lines). Turning on surface recombination, the maximum efficiency drops

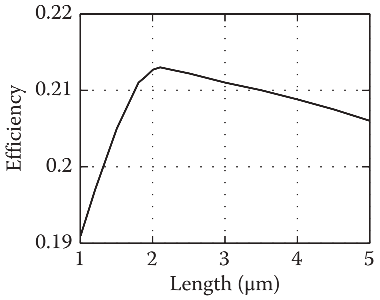


FIGURE 42.10 Cell efficiency of a radial junction nanowire array versus nanowire length.

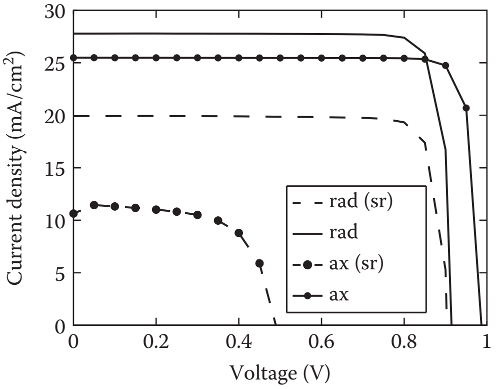


FIGURE 42.11 Current versus voltage relationship of a radial (rad) and axial (ax) junction nanowire array with and without surface recombination (sr).

from 21% to 15.5% for the radial structure and from 22% to 3.5% for the axial structure. It can be shown that the maximum surface recombination rate is achieved for the condition $v_n n_s = v_p p_s$ [32]. Assuming similar recombination velocities for electrons and holes for simplicity, the maximum surface recombination is hence located where electron and hole densities are both high at the same location. For a p-n junction, this occurs in the depletion region. It explains why the axial junction solar cell shows a substantial decrease in efficiency; in this geometry, the depletion region is exposed to the surface. This can be avoided by adding a passivation layer to the surface. For the radial junction, the depletion region thickness is smaller than the outer n-doped shell thickness, which makes the surface a quasi-neutral region. In that respect, the radial junction arrangement is vastly superior and much less prone to degradation from surface recombination effects.

42.6 Summary

Photovoltaic cells have become a fast growing technology for electrical power generation. Conversion efficiency and cost have been major drivers for the development. In that context, the use of nanostructures in solar cells has not lead to a major performance advantage up to now, and most implementations rely on classical p-n junction electronics and classical ray optics. In this chapter, the use of periodic nanowire arrays and their impact on solar cell efficiency has been analyzed. An efficiency analysis is presented that includes the nature of wave optics based on electromagnetic propagation and the optical DOS. The former leads to resonance and interference effects, which in turn causes localization of the electromagnetic energy and spectral dispersion. As main result, the efficiency of a bulk cell with ideal antireflective coating can be achieved with wires of only 20% of material without antireflective coating. In addition, efficient light management improves the solar cell efficiency via a modification of the optical DOS. This leads to a higher open-circuit voltage compared to a classical solar cell.

Acknowledgements

This work has been supported by the EU FP7 grant AMON-RA (grant agreement FP7-214814-1). The author acknowledges the contribution of J. Kupec, S. Yu on the numerical simulation results.

Appendix 42.A: Material Parameters

The InP material parameters used in the simulations are listed here. The refractive index dispersion is shown in Figure 42A.1.

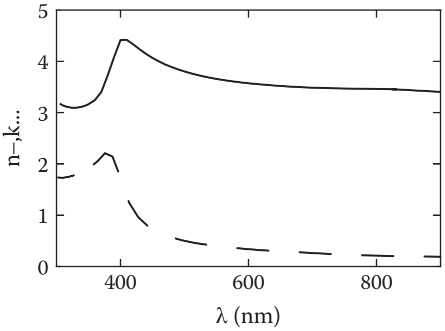


FIGURE 42A.1 InP refractive index *n* and extinction *k* versus wavelength (From S. L. Chuang, *Physics of Photonic Devices. Wiley Series in Pure and Applied Optics*, 2009. Copyright Wiley-VCH Verlag GmbH & Co. KGaA).

As mobility model for the electronic simulations, the Arora model has been used:

$$\mu = \mu_{\text{min}} + \frac{\mu_{\text{D}}}{1 + (N_i/N_0)^A}, \tag{42A.1}$$

with N_{i} being the impurity density, and the remaining parameters have been chosen as [33].

	$\mu_{min}(\text{m}^2/\text{V}\cdot\text{s})$	$\mu_D(\text{m}^2/\text{V}\cdot\text{s})$	$N_0(\text{m}^{-3})$	A
el	0.040	0.520	3.00e23	0.47
hl	0.001	0.017	4.87e23	0.62

The intrinsic carrier density is $n_{\text{i}} = 1.2 \times 10\ \text{m}^{14}\text{m}^{-3}$, the bandgap energy is $E_{\text{G}} = 1.34$ eV, and the radiative recombination coefficient is $1.2 \times 10\ \text{m}^{-16}\text{m}^3/\text{s}$.

References

1. L. Tsakalakos, J. Balch, J. Fronheiser, M.-Y. Shih, S. F. Leboeuf, M. Pietrzykowski, P. J. Codella, B. A. Korevaar, O. Sulima, J. Rand, and A. Davuluru. Strong broadband optical absorption in silicon nanowire films. *Instrumentation*, 1:1–10, 2007.
2. E. Garnett and P. Yang. Light trapping in silicon nanowire solar cells. *Nano Letters*, 10:1082–1087, 2010.
3. J. Wallentin, N. Anttu, D. Asoli, M. Huffman, I. Berg, M. H. Magnusson, G. Siefer, P. Fuss-Kailuweit, F. Dimroth, B. Witzig-Mann, H. Q. Xu, L. Samuelson, K. Deppert, and M. T. Borgstrm. InP nanowire array solar cells achieving 13.8% exceeding the ray optics limit. *Science*, 339(6123):1057–1060, 2013.
4. B. Kayes, H. Atwater, and N. Lewis. Comparison of the device physics principles of planar and radial pn junction nanorod solar cells. *Journal of Applied Physics*, 97(11):114302, 2005.
5. C. Boecklin, R. G. Veprek, S. Steiger, and B. Witzigmann. Computational study of an InGaN/GaN nanocolumn light-emitting diode. *Physical Review B*, 81:155306, 2010.
6. G. Mariani, A. C. Scofield, C.-H. Hung, and D. L. Huffaker. GaAs nanopillar-array solar cells employing in situ surface passivation. *Nature Communications*, 4:1497, 2013.

7. W. Shockley and H. J. Queisser. Detailed balance limit of efficiency of pn junction solar cells. *Journal of Applied Physics*, 32(3):510–519, 1961.
8. E. M. Purcell, H. C Torrey, and R. V Pound. Resonance absorption by nuclear magnetic moments in a solid. *Physical Review*, 69(1–2):37–38, 1946.
9. Ivan S. Nikolaev, Willem L. Vos, and A. Femius Koenderink. Accurate calculation of the local density of optical states in inverse-opal photonic crystals. *Journal Optical Society of America B*, 26(5):987–996, 2009.
10. E. Yablonovitch. Statistical ray optics. *Journal of the Optical Society of America*, 72(7):899–907, 1982.
11. J. Nelson. *The Physics of Solar Cells*. London: Imperial College Press, 2003.
12. S. L. Chuang. *Physics of Photonic Devices. Wiley Series in Pure and Applied Optics*, 2nd edition. New York, NY: Wiley, 2009.
13. K. Barnham and G. Duggan. A new approach to high-efficiency multi-band-gap solar cells. *Journal of Applied Physics*, 67:3490–3493, 1990.
14. A. Luque and A. Marti. Increasing the efficiency of ideal solar cells by photon induced transitions at intermediate levels. *Physical Review Letters*, 78:5014–5017, 1997.
15. R. T. Ross and A. J. Nozik. Efficiency of hot-carrier solar energy converters. *Journal of Applied Physics*, 53(5):3813–3818, 1982.
16. J. Kupec and B. Witzigmann. Dispersion, wave propagation and efficiency analysis of nanowire solar cells. *Optics Express*, 17(12):10399–10410, 2009.
17. A. Taove and S. C. Hagness. *Computational Electrodynamics: The Finite-Difference Time-Domain Method*, 3rd edition. Norwood, MA: Artech House, 2005.
18. B. Cockburn and C.-W. Shu. The Runge-Kutta discontinuous Galerkin method for conservation laws v: Multidimensional systems. *Journal of Computational Physics*, 141:199–224, 1998.
19. F. Roemer, B. Witzigmann, O. Chinellato, and P. Arbenz. Investigation of the Purcell effect in photonic crystal cavities with a 3D finite element nanowire solar cells: Electro-optical performance Maxwell solver. *Optical and Quantum Electronics*, 39(4–6):341–352, 2007.
20. F. Roemer and B. Witzigmann. Spectral and spatial properties of the spontaneous emission enhancement in photonic crystal cavities. *Journal Optical Society of America B*, 25(1):31–39, 2008.
21. O. D. Miller, E. Yablonovitch, and S. R. Kurtz. Strong internal and external luminescence as solar cells approach the Shockley–Queisser limit. *IEEE Journal of Photovoltaics*, 2(3):303–311, 2012.
22. D. Knig, K. Casalenuovo, Y. Takeda, G. Conibeer, J. F. Guillemoles, R. Patterson, L. M. Huang, and M.A. Green. Hot carrier solar cells: Principles, materials and design. *Physica E: Low-Dimensional Systems and Nanostructures*, 42(10):2862–2866, 2010.
23. M. Luisier and G. Klimeck. Atomistic full-band simulations of silicon nanowire transistors: Effects of electron-phonon scattering. *Physical Review B*, 80:155430, 2009.
24. W. Frensley. Wigner-function model of resonant-tunneling semiconductor device. *Physical Review B*, 36(3):1570–1580, 1987.
25. U. Aeberhard and R. Morf. Microscopic nonequilibrium theory of quantum well solar cells. *Physical Review B*, 77(12):125343, 2008.
26. U. Aeberhard. Simulation of nanostructure-based and ultra-thin film solar cell devices beyond the classical picture. *Journal of Photonics for Energy*, 4(1):042099, 2014.
27. M. Auf der Maur, G. Penazzi, G. Romano, F. Sacconi, A. Pecchia, and A. Di Carlo. The multiscale paradigm in electronic device simulation. *IEEE Transactions on Electron Devices*, 58(5):1425–1432, 2011.
28. J. Kupec, L. R. Stoop, and B. Witzigmann. Light absorption and emission in nanowire array solar cells. *Optics Express*, 18(26):27589–27605, 2010.
29. S. Yu, J. Kupec, and B. Witzigmann. Efficiency analysis of III-V axial and core-shell nanowire solar cells. *Journal of Computational and Theoretical Nanoscience*, 9(5):1–8, 2012.
30. F. Roemer and B. Witzigmann. Modelling surface effects in nanowire optoelectronics. *Journal of Computational Electronics*, 11(4):431–439, 2012.

31. S. Bothra, S. Tyaqi, S. K. Ghandhi, and J. M. Borrego. Surface recombination velocity and lifetime in InP. *Solid-State Electronics*, 34:47–50, 1991.
32. P. P. Altermatt, A. G. Aberle, J. Zhao, A. Wang, and G. Heiser. A numerical model of p-n junctions bordering on surfaces. *Solar Energy Materials and Solar Cells*, 74:165–174, 2002.
33. M. Sotoodeh, A. H. Khalid, and A. A. Rezazadeh. Empirical low-field mobility model for III–V compounds applicable in device simulation codes. *Journal of Applied Physics*, 87(6), 2000.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

43

Thin-Film Solar Cells

43.1 Introduction 497
43.2 Technology Overview 498
Silicon-Based Cells • CdTe-Based Cells • Copper Indium Gallium Selenide–Based Cells • Dye-Sensitized Solar Cells • Organic Solar Cells • Perovskites-Based Cells
43.3 Optical Modeling 503
43.4 Transport Modeling 511
Drift-Diffusion Model • Kinetic Monte Carlo • Example: Transport through 1D Disordered System by kMC
43.5 Specific Modeling Issues 527
Dye Solar Cells • Organic Photovoltaics • Perovskites • Modeling of Morphology and Grain Effects

Matthias Auf der Maur

Tim Albes

and

Alessio Gagliardi

43.1 Introduction

In this chapter, we review simulation models for thin-film solar cells, and especially the specific challenges encountered in modeling of thin-film cells based on emerging technologies. From a technological point of view, the numerous currently available photovoltaic (PV) technologies can be roughly subdivided into two classes. One is based on conventional (mono-)crystalline semiconductor technology, to which belong the first generation of solar cells such as monocrystalline silicon or gallium arsenide p-n junction cells, and the other is based on thin-film technology [1]. The term *thin-film* in this context does not primarily refer to the thickness of the devices, but to the type of production technologies used, although their physical thickness is typically small in the range of few tens of micrometer to well below 1 μm [19,90].

Thin-film technology is characterized by several advantages over conventional technology [4]. First of all, material deposition techniques are often more simple. Depending on the material system, there is a vast diversity in techniques. Many of these are low-temperature methods, when compared to standard crystalline semiconductor technology, and many of them are easily scalable to large-scale production and large device areas. In particular, solar cells based on organic polymers, e.g., can be produced from liquid solutions with processes similar to printing technology, which opens a way to very low-cost large-scale roll-to-roll production [5]. Overall, simpler and lower cost production methods and often cheaper, more abundant, or environmentally unproblematic materials are the main driving factors for development and optimization of thin-film PV concepts. Still, these technologies have to compete with costs of traditional PV, other renewable energy sources, and most of all with that of fossil energy sources in order to penetrate the energy market. In 2014, thin-film technologies accounted for roughly 9% of the global PV module production [6].

The “classical” thin-film solar cell technologies include the second generation cells like amorphous and microcrystalline silicon (a-Si, μc-Si), the chalcogenide alloy copper indium gallium diselenide (CIGS), and cadmium telluride (CdTe), which together count up for almost all of the current thin-film solar cell market.

Although the record efficiency map regularly published by the National Renewable Energy Laboratory (NREL) [7] groups only these technologies under the term "thin-film," we assume a broader interpretation by including and actually focusing on emerging concepts. These concepts are organic-based (bulk) heterojunction cells (organic photovoltaics [OPV]), dye solar cells (DSCs), and Perovskite-based cells. Recent reviews on the state of the art of these emerging technologies, also known as third-generation cells [8], can be found in [9–11].

Modeling of thin-film solar cells requires three essential ingredients, namely a model for the optical absorption, i.e., carrier or charge generation, for the charge separation, and for electronic transport. While the details of the latter depend mostly on the material system and technology of a given cell, the optical absorption is particularly critical for thin-film devices compared to devices employing thick absorbers. This is because they typically consist of several layers of different materials with thicknesses comparable to or smaller than the optical wavelength. Therefore, the optical absorption profile can be significantly determined by interference effects, so that the calculation of the optical absorption requires basically the explicit solution of the full Maxwell equations. Even if the optical absorption in some planar devices with thicker absorber layers, like DSCs, may be calculated with simplified approaches, the inclusion of light management–improving measures like scattering elements or back contact gratings usually requires a scheme based on the solution of Maxwell's equations, too.

Most of the models described in literature are using for the optical modeling constant (but material and wavelength dependent) bulk absorption coefficients, such that the optical absorption can be calculated once, providing the carrier generation profile for the electronic transport simulation. Analytical models for carrier transport can be found under suitable simplifying assumptions, as described e.g., in [1], and such models are very useful to understand basic features of solar cell behavior. Here, however, we are interested in physics-based simulation beyond analytical approximations, which provides more detailed insight into device operation and which, in particular, can be used for device optimization. Moreover, analytical models can typically be formulated only for one-dimensional (1D) problems, while numerical simulations can treat higher dimensions. This allows for simulation-based optimization of the device geometry, e.g., related to contacting schemes, which is hardly possible using only 1D simulations.

The chapter is organized as follows. First, we give a short overview on the basic features of the different thin-film technologies. Then we review the basic physics-based modeling approaches typically used for thin-film solar cells, both for optical absorption and electronic transport. Then we will describe critical details and modeling issues for the different third-generation thin-film concepts. More advanced concepts based on low-dimensional structures, hot carriers or using plasmonic effects, which generally require more complex simulation models, are not covered in this chapter.

43.2 Technology Overview

In the following sections, we shortly review the main characteristics of different thin-film technologies. For more in-depth discussions and further references, we refer, e.g., to [2,12].

43.2.1 Silicon-Based Cells

The currently most important silicon thin-film technologies are based on amorphous (a-Si) and microcrystalline (μc-Si) form. In contrast to crystalline silicon, which forms a regular tetrahedral configuration in a diamond lattice such that every silicon atom has four neighboring atoms, amorphous silicon does not show the same regular structure. The tetrahedral coordination of the atoms is mostly maintained, but random variations in bond angles and random appearance of dangling bonds, i.e., atoms having less than four neighbors, destroys the regular lattice and the associated periodicity.

The dangling bonds in amorphous silicon lead to a large number of defect states inside the bandgap, which adversely influence its electronic properties. In particular, the material can hardly be doped.

However, passivation of at least part of these dangling bonds by hydrogen atoms, known as *hydrogenation*, has been found to be necessary and sufficient to improve considerably material quality and allow it to be doped [13]. The hydrogenated form of a-Si is known as a-Si:H.

μc-Si (or more correctly μc-Si:H) is a more complex material, consisting of amorphous phases and conglomerates of silicon nanocrystals. Their microstructure depends largely on deposition methods and underlying layers, and the presence of grain boundaries between crystalline domains will influence further electronic properties.

As expected, the electronic properties of such disordered materials are worse than those of crystalline silicon. In particular, carrier mobilities are much lower. However, both materials are good enough to build working solar cells.

Both a-Si and μc-Si can be produced by a plasma-enhanced chemical vapor deposition (PECVD) technique introduced around 1970 [13,14]. The advantages of this technique over crystalline silicon deposition methods is that the required substrate temperature is significantly lower ($<600\,^\circ$C) and that it can be employed for large-area deposition. The former allows the use of cheap substrate materials such as plastic, while the latter allows to process large modules at once. Moreover, a-Si and μc-Si have two material intrinsic advantages. First, the loss of long-range order effectively leads to a transition from an indirect bandgap to a direct one, resulting in higher optical absorption coefficient. This allows to use thinner absorber layers without loss of absorption and thus to a cost reduction. Second, because of the different electronic transport properties these cells actually perform better at higher temperatures, contrary to crystalline silicon solar cells.

Typically, a-Si and μc-Si solar cells are built as p-i-n structures as shown in Figure 43.1a rather than pn-junctions, i.e., with an intrinsic layer between the n- and p-doped contact layers, since the carrier diffusion length in doped material are rather short due to the relatively poor transport properties. The transparent conductive oxide (TCO) layer in Figure 43.1a is e.g., indium tin oxide (ITO) or SnO_x. This layer is common to all thin-film solar cell technologies and serves to achieve a sufficiently low front contact resistivity without optical absorption. Cell thickness is usually rather small (around 400 nm for a-Si, few μm for μc-Si) in order to reduce recombination losses. Current best research cell efficiency is around 10% for a-Si:H and 12% for μc-Si single junction cells, compared to 25% of single-crystal silicon cells [15].

One of the most important problems in hydrogenated silicon cells at current is a light-induced performance degradation by prolonged illumination, known as the Wronski–Staebler effect, which is believed to be related to light-induced diffusion of the hydrogen atoms [16].

A recent review on thin-film Si solar technology can be found, e.g., in [17].

43.2.2 CdTe-Based Cells

CdTe is a very suitable material for PV applications. It has a direct bandgap of $E_g \approx 1.45$ eV, which is in the optimum range of values, and it has a high optical absorption coefficient $>10^4$ cm^{-1} over the whole solar spectrum, allowing for thin absorbers.

Both monocrystalline and polycrystalline CdTe cells can be produced. Polycrystalline CdTe can be deposited by a vast range of techniques, including vacuum techniques such as sublimation, evaporation, atomic layer deposition, chemical vapor deposition, but also nonvacuum techniques based on solution processes. Low-temperature processes are especially interesting since they allow for deposition on cheap and flexible substrates and for large-scale roll-to-roll process technology. Deposition is followed by a $CdCl_2$ treatment to improve material quality.

Some disadvantages of CdTe are the inability to tune the bandgap by alloying, and the presence of native defects to an amount that makes it difficult to achieve high-doping levels. The latter is amended by smart contacting schemes. At the front contact, an n-doped CdS buffer layer is introduced before the p-type CdTe absorber layer, which also acts as a buffer layer for CdTe deposition. The back contact is challenging because of the impossibility to achieve high p-doping density and since CdTe has a high electron affinity leading to high Schottky barriers for usual contact metals. Several solutions exist, e.g., suitable combinations of

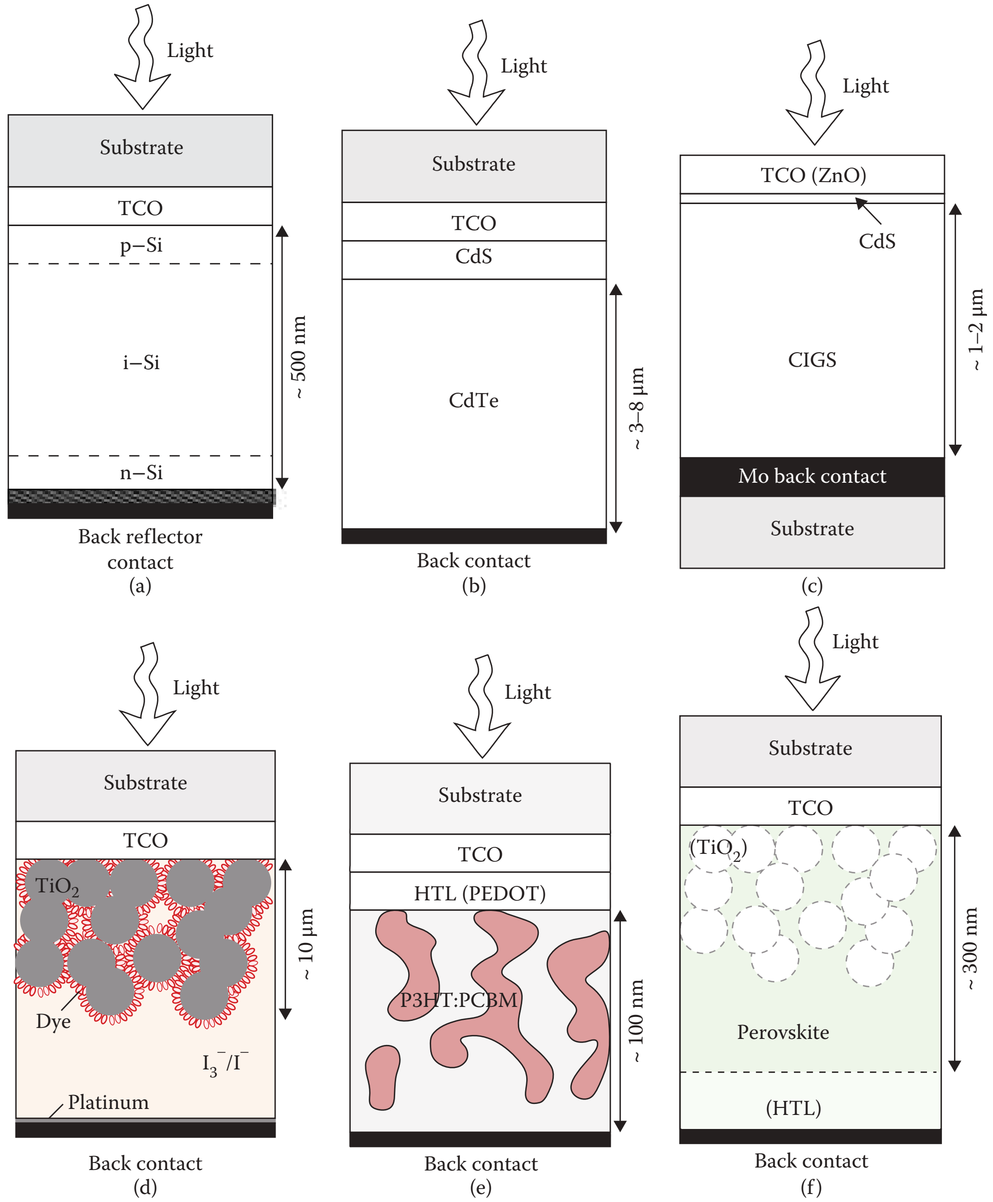

Light
Substrate
TCO
p–Si
i–Si
n–Si
~ 500 nm
Back reflector contact
(a)
Light
Substrate
TCO
CdS
CdTe
~ 3–8 μm
Back contact
(b)
Light
TCO (ZnO)
CdS
CIGS
~ 1–2 μm
Mo back contact
Substrate
(c)
Light
Substrate
TCO
TiO_2
Dye
I_3^-/I^-
Platinum
~ 10 μm
Back contact
(d)
Light
Substrate
TCO
HTL (PEDOT)
P3HT:PCBM
~ 100 nm
Back contact
(e)
Light
Substrate
TCO
(TiO_2)
Perovskite
(HTL)
~ 300 nm
Back contact
(f)

FIGURE 43.1 Typical schematic structures of different thin-film solar cells. (a) Amorphous silicon, with intrinsic absorber layer (i-Si) and doped contact layers (n/p-Si); (b) cadmium telluride (CdTe); (c) copper indium gallium selenide (CIGS); (d) dye solar cell (DSCs): TiO_2 nanoparticles are in the order of 10 nm diameter; (e) organic solar cell: the characteristic dimension of the absorber blend is in the order of 10 nm; and (f) perovskite cell, in some versions including a mesoporous TiO_2 scaffold as used in DSCs. Typical absorber thicknesses are indicated in the figures. TCO is the transparent conducting oxide, like indium tin oxide (ITO) or zinc oxide (ZnO), HTL is the hole transport layer.

buffer layers with metallization or particular surface treatments. Transport and recombination properties in CdTe films depend much on deposition technology and on the properties of the polycrystallinity of the film. The best research cell efficiency reported so far is 21% [15]. A typical CdTe solar cell scheme is shown in Figure 43.1b.

CdTe thin-film solar cells have been commercialized successfully and in 2014 they covered roughly 44% of the market share of all thin-film technologies [6]. It has to be noted that the presence of cadmium in CdTe has induced some doubts about the environmental compatibility of this technology. Recent reviews can be found in [18,19].

43.2.3 Copper Indium Gallium Selenide–Based Cells

Chalcopyrite-based solar cells is the second most important thin-film technology nowadays with a market share of 39% [6]. The most relevant compounds of this material class for PV applications are $CuInSe_2$, $CuInS_2$, and $CuGaSe_2$ having a bandgap of 1.0, 1.5, and 1.7 eV, respectively. These compounds can be easily alloyed, providing an extended range of bandgaps and lattice constants. Typically, the absorber consists of a $Cu(In,Ga)(S,Se)_2$ alloy, giving the name "CIGS," or of $CInS_2$ which has a gap of 1.5 eV. The highest cell efficiency reported in 2015 was around 21%, higher than multicrystalline Si cells [15]. Beside significantly higher efficiency than a-Si thin-film cells, CIGS cells show better stability and have no light-induced degradation.

A typical CIGS cell is shown in Figure 43.1c [20]. The nontransparent back contact is typically molybdenum sputtered onto a suitable substrate. Then follows a polycrystalline p-type CIGS absorber layer of few micrometer thickness. A heterointerface with a very thin n-type CdS is subsequently formed, followed by a TCO-like heavily doped ZnO as front contact and window layer.

Like CdTe and a-Si, CIGS cells can be produced on flexible substrates, allowing for specialized application fields. Reviews can be found, e.g., in [21,22].

43.2.4 Dye-Sensitized Solar Cells

A dye-sensitized solar cell (DSC) is an electrochemical system where the absorption and charge transport components are spatially separated [23]. The standard device is made of a mesoporous oxide, usually titanium dioxide (TiO_2), composed of nanoparticles of 15–20 nm in diameter. The surface of the mesoporous oxide is covered by a monolayer of molecules, the dye, conventionally organometallic complexes like the N719 [24] based on ruthenium. Finally, the oxide is sourrounded by a liquid electrolyte containing a redox pair. The most common one is triiodide/iodide pair (I_3^-/I^-), but many other redox pairs have been tried. In particular, cobalt-based redox pairs [25] have shown interesting properties. The contacts also are made of a glass with a transparent conducting oxide layer like ITO or fluorine-doped tin oxide (FTO) on the TiO_2 side, the anode, while the cathode is a thin layer of platinum deposited on a metallic substrate. In the device, the dye is the active material which absorbs light, while the redox pair in the electrolyte and the TiO_2 are the hole and electron transporting materials, respectively. The mesoporous structure of the TiO_2 is necessary in order to increase the surface for the dye and thus enhance dramatically the absorption efficiency, since necessarily only a monolayer of dye molecules can be used in order to efficiently extract the electrons and regerenate the dye.

The photoconversion in a DSC is far more complex than in conventional PV devices. After the absorption of the photon by the dye, the electron is transferred to the TiO_2 by an extremely fast charge transfer. Then the electron percolates in the mesoporous TiO_2 until it is collected by the ITO or FTO at the anode. The oxidized dye is regenerated by an electrochemical process with the redox pair. In a standard DSC this means that iodide is consumed and triiodide is generated nearby of the dye. The process generates an excess of triiodide with respect to the equilibrium concentration of the I_3^-/I^- couple in the bulk electrolyte. This concentration gradient moves triiodide toward the cathode where it is transformed back into iodide by taking two electrons from the catalyst, the platinum, closing the circuit. The platinum takes electrons directly from the external circuit.

There are several advantages of a DSC: it is a thin-film device, with a thickness in the range of tens of micrometer; it is quite cheap as the most expensive components (platinum and dye) are used in tiny amounts (few nanometer in thickness and a single molecular layer, respectively); the decoupling of the absorption and transport components strongly reduces recombination processes as electron and hole pairs are immediately separated on different materials at the very beginning of the photoconversion process; and fabrication processes are very cheap.

The main drawback of the device is the liquid component that can easily leak out from microcracks in the encapsulation material or be polluted by the external environment. Moreover, the most common redox pair, triiodide/iodide, is highly corrosive leading to problems in the metals used to make the external contacts and degradation of the dye layer due to desorption of molecules from the surface. All these effects are obviously detrimental for the life span of the device. In order to correct these problems, an alternative DSC has been proposed [26], where the liquid electrolyte for hole transport is substituted by an organic semiconductor (e.g., Spiro-OMeTAD or P3HT [27]). The advantage is that the new hole transporter is more stable and noncorrosive; however, the charge transport in these organic semiconductors is worse leading to higher recombinations at the interface. Moreover, the intermixing between the organic material and the mesoporous TiO_2 is less efficient. DSCs have reached 12% of confirmed efficiency [15]. Recent reviews can be found in [28,29].

43.2.5 Organic Solar Cells

OPV cells are classified as organic because their active layer consists entirely of carbon-based materials, such as polymers or small organic molecules. A major advantage of OPVs is their compatibility with techniques suitable for large-scale and low-cost fabrication, such as roll-to-roll processing. The ability to chemically engineer organic materials allows to create a variety of compounds suitable for OPV applications. Especially due to the high absorption coefficient of specific polymers, very thin layers of 50–300 nm provide sufficient absorption and make OPVs a favorable choice for thin-film solar cells. Additional features, such as mechanical flexibility and low weight, open up the possibility for a variety of applications.

A peculiarity of the organic materials used for OPVs is their low permittivity ($\varepsilon_r \approx 3 - 4$), which leads to a weak screening of electrostatic interactions. On absorption of a photon, an electron–hole pair is created. Due to the strong Coulomb interaction between the electron and the hole, the pair forms an exciton with binding energies of several hundred millielectron volts (meV). The thermal energy of 25 meV at room temperature is not large enough to split the exciton and separate the charges. Therefore, the concept of a heterojunction composed of a (electron) donor, usually a polymer or small molecule, and an (electron) acceptor was introduced. The most prominent acceptor material is a derivative of the Buckminsterfullerene, [6,6]-phenyl-C61 butyric acid methyl ester (PCBM), because its large electronegativity has shown major improvement in exciton dissociation [30].

The first organic solar cell with efficiency of >1% was developed by Tang [31] and consisted of a planar heterojunction between donor and acceptor materials. Planar architectures have the disadvantage of a low exciton splitting efficiency, since excitons typically have a very small diffusion length and thus only excitons created nearby the junction can diffuse to the interface and become separated before recombination. A boost in device performance was achieved by introducing the concept of a bulk heterojunction (BHJ) [32], which is an intermixed morphology between donor and acceptor materials, as indicated schematically in Figure 43.1c. It provides a trade-off between efficient exciton splitting and good charge transport. The spatial structure of the BHJ is crucial for device performance and is still object of intense study.

The pioneering organic devices nowadays are based on the polymer/fullerene bulk-heterojunction concept with efficiencies of up to 10.8% [15,33]. Although much progress has been achieved in the last decade, higher efficiencies are needed in order for organic solar cells to be commercially competitive with established (inorganic) technologies. Since the experimental investigation of OPVs during operation is difficult, simulations can give valuable insights into the still not fully understood charge generation and transport processes.

More details and reviews on OPV can be found, e.g., in [9,34,35].

43.2.6 Perovskites-Based Cells

Since the first attempts in 2009 and a breakthrough in 2012 regarding efficiency and stability [36], the organometal lead halide perovskite $CH_3NH_3PbX_3$ (with X = Cl, Br, or I) has attracted great attention as absorber material in PV devices. Originally, perovskites have been used as alternative dyes in DSCs, but then perovskite-based solar cells started forming a PV technology on their own. Structurally, these perovskites consist of an ionic cage built from PbX in a tetragonal structure, with an organic methylammonium (CH_3NH_3) molecule at the center [37].

One of the most astonishing properties of perovskite-based solar cells is the fact that perovskite layers can be very easily deposited with simple solution processes, and at the same time easily reaching high photoconversion efficiencies. In fact, recent perovskite cells show efficiencies comparable to the best CdTe and CIGS cells, and thus much higher than all other emerging PV concepts [15]. Such a good performance is due to good absorption properties and good, balanced electron and hole transport properties.

A further useful property of methylammonium lead halide perovskites ($MAPbX_3$) is the ability to tune the bandgap by alloying $MAPbI_3$ with $MAPbBr_3$, which allows to extend the absorption spectra up to 800 nm.

Since perovskite solar cells originally emerged from the DSC technology, their structure often resembles a DSC. Usually, a perovskite layer of few hundred nanometer thickness is deposited on a TiO_2 layer, with or without mesoporous phase, followed by a hole transport layer (HTL), as shown in Figure 43.1f. However, the latter may even be omitted [38].

Substantially, perovskite technology is "dirty," cheap but performing at the same time, reaching efficiencies as high as 20% at cell level [15]. However, it struggles with several open issues. Most importantly, stability is still very poor [39], which is mostly due to the sensitivity of the $MAPbI_3$ to humidity [40]. Second, often the measured current–voltage characteristics show pronounced hysteresis, which is nowadays believed to be due to migration of charged defects, especially lead vacancies, in the perovskite. This poses difficulties for the accurate experimental evaluation of cell efficiencies. Finally, toxicity of lead halide perovskites may be an issue. In fact, lead compounds are toxic, and therefore often banned by law and certainly unpopular in industry. Identification of lead-free perovskites, however, did not yet lead to satisfactory candidates. It seems, that lead-based halides provide best performance.

Recent reviews on perovskite-based solar cells can be found in [37,38,40,41].

43.3 Optical Modeling

The calculation of the optical field distribution in a solar cell is a critical modeling step for any solar cell, as it provides the carrier generation profile, which is the most important factor determining the photocurrent. The generation of carriers is related to the optical absorption of a material for a given wavelength λ and can be written as [42,43]

$$G(\lambda; x) = \alpha(\lambda; x) I(\lambda; x) \tag{43.1}$$

where G is the carrier generation rate per wavelength interval at a given incident free-space wavelength λ at position x, usually given in units of $cm^{-3}s^{-1}nm^{-1}$; α is the optical absorption coefficient at wavelength λ, with units cm^{-1}; and $I(\lambda; x)$ is the local light intensity given as photon flux in $cm^{-2}s^{-1}nm^{-1}$. In Equation 43.1, we assume that every absorbed photon produces a carrier, which may be an unbound electron–hole pair or an exciton. The total generation profile is then obtained by summing the contributions of all relevant wavelengths present in the incident light. The knowledge of the local absorption requires the knowledge of the local wavelength-dependent absorption coefficient, which is usually taken as a material constant, and of the local intensity spectrum. The latter is related to the local strength of the electromagnetic field, which can be written as [44]

$$I(\lambda; x) = \frac{\lambda \varepsilon_0 n}{2h} |E(\lambda; x)|^2 \tag{43.2}$$

where ϵ_0 is the vacuum permittivity, n the refractive index, h Planck's constant, E the amplitute of the electric field, and λ the free space wavelength. Therefore, the knowledge of the electromagnetic field distribution in a solar cell device allows to calculate the local generation rate G.

While the local electromagnetic field is accessible only via simulation, it is possible to both calculate and measure the absorption spectrum α. The usual way to calculate α is by using Fermi's golden rule and evaluating the optical transition matrix elements between the involved electronic states [1,45]. In this approximation, the transition probability between two states i and f with different energies E_i and E_f is given by

$$P_{i \to f} = \frac{2\pi}{\hbar} |\langle i|H'|f\rangle|^2 \delta(E_f - E_i \pm \hbar\omega), \tag{43.3}$$

where H' is the Hamiltonian describing the light-matter interaction, treated as a perturbation, and $\hbar\omega$ is the photon energy. $\langle i|H'|f\rangle$ in Equation 43.2 is the matrix element coupling initial and final states i and f, and its calculation requires an explicit representation of these states in some suitable basis. The delta function is an expression of the conservation of energy, and the minus sign describes photon absorption, while a positive sign describes photon emission, i.e., radiative recombination. The net rate of photogeneration is then given by the product of transition probability with the number of occupied initial and available final states, and by subtracting emission from absorption. The occupation of the states is given by the carriers' Fermi–Dirac distribution functions f_i and f_f. Dividing the net rate by the incident photon flux leads to the absorption coefficient [45]

$$\alpha(\hbar\omega) = \frac{\pi e^2}{n_r c \varepsilon_0 m_0^2 \omega V} \sum_{\substack{\mathbf{k}_i, \mathbf{k}_f, \\ \sigma_i, \sigma_f}} |\hat{e} \cdot \mathbf{p}_{fi}|^2 \delta(E_f - E_i - \hbar\omega)(f_i - f_f) \tag{43.4}$$

Here, e is the elementary charge, n_r is the relative dielectric constant, ε_0 is the permittivity, m_0 is the electron mass, V is the normalization volume, and the optical matrix element has been substituted by the momentum matrix element applying the dipole approximation [45]. The summation has to be done over the crystal momentum **k** and all quantum numbers σ of both initial and final states. For absorption and at not too high optical excitation, one can usually assume all initial states as occupied and all final states as empty, i.e., $f_i \approx 1$ and $f_f \approx 0$.

Radiative recombination is obtained in a similar way; however, in that case only transitions between occupied electron and hole states are relevant. Therefore, the summation can be restricted to states with energies near the band extrema. For absorption, however, all transitions are relevant, and therefore the electronic states have to be known with sufficient accuracy for all energy bands and in the whole reciprocal k-space (or more apropriately in the full first Brillouin zone). This makes calculation of absorption spectra more difficult, so that usually measured spectra are used.

In principle, the local electric field strength at all relevant wavelengths can be obtained by solving Maxwell's equations using one of the standard methods such as finite difference time domain (FDTD) [46], finite elements (FEM) [47], or rigorous coupled-wave analysis (RCWA) [48], but this is not always necessary. For a homogeneous thick absorber, in the absence of coherent effects, the Beer–Lambert approximation can be used [1]. It allows to calculate the local optical intensity simply considering the absorption along the optical path, starting from Equation 43.1. Assume a monochromatic light beam with wavelength λ with intensity I traveling in a material with absorption coefficient α. When it travels an infinitesimal distance dx, the intensity of the beam reduces by $\alpha I \mathrm{d}x$, since a part of the photons are absorbed. We can therefore deduce a continuity law for the light intensity given by

$$\frac{\mathrm{d}I}{\mathrm{d}x} = -\alpha I \tag{43.5}$$

The solution to this is an exponential function, describing exponential decay of the intensity known as Beer–Lambert law, $I(x) = I_0 \mathrm{e}^{-\alpha x}$, where I_0 is the intensity of the incident beam (after subtraction of reflection losses, i.e., immediately inside the device) and x the distance from the point of incidence. In the case of spatially inhomogeneous absorption, the result is easily generalized to

$$I(\lambda; x) = I_0(\lambda)\mathrm{e}^{-\int_0^x \alpha(\lambda;x)\,\mathrm{d}x} \tag{43.6}$$

Equation 43.6 together with Equation 43.1 are used succesfully for the simulation of solar cells where layer thicknesses are larger than the optical wavelength. However, since this model is based on a ray optics description of light, it will break down whenever effects like interference or diffraction are important. In that case a more rigorous approach is needed to calculate the optical intensity profile.

Under the assumption of a flat, planar layered structure, a more precise model can be formulated, without increasing much the computational efforts. This model is based on a 1D approximation along the optical path of a plane wave, including sequential reflections at material interfaces and transmissions in homogeneous portions of the device, and is known as the *Transfer Matrix Method* (TMM) [49,50]. Today, this is probably the most used model for one-dimensional optical modeling of thin-film devices. The model has been first applied to organic thin-film solar cells in the late 1990s [51–53], and since then it has become a standard tool [54–60].

The basic principle of the TMM is shown schematically in Figure 43.2. To apply the method, we assume that the device can be sliced into layers of homogeneous properties with flat, parallel interfaces. We do not include a possible thick substrate that should be treated with an incoherent method (details on this can be found, e.g., in [44]). Further, we assume a monochromatic optical field written as $E(x)\exp(i\omega t)$, where $E(x)$ is a complex local field amplitude, where x is the coordinate along the stack, with $x = 0$ being the point of incidence. For simplicity, we assume a transverse electromagnetic (TEM) field with orthogonal incidence here, and in the following we drop the harmonic time dependence.

The amplitudes of the electric and magnetic field at each position can be found by using two relations derived from Maxwell's equations. The first relates the fields at the left interface of a layer to that at the right interface by means of the free propagation in a homogeneous medium, while the second relates the fields at the left and the right of each interface by means of continuity conditions. Maxwell's equations in

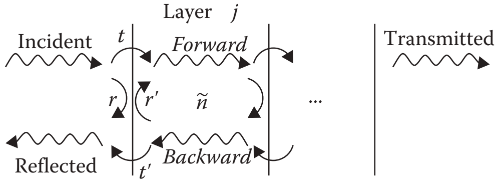


FIGURE 43.2 Scheme of the transfer matrix method. The device is assumed to be sliced in homogeneous layers, and monochromatic light is incident from the left. We decompose the optical field in a forward and backward traveling component. At each interface between two layers, these two components undergo reflection and transmission, r and t, while inside the layer the waves propagate freely with propagation constant according to the complex refractive index $\tilde{n}$.

homogeneous form (without source terms) are given by [49,61]

$$\nabla \times \mathbf{E} = -\frac{\partial \mathbf{B}}{\partial t}, \quad \nabla \cdot \mathbf{D} = 0 \tag{43.7}$$

$$\nabla \times \mathbf{H} = \frac{\partial \mathbf{D}}{\partial t}, \quad \nabla \cdot \mathbf{B} = 0 \tag{43.8}$$

$$\mathbf{B} = \mu \mathbf{H}, \quad \mathbf{D} = \varepsilon \mathbf{E} \tag{43.9}$$

Here, **E**, **H**, **D**, and **B** are the electric field, the magnetic field (or magnetizing field, magnetic field strength), the electric displacement, and the magnetic flux density (or magnetic induction, magnetic field), respectively. The two electric fields and magnetic fields are related by the constitutive relations (Equations 43.7 through 43.9), where ε and μ are the electric permittivity and the magnetic permeability, respectively. The latter is usually equal to the vacuum permeability μ_0 for the materials of interest in this context. Since we are interested in time harmonic fields, Equations 43.7 through 43.9 are conveniently transformed to Fourier space, leading to

$$\nabla \times \mathbf{E} = -i\omega \mathbf{B} \tag{43.10}$$

$$\nabla \times \mathbf{H} = \quad i\omega \mathbf{D} \tag{43.11}$$

while the other expressions remain the same. Note that the fields are now complex quantities. Moreover, the permittivity, which we write as product of vacuum permittivity ε_0 and relative permittivity $\tilde{\varepsilon}$, is frequency dependent (dispersive) and complex in case of lossy, i.e., absorbing materials.

In a homogeneous medium, the solution of Equations 43.7 through 43.9, or equivalently Equations 43.10 and 43.11, is found to be a propagating wave $E(x) = E_0 \exp(\pm ikx)$, where k is the propagation constant (in our case the component along x). The latter is related to the relative permittivity $\tilde{\varepsilon}$, or, equivalently, to the refractive index $\tilde{n} = \sqrt{\tilde{\varepsilon}}$ by [43]

$$k = \frac{2\pi}{\lambda} = \frac{\omega}{c} = \frac{\omega\sqrt{\tilde{\varepsilon}}}{c_0} = \frac{2\pi}{\lambda_0}\tilde{n} \tag{43.12}$$

where c_0 and λ_0 are the vacuum light speed and the free space wavelength, respectively; $\omega = 2\pi f$ is the angular frequency, and we have used the expression for the light speed $c = 1/\sqrt{\varepsilon\mu}$. Note that the refractive index $\tilde{n} = n_0 - i\kappa$ is a complex quantity, where the imaginary part is related to absorption.

It is convenient to decompose the electric field into *forward* and *backward* travelling components E^+ and E^-, as indicated in the figure. This is written as

$$E(x) = E^+(x) + E^-(x) \tag{43.13}$$

where once again all amplitudes are complex quantities.

We can now relate the field amplitudes at the left and right interface of the layer j by

$$E^+_{j;\mathrm{R}} = E^+_{j;\mathrm{L}} \mathrm{e}^{-ik_j t_j} \tag{43.14}$$

$$E^-_{j;\mathrm{R}} = E^-_{j;\mathrm{L}} \mathrm{e}^{ik_j t_j} \tag{43.15}$$

where k_j and t_j are the propagation constant and thickness of layer j, respectively. This can be put into matrix form as

$$\begin{pmatrix} E^+ \\ E^- \end{pmatrix}_{j;\mathrm{R}} = \mathbf{M}_j \begin{pmatrix} E^+ \\ E^- \end{pmatrix}_{j;\mathrm{L}}, \quad \mathbf{M}_j = \begin{pmatrix} \mathrm{e}^{-ik_jt_j} & 0 \\ 0 & \mathrm{e}^{ik_jt_j} \end{pmatrix} \tag{43.16}$$

where we have defined the transmission matrix of layer j as $\mathbf{M}_j$. Comparing Equation 43.16 with Equation 43.6 and considering Equation 43.2, the relation between Im(k), commonly called *extinction coefficient*, and the absorption coefficient α can be immediately found as

$$\alpha = \frac{4\pi}{\lambda}\kappa \tag{43.17}$$

Based on Maxwell's equations, one can derive the boundary conditions for the in-plane and orthogonal components of the electric and magnetic fields at planar material interfaces (orthogonal to the x-axis) as shown in Figure 43.3, which are well known and given by

$$\mathbf{n}_{lm} \times (\mathbf{E}_m - \mathbf{E}_l) = 0 \tag{43.18}$$

$$\mathbf{n}_{lm} \times (\mathbf{H}_m - \mathbf{H}_l) = 0 \tag{43.19}$$

i.e., the tangential components of **E** and **H** need to be continuous. In the above equations, $\mathbf{n}_{lm}$ denotes the normal on the interface between layers l and m. The orthogonal components of **D** and **B** have to be continuous, which is not relevant in our case of normal incidence.

Furthermore, using Equation 43.10 we can express the magnetic field in terms of the electric field as

$$B^{\pm} = \pm\frac{\tilde{n}}{c_0}E^{\pm} \tag{43.20}$$

which is obtained by evaluating the curl of the electric field and using Equation 43.12. With this we can rewrite Equations 43.18 and 43.19 in terms of the forward and backward traveling electric field components as

$$E_l^+ + E_l^- = E_m^+ + E_m^- \tag{43.21}$$

$$\tilde{n}_l(E_l^+ - E_l^-) = \tilde{n}_m(E_m^+ - E_m^-) \tag{43.22}$$

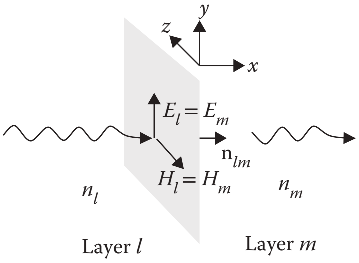


FIGURE 43.3 A material interface and the respective boundary conditions for the in-plane field components.

which can be rearranged such as to provide the transfer matrix of the material interface

$$\begin{pmatrix} E^+ \\ E^- \end{pmatrix}_{l,\mathrm{R}} = \underbrace{\frac{\tilde{n}_l + \tilde{n}_m}{2\tilde{n}_l} \begin{pmatrix} 1 & \frac{\tilde{n}_l - \tilde{n}_m}{\tilde{n}_l + \tilde{n}_m} \\ \frac{\tilde{n}_l - \tilde{n}_m}{\tilde{n}_l + \tilde{n}_m} & 1 \end{pmatrix}}_{\mathbf{I}_{lm}} \begin{pmatrix} E^+ \\ E^- \end{pmatrix}_{m,\mathrm{L}} \tag{43.23}$$

In the expressions for the elements of the transfer matrix $\mathbf{I}_{lm}$, we can identify the reflection and transmission coefficients of the interface between adjacent layers l and $m = l + 1$ with respect to the amplitudes

$$r = r_{lm} = \frac{E^-_{l;\mathrm{R}}}{E^+_{l;\mathrm{R}}} = \frac{n_l - n_m}{n_l + n_m} \tag{43.24}$$

$$t = t_{lm} = \frac{E^+_{m;\mathrm{L}}}{E^+_{l;\mathrm{R}}} = \frac{2n_l}{n_l + n_m} = 1 + r_{lm} \tag{43.25}$$

$$r' = r_{ml} = \frac{E^+_{m;\mathrm{L}}}{E^-_{m;\mathrm{L}}} = -r_{lm} \tag{43.26}$$

$$t' = t_{ml} = \frac{E^-_{l;\mathrm{R}}}{E^-_{m;\mathrm{L}}} = \frac{1 - r^2_{lm}}{t_{lm}} \tag{43.27}$$

so that we can write

$$\mathbf{I}_{lm} = \frac{1}{t_{lm}} \begin{pmatrix} 1 & r_{lm} \\ r_{lm} & 1 \end{pmatrix} \tag{43.28}$$

Combining Equations 43.16 and 43.28 we obtain the transfer matrix for the fields at the left interface between layers l and $m = l + 1$ as

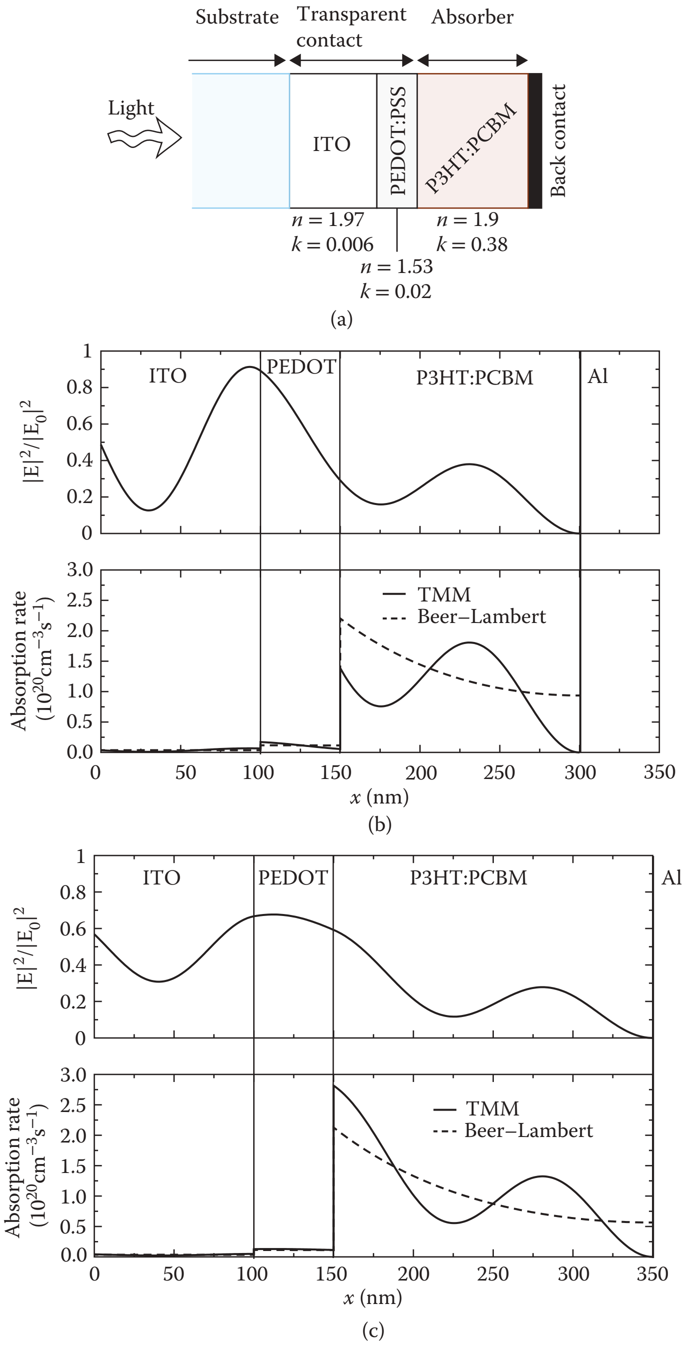
$$\begin{pmatrix} E^+ \\ E^- \end{pmatrix}_{l} = \mathbf{M}_l \mathbf{I}_{lm} \begin{pmatrix} E^+ \\ E^- \end{pmatrix}_{m} \tag{43.29}$$

and finally the transfer matrix $\mathbf{T}$ relating the fields to the left and to the right of stack as a product

$$\mathbf{T} = \prod_l \mathbf{M}_l \mathbf{I}_{l,l+1} \tag{43.30}$$

Note that TMM can be implemented in several ways, and that it is inherently numerically unstable for larger layer thicknesses [62]. Stabilization is possible, e.g., by resorting to alternative representations of the transfer matrix [63].

As an example for the importance of interference effects in layered structures, we present a comparison of the local field strength and generation rate between results of a TMM calculation and those of an incoherent calcuation. For the latter, we considered the reflectance and transmittance at each material interface and summed up the intensities of the forward and backward traveling waves. The test structure, similar to the one used in [64], is shown in Figure 43.4a. For simplicity, we assumed total reflection at the aluminum back contact, and we did not consider the glass substrate. The results have been obtained for a wavelength of 500 nm. Figure 43.4b and c presents the field and generation profiles for two different



(a)

(b)

(c)

FIGURE 43.4 Structure for comparison of coherent and incoherent optical field calculation. The parameters used in the simulation are given in (a). In (b) and (c) the optical field intensity and absorption profiles for two different absorber thicknesses is shown. The local absorption rate considerably deviates from the Beer–Lambert prediction, in particular for thin layers. For thick layers Beer–Lambert behavior is recovered. (ITO, indium tin oxide; PCBM, [6,6]-phenyl-C61 butyric acid methyl ester;TMM, transfer matrix method.)

absorber thicknesses, showing that the profiles differ significantly from what would be expected from the Beer–Lambert model, and that it depends critically on layer thicknesses. In particular, absorption maxima can be located deeply inside the layer instead of near the interface. When the layers are becoming optically thick, the local absorption rate mediated over a wavelength converges to the result obtained by the Beer–Lambert law.

Figure 43.5 shows another example, borrowed from Reference [53], where the obtained results have been compared with experimental data. In this case, the donor–acceptor heterojunction is formed by the

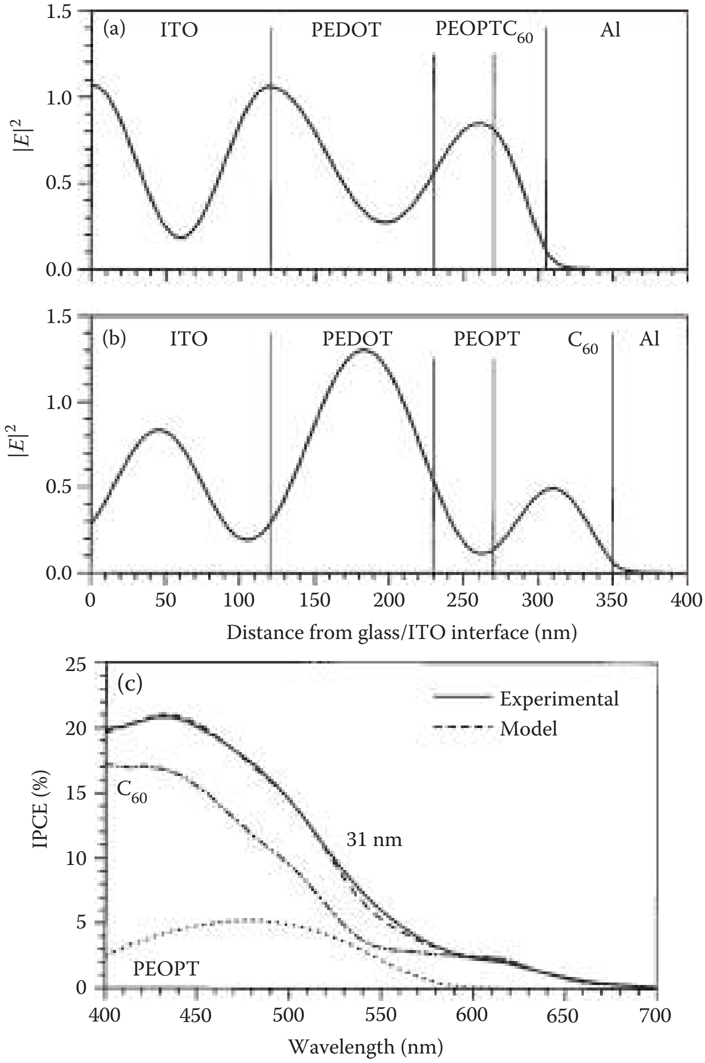


FIGURE 43.5 Field intensity profiles in an organic heterojunction solar cell for two different thicknesses of the C_{60} layer (a, b). The quality and accuracy of the optical simulation is proven by the comparison of experimental and simulated quantum efficiency in (c). Note that the simulation provides the information on where absorption happens, which cannot be obtained from the measurement. For example, in (c) the wavelength resolved contributions to the total absorption in the C_{60} and PEDOT layers is reported for an optimized C_{60} layer thickness. (Reproduced with permission from Leif A.A. Pettersson et al. Modeling photocurrent action spectra of photovoltaic devices based on organic thin films. *Journal of Applied Physics*, 86(1):487—496, 1999 Copyright 1999, American Institute of Physics Publishing LLC.)

PEOPT/C_{60} interface. To achieve maximum efficiency, a peak in field intensity should be obtained near this interface. Figure 43.3 presents the intensity profile at a wavelength of 460 nm for a C_{60} layer thickness of 35 and 80 nm, respectively, indicating that device optimization can benefit substantially from optical simulations. Note that the field profile is optimized in this way for a single wavelength only. Although the field profile cannot be measured directly experimentally, the accuracy of the optical simulation can be assessed by measuring the incident photocurrent efficiency (IPCE) (also denoted as external quantum efficiency, EQE), which is the product of quantum absorptivity and the current collection efficiency at each wavelength and gives the ratio of the number of collected electrons to the number of incident photons [43]. This quantity has been simulated in [53] by combining optical and current transport simulations, and the comparison between measurement and simulation is presented in Figure 43.3 for 31-nm thick C_{60} layer, showing excellent agreement. This example also shows how a simulation can give accurate insight into internal device behavior that is inaccessible by measurement alone, opening the way to efficient device optimization.

While 1D TMM-based modeling is a powerful tool to describe the optical field in a planar layered structure, it is less suitable to treat devices where light management techniques are employed, such as (possibly random) front or back surface texturing. Such approaches are becoming more and more important in order to approach the theoretical limits of light-to-electricity conversion, especially for technologies that are already optimized from an electrical point of view [65]. For organic solar cells, e.g., there is interest in grating structures on the back contact in order to modulate the electric field profile and thus to optimize optical generation and efficiency. Plasmonic effects are also studied for improving solar cell performance [66]. To accurately model texturing, gratings, or plasmonic effects, usually full 3D Maxwell solvers are employed [67,68]. Unfortunately, such calculations are computationally quite expensive, so that approximate models formulated in one dimension are also of interest [69]. Special approaches have been devised also for other cases, like thick layers with incoherent coupling [70].

Also note that from Figure 43.1 it might seem that TMM is not suited for devices based on mesoporous or blended materials, since the materials are not really piecewise homogeneous and there are apparently no planar interfaces. However, since the characteristic dimensions are in the order of at most tens of nanometers, these composite materials can be described with sufficient accuracy as effective materials so that TMM can be applied without particular restrictions.

43.4 Transport Modeling

While optical calculations are important for the simulation of solar cells, taken alone they can only provide upper limits for cell efficiency. For a comprehensive view on cell operation, the transport of the generated carriers and their collection at the electrical contacts has to be taken into account. This section reviews the two carrier transport models that are used most often for solar cell simulations, namely the drift-diffusion model and kinetic Monte Carlo (kMC).

43.4.1 Drift-Diffusion Model

Modeling of electronic transport in thin-film solar cells is mostly based on the drift-diffusion model, which has been the work-horse of physics-based electronic device simulation since decades. A thorough introduction, derivation of the model and implementation hints can be found in different textbooks such as [71], and a discussion of theoretical and mathematical aspects can be found, e.g., in [72]. A more detailed description is provided in Chapter 50.

Since the drift-diffusion model is well known and is described in numerous books and articles, we will provide only a short generic description and concentrate more on the features important for thin-film solar cell simulation.

For a “normal” semiconductor where only electrons and holes are mobile carriers, the drift-diffusion model consists of a set of three continuity equations for the electric displacement $D = -\varepsilon\nabla\varphi$ and the two particle currents J_n and J_p:

$$\nabla(\varepsilon\nabla\varphi) = q(n - p - C) \tag{43.31}$$

$$\frac{\partial n}{\partial t} - \frac{1}{q}\nabla J_\mathrm{n} = -R, \qquad J_\mathrm{n} = q(D_\mathrm{n}\nabla n - \mu_\mathrm{n} n\nabla\varphi) \tag{43.32}$$

$$\frac{\partial p}{\partial t} + \frac{1}{q}\nabla J_\mathrm{p} = R, \qquad J_\mathrm{p} = -q(D_\mathrm{p}\nabla p + \mu_\mathrm{p} p\nabla\varphi) \tag{43.33}$$

This nonlinear system of partial differential equations is also known as the *basic semiconductor device equations* or the van Roosbroeck system, who was the first to formulate this set of equations in the framework of electronic transport in semiconductors [73]. The first of the above equations (Equation 43.31) is the Poisson equation, relating the electrostatic potential φ with the total charge density $\rho = q(p - n + C)$ where n and p are the electron and hole densities, respectively, while ε is the permittivity. C denotes the net density of fixed charges such as dopants or traps. The other two equations are the continuity equations for the electron and hole currents, where the constitutive equations for the currents are given on the right (which are also known as *drift-diffusion equations*) where q is the electron's charge, $D_\mathrm{n,p}$ are the electron and hole diffusion coefficients, and $\mu_\mathrm{n,p}$ are the electron and hole mobilities. R is the net recombination rate, including all processes which lead to destruction or generation of electron–hole pairs.

The three most important recombination models are the trap-assisted (Shockley–Read–Hall [SRH]) recombination, bimolecular (direct, radiative) recombination, and Auger recombination, given by [71,74]

$$R_\mathrm{SRH} = \frac{np - n_i^2}{\tau_\mathrm{p}(n + n_t) + \tau_\mathrm{n}(p + p_t)} \qquad \text{SRH} \tag{43.34}$$

$$R_\mathrm{rad} = B(np - n_i^2) \qquad \text{Radiative} \tag{43.35}$$

$$R_\mathrm{auger} = (C_\mathrm{n} n + C_\mathrm{p} p)(np - n_i^2) \qquad \text{Auger} \tag{43.36}$$

Here, n_i^2 is the square of the intrinsic carrier density, given by the product of the equilibrium densities n_0 and p_0. n_t and p_t are parameters depending on the energetic position inside the gap of the trap. Note that Equation 43.34 is valid for a single discrete trap level and in the absence of an electric field, however, it can be generalized to other cases [75,76]. The other parameters in the expressions are mostly material related. The relative importance of the different recombination mechanisms depends on the operating conditions and on material quality. SRH recombination is important at low carrier densities and in presence of defects, while Auger recombination usually becomes dominant only at high carrier densities.

Written in standard form (Equations 43.31 through 43.33), the particle currents are seen to be decomposed into a diffusion term proportional to the gradient of the carrier density and a drift component proportional to the conductivity times the electric field. This is an intuitive description of particle transport, whenever the system is dominated by scattering, that is when the particles undergo Brownian motion with some thermal velocity, and the external perturbation (electric field or density gradient) induces only a small nonzero mean carrier velocity leading to a net current flow.

The system of Equations 43.31 through 43.33, supplemented with suitable boundary conditions, is usually solved numerically, although under certain circumstances analytic solutions exist in one dimensions. For implementation details and numerical methods, we refer to [71] and the vast existing literature on the topic.

The device equations (Equations 43.31 through 43.33) are still the most important model for transport in semiconductor devices, since they can be solved much faster than any other transport model and therefore represent a good compromise between computational efficiency and physical accuracy. While the equations allow easy interpretation, and can be solved relatively easily, they are based on a number of assumptions that are important when applying them to systems other than crystalline semiconductors. The other important point to note is that they give a good description of device behavior only when good models for the quantities appearing as parameters in Equations 43.31 through 43.33 can be provided. In particular, the models for the recombination and generation processes and for the carrier mobilities, in combination with accurate material-dependent parameters, are important in order to reach satisfactory results with the drift-diffusion model. Most often this implies careful parameter fitting against experimental data, careful checking whether the fitted parameters provide a predictive model or not.

In the following, we discuss some of the basic assumptions underlying the semiconductor device equations (Equation 43.31 through 43.33) and possible consequences when applying them to the materials of interest in this chapter. A discussion of technology-specific issues regarding the basic equations or the parameter models is deferred to Section 43.5.

The standard way to derive the drift-diffusion equations is to start from the semiclassical Boltzmann transport equation [72]. These derivations include implicitly or explicitly assumptions on the distribution function such that the carrier density takes the form of the well-known Boltzmann approximation, i.e.,

$$n = N_\mathrm{c}\mathrm{e}^{\frac{E_\mathrm{F,n}-E_\mathrm{c}}{k_\mathrm{B}T}}, \quad p = N_\mathrm{v}\mathrm{e}^{-\frac{E_\mathrm{F,p}-E_\mathrm{v}}{k_\mathrm{B}T}} \tag{43.37}$$

Here, N_c and N_v are the effective density of states (DOS) in the conduction and valence band, $E_\mathrm{F,n}$ and $E_\mathrm{F,p}$ are the electron and hole quasi Fermi levels, and E_c and E_v are the conduction and valence band edges. The latter can be expressed as $E_\mathrm{c,v} = E^{(0)}_\mathrm{c,v} - q\varphi$, where $E^{(0)}_\mathrm{c,v}$ are material band edges in the absence of an electric field, identifiable with electron affinity and ionization energy, k_B is the Boltzmann constant, and T the temperature.

The diffusion constants and mobilities in Equations 43.31 through 43.33 are not independent, which can be easily seen when substituting Equation 43.27 into Equations 43.31 through 43.33 and solving for equilibrium, i.e., zero current under no external bias nor carrier generation. It is found that the two quantities are related by a so-called Einstein relation [74]:

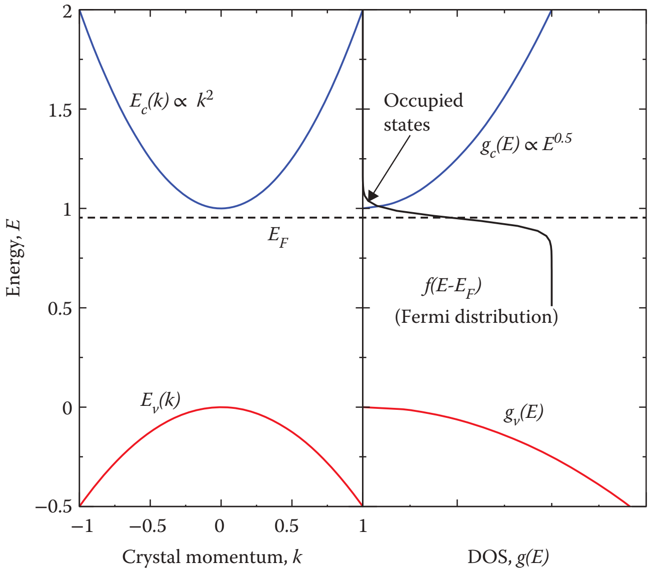
$$D_\mathrm{n} = \frac{k_\mathrm{B}T}{q}\mu_\mathrm{n}, \quad D_\mathrm{p} = \frac{k_\mathrm{B}T}{q}\mu_\mathrm{p} \tag{43.38}$$

which in actual implementations of the drift-diffusion model is usually used to obtain the diffusion coefficients from the mobilities.

The expressions (Equation 43.37) provide an adequate description of the carrier densities for most nondegenerate crystalline semiconductors, i.e., at not too high carrier densities. This is because the same expressions are obtained when calculating the carrier density from the DOS and the Fermi–Dirac distribution, and making the Boltzmann approximation for small densities. Generally, the density, e.g., of electrons is found under the assumption of a local equilibrium distribution by occupying the available states, given by the DOS, with a local quasi Fermi level and integrating over all energies, i.e.,

$$n(\mathbf{r}) = \int g(E) f\left(\frac{E - E_\mathrm{F,n}(\mathbf{r})}{k_\mathrm{B}T}\right) \mathrm{d}E \tag{43.39}$$

where $f(x)$ is the Fermi–Dirac distribution and $g(E)$ the DOS. The DOS of a crystalline semiconductor at the relevant energies, i.e., near the band extrema, is obtained from the energy dispersion, which is approximately parabolic in this range, as shown schematically in Figure 43.6. Details on the calculation of the carrier densities can be found, e.g., in [42,71,74,77].



$$E(k) = E_c^{(0)} + \frac{\hbar^2 k^2}{2m^*}, \qquad f(x) = \frac{1}{1+e^x}$$

$$g_c(E) = \frac{1}{\pi^2}\left(\frac{2m^*}{\hbar^2}\right)^{3/2}\sqrt{E}$$

$$\begin{aligned} n &= \frac{1}{(2\pi)^3}\int f\left(\frac{E_c(k)-E_F}{k_\mathrm{B}T}\right) \mathrm{d}k^3 \\ &= \int_{E_c}^{\infty} g_c(E) f\left(\frac{E-E_F}{k_\mathrm{B}T}\right) \mathrm{d}E \\ &= \underbrace{2\left(\frac{m^*kT}{2\pi\hbar^2}\right)^{3/2}}_{} F_{1/2}(E_F/k_\mathrm{B}T) \\ &\approx \quad N_c \quad \cdot \quad e^{(E_F-E_c)/k_\mathrm{B}T} \end{aligned}$$

FIGURE 43.6 Parabolic approximation to the energy dispersion $E(k)$ of a simple semiconductor and the DOS resulting from this dispersion, obtained by setting $1/(2\pi)^3 \int \mathrm{d}k^3 = \int g(E)\mathrm{d}E$. On the right, some of the expressions are given, and the calculation of the density is sketched ($m^* = \hbar^2(\partial^2 E(k)/\partial k^2)^{-1}$ is the effective mass, E_F is the quasi Fermi level).

The key point is that the derivations are based on the assumption that the carriers are moving freely in a periodic potential of the underlying crystal lattice, which allows to define the momentum k of the carriers and which leads to the parabolic dispersion valid at least in the vicinity of the band extrema. While this is a good approximation for crystalline semiconductors, it does not hold in many materials of interest in thin-film PV technology. In particular, whenever the periodicity of the crystal structure is lost, the DOS will become substantially different. In an amorphous semiconductor such as a-Si, e.g., the crystalline lattice structure is maintained only locally and only approximately: a-Si is mostly tetrahedrally bonded, but with a statistical distribution of bond angles and with certain number of silicon atoms with three or five neighbors. This modifies the DOS by washing out the band edge and appearance of approximately exponential band tails, known as Urbach tails [78], and appearance of localized states in the gap even in absence of impurity atoms, as shown schematically in Figure 43.7b. Similarly, organic semiconductors are usually characterized by a certain energetic disorder in the molecular energy levels, and the definition of carrier momentum is largely meaningless. In this case, the DOS is completely different from the “standard” semiconductor bulk DOS, and is often well described by a Gaussian distribution with a certain width σ characterizing the statistical scattering of the molecular levels [79], shown in Figure 43.7c.

A consequence of a DOS, which differs from that resulting from a parabolic dispersion, is that the Einstein relation relating the diffusion constants and the mobilities needs to be generalized. This general form can be easily found from Equations 43.31 through 43.33 as

$$D_\mathrm{n} = n\left(\frac{\partial n}{\partial \varphi}\right)^{-1}\mu_\mathrm{n}, \quad D_\mathrm{p} = -p\left(\frac{\partial p}{\partial \varphi}\right)^{-1}\mu_\mathrm{p} \tag{43.40}$$

An accurate description of semiconductor transport therefore requires to take the functional form of the DOS into account, although it does not appear explicitly in the basic semiconductor equations. Beside the formal modification of the Einstein relation, the DOS also influences some directly measurable quantities.

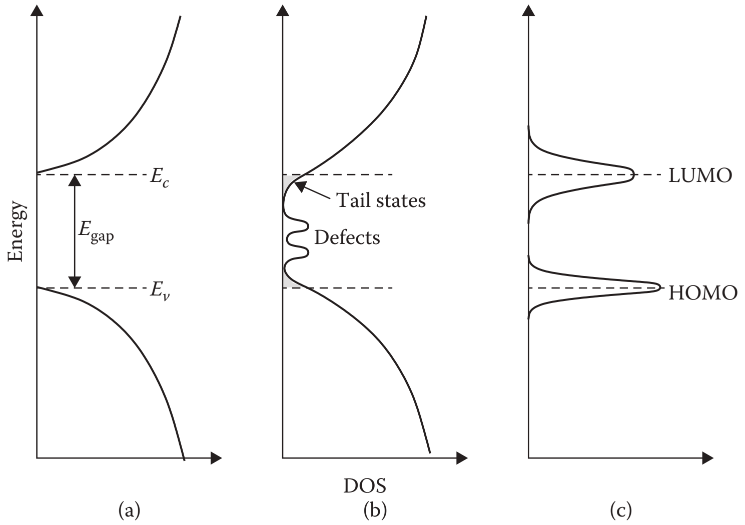


FIGURE 43.7 Sketch of different density of states (DOS): (a) crystalline semiconductor with parabolic band dispersion, (b) amorphous semiconductor with band tails and defect states, (c) disordered semiconductor, e.g., organic, with the DOS formed from energetic and spatial disorder of molecular orbital energies. In the last case, the bands are usually denoted as HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital).

Assume, e.g., a bulk absorber in flat band condition subject to homogeneous optical carrier generation, so that there is no current flow, and assume that only radiative recombination is relevant. In this case, all generated electron–hole pairs have to recombine radiatively according to Equation 43.35, i.e., $G = Bnp$. The splitting of the electron and hole quasi Fermi levels induced by the illumination is a measure for the open-circuit voltage V_{oc} that can be expected. Using Equation 43.37, an analytic formula can be easily found as [80]

$$V_{\text{oc}} = \frac{1}{q}(E_{\text{F,n}} - E_{\text{F,p}}) = \frac{1}{q}E_g + \frac{k_{\text{B}}T}{q}\ln\left(\frac{G}{BN_{\text{c}}N_{\text{v}}}\right) \tag{43.41}$$

and apparently the V_{oc} should vary by a constant rate of $2.3k_{\text{B}}T/q$ per decade in intensity. A different result is obtained if the same quantity is calculated using a different expression for the DOS, e.g., using a Gaussian DOS (GDOS) as typical for organic semiconductors and given by

$$g(E) = \frac{N_0}{\sigma\sqrt{2\pi}}\exp\left(-\frac{1}{2}\frac{(E-E_0)^2}{2\sigma^2}\right) \tag{43.42}$$

Here N_0 is the total density, σ the width (variance) of the Gaussian, and E_0 the center of Gaussian. Using approximations of the formulas for the densities [81], and assuming the same parameters for the electron and hole DOS, we can obtain the open-circuit voltage for sufficiently large carrier densities in this case as

$$V_{\text{oc}} = \frac{1}{q}E_{\text{g}} - 2\frac{\sigma\sqrt{2}}{qH(\sigma)}\text{erfc}^{-1}\left(2\sqrt{\frac{G}{BN_0^2}}\right) \tag{43.43}$$

Here, E_{g} is defined as the difference of the center energies of the electron and hole DOS, and $H(\sigma) = k_{\text{B}}T\sqrt{2}/\sigma \cdot \text{erfc}^{-1}(\exp(-\frac{1}{2}\sigma^2/(k_{\text{B}}T)^2))$. Figure 43.8 compares the derivative of V_{oc} with respect to the logarithm of the generation rate for a parabolic DOS and a GDOS, showing that for typical Gaussian widths

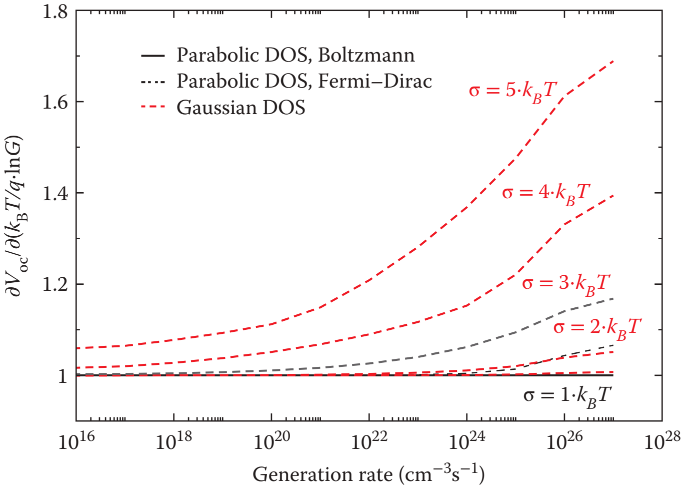


FIGURE 43.8 Variation of the bulk quasi Fermi level splitting due to a homogeneous generation rate. Deviation from the exponential form of the carrier density leads to considerable deviation from the expected rate of change. For the Gaussian density of states (DOS), $N_0 = 10^{21}$ cm^{-3} has been used, for the parabolic case $N_c = N_v = 10^{19}$ cm^{-3}. The recombination rate constant was $B = 10^{-10}$ cm^3·s^{-1}.

the DOS influences the rate even at moderate generation rates. The effect becomes stronger with increasing intensity. The measurable effect of energetic disorder is described, e.g., in [82].

Note that a further modification of the form (Equations 43.31 through 43.33) of the drift-diffusion equations is needed whenever the material parameters in the formula for the carrier densities (i.e., effective DOS or the band edges, see Figure 43.6), or generally the DOS, change spatially. This can be seen easily when evaluating the gradient of the densities, which then leads to additional terms that can be interpreted as effective potentials. These corrections are simple in the case of the standard exponential form (Equation 43.37) of the carrier densities, but they become more complicated in the degenerate case and for other DOS. It is therefore sometimes preferred to write the drift-diffusion model in a different but equivalent form, derived from thermodynamics [61,83]:

$$\nabla(\varepsilon\nabla\varphi) = q(n - p - C) \tag{43.44}$$

$$\frac{\partial n}{\partial t} - \frac{1}{q}\nabla J_\mathrm{n} = -R, \qquad J_\mathrm{n} = -q\mu_\mathrm{n} n\nabla\phi_\mathrm{n} \tag{43.45}$$

$$\frac{\partial p}{\partial t} + \frac{1}{q}\nabla J_\mathrm{p} = -R, \qquad J_\mathrm{p} = -q\mu_\mathrm{p} p\nabla\phi_\mathrm{p} \tag{43.46}$$

Here, $\phi_{\{n,p\}}$ are the electrochemical potentials of the electrons and holes, respectively, related to the quasi Fermi levels by $E_{F,\{n,p\}} = -q\phi_{\{n,p\}}$. While in Equations 43.31 through 43.33, the dependent variables are the electrostatic potential and the carrier densities, in (Equations 43.44 through 43.46) they are the electrostatic and the electrochemical potentials, and the particle currents are driven by the gradient of the electrochemical potential. The advantage of this formulation with respect to the one in Equation 43.31 through 43.33 is that it is independent on the functional form of the density, and the Einstein relation is not needed explicitly. For the rest, it has the same limitations of applicability as Equations 43.31 through 43.33.

From a technical point of view, note that the decoupled carrier transport equations in Equations 43.44 through 43.46 are quasilinear, with the differential operator being proportional to the carrier density, while in Equations 43.31 through 43.33 the transport equations are linear. This makes the numerical solution procedure more involved.

If we are using the thermodynamic formulation (Equations 43.44 through 43.46) for simulation of solar cells, then we need an adequate model for the DOS or, respectively, the carrier density for all the materials involved, and we need models for the carrier mobilities and the recombination rates. Especially, the mobility is a critical parameter. In a crystalline semiconductor, the mobility is related to the scattering the carriers are subject to. So it depends strongly on temperature and defect or doping densities, and usually mobility increases with decreasing temperature and decreasing defect densities, since the associated scattering rates decrease. Mobility also depends on the electric field, typically by decreasing with increasing field due to a saturation of the mean drift velocity. In fact, the proportionality between drift velocity and electric field, given by $v = \mu_n E$, holds only at low field strength [84]. In noncrystalline materials such as amorphous or organic semiconductors, these dependencies can be radically different. For example, mobility may increase with temperature because carriers can escape more easily local energy minima or overcome small energy barriers. Or it may increase with electric field because the field reduces the effective energy barriers hindering carrier flow. In some cases, such behavior is a manifestation of an underlying microscopic transport that deviates from a diffusive one.

In disordered organic materials, e.g., the deviation from a periodic crystalline structure makes the concept of band-transport inapplicable. Instead, charges are spatially localized and charge transport occurs by transitions between localized states, a process termed as *hopping*. In reality, disordered systems show both disordered and (semi-)crystalline regions so that charge transport is always an interplay between hopping and band conduction. Higher contributions of hopping transport usually lead to lower charge mobilities. Localized states can be imagined as local energy valleys in which charges get confined. Transport can either take place by surmounting the energy barrier to a neighboring localized state by absorbing phonons, or by tunneling through the barrier. The concept of phonon-assisted hopping was introduced by Mott [85]. In 1960, Miller and Abrahams described a hopping model between two localized states consisting of both a contribution due to thermal activation and tunneling. After their model, the hopping process between the states i and j with energy difference $\Delta E = E_j - E_i$ and distance r_{ij} is of the form [86]

$$a_{i\to j} = a_0 \exp\left(-2\gamma r_{ij}\right) \begin{cases} \exp\left(-\frac{\Delta E}{k_B T}\right) & \text{if } \Delta E > 0 \\ 1 & \text{if } \Delta E \leq 0 \end{cases} \tag{43.47}$$

where a_0 is given by a typical phonon frequency in the system and γ is the inverse localization constant. The formula combines the contribution of two different processes to hopping: the first exponential describes a tunneling contribution through a barrier given by the distance r_{ij} of the localized states, and the latter term stands for the thermally activated contribution with the relation between barrier height ΔE_{ij} and the thermal energy $k_B T$.

Another model that takes into account strong coupling between charges and lattice is the polaron model. In organic materials with low permittivities, a charge on a molecule can considerably deform the structure of the molecule. The complex of charge and the distortion of the lattice it induces is known as a quasiparticle called *polaron*. Marcus [87] developed a model for the transition of polarons:

$$a_{i\to j} = \frac{\left|J_{ij}\right|^2}{\hbar} \sqrt{\frac{\pi}{\lambda k_B T}} \exp\left(-\frac{(\Delta G_{ij} - \lambda)^2}{4\lambda k_B T}\right) \tag{43.48}$$

where J_{ij} is the transfer integral for electron and hole transfer, ΔG_{ij} is the free energy difference between state, i and j and λ the reorganization energy, which accounts for the reconfiguration of the molecule due to the deformation by the charge.

In order to derive a mobility model for disordered materials, a description for the distribution of available states must be available. The spatial irregularity of these systems leads to fluctuations in the intermolecular interactions. A successful model to represent these energetic fluctuations is the GDOS as given in

Equation 43.42, in which the energy distribution of the localized states is spread by a Gaussian distribution function around the HOMO/LUMO energy levels of the organic material. The GDOS in combination with the Miller–Abrahams hopping was applied by Bässler [79] by Monte Carlo simulations to study the effect of temperature and electric field on the charge mobility. He found a non-Arrhenius temperature dependence $\mu(T) = \mu_0 \exp\left(-(t\hat{\sigma})^2\right)$ with $\hat{\sigma} = \sigma/k_B T$, and, Poole–Frenkel type behavior $\mu(F) = \mu_0 \exp\left(\gamma\sqrt{F}\right)$ for the dependence on the electric field F for a limited range of electric fields. To include the influence of charge carrier densities on the mobility, Pasveer et al. [88] parametrized the results based on a direct solution of the Master equation in a system with a GDOS. They found as expression for the hopping mobility:

$$\mu(T, \rho, F) = \mu(T, \rho) f(T, F) \text{ with} \tag{43.49}$$

$$\mu(T, \rho) = \mu_0 c_1 \exp\left[-c_2\left(\frac{\sigma}{k_B T}\right)^2\right] \exp\left[\frac{1}{2k_B T}\left(\frac{\sigma^2}{k_B T} - \sigma\right)(2\rho a^3)^\delta\right] \tag{43.50}$$

$$f(T, F) = \exp\left\{0.44\left[\left(\frac{\sigma}{k_B T}\right)^{\frac{3}{2}} - 2.2\right]\left[\sqrt{1 + 0.8\left(\frac{eFa}{\sigma}\right)^2} - 1\right]\right\} \tag{43.51}$$

where ρ is the charge carrier density, either electrons or holes, F is the electric field, $k_B T$ the thermal energy, $a = N_0^{-\frac{1}{3}}$ the distance between localized states, σ the energetic disorder in the GDOS, $c_1 = 1.8 \cdot 10^{-9}$, $c_2 = 0.42$, $\mu_0 = a^2 \nu_0 e/\sigma$, with ν_0 the prefactor in the Miller–Abrahams formula, and

$$\delta = 2\frac{\log(s^2 - s) - \log(\log(4))}{s^2} \text{ with } s = \frac{\sigma}{k_B T} \tag{43.52}$$

This parametrization can be used in drift-diffusion models to account for the hopping nature of transport.

43.4.2 Kinetic Monte Carlo

In this section, we give an overview of the kinetic Monte Carlo (kMC) method [89,90] and how it can be applied to set up a model for organic solar cells.

Monte Carlo methods offer a simple yet powerful tool to solve real world problems by the use of random numbers. In the hierarchy of transport models shown in Figure 43.9, kMC bridges the gap between macroscopic models such as drift-diffusion and molecular dynamics (MD) and quantum mechanical models.

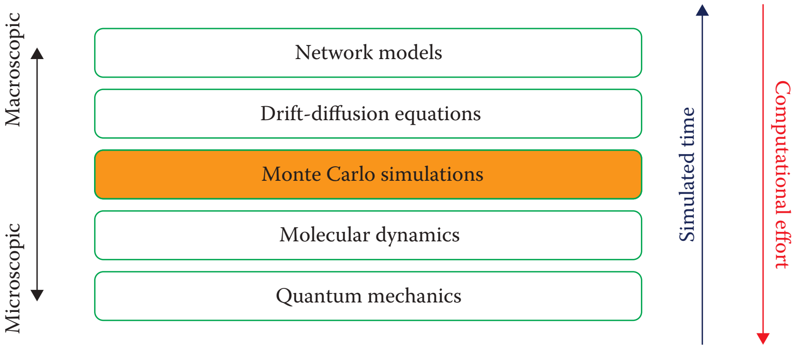


FIGURE 43.9 Hierarchy of transport models. Kinetic Monte Carlo simulations bridge the gap between macroscopic and microscopic modeling techniques.

The idea for the use of random numbers in an algorithm came from Fermi [91] and was further pursued by Ulam and Metropolis in the 1940s studying neutron diffusion in fissionable materials. They developed the most used method of this kind, the Metropolis algorithm [92], during a study about the calculation of state equations for a system of interacting particles described by a hard sphere cutoff potential.

In the following decades, Monte Carlo methods have proven to be a versatile tool and much progress in the development of calculations for a variety of areas, ranging from surface adsorption processes, radiation damage annealing, and statistical physics [93–95], has been achieved. Especially, the development of a Monte Carlo algorithm class that models physical systems as dynamically evolving from state to state by a certain set of transition rates obtained large interest as it allowed to study time-dependent properties of physical systems. This class of algorithms is today commonly known as *kinetic Monte Carlo*. It has received great benefit in the growing computational power of modern processors and provides a flexible tool to simulate dynamical behavior of physical systems, e.g., transport processes in inorganic semiconductor devices [96]. Major contributions to the modern kMC algorithm, as it is described here, have been made by Bortz et al. in 1975 [95] and Gillespie in 1976 [89] and 1977 [97].

Compared to MD simulations, which include the interactions of single atoms with all dynamics of the molecules like vibrations, occurring on very small time scales (fs), kMC characterizes the state of a system by a set of more macroscopic variables which do not include underlying dynamics like fast vibrational modes. These motions on smaller time scales can be ignored as long as the current state can still be assigned to a long-time state. The dynamical evolution of the system is only seen by transitions between such long-time states, called events. Such a system is called an infrequent-event system [90]. Figure 43.10 illustrates this concept. By ignoring the underlying fast motions and only considering the transitions between long-time states, a sufficient amount of processes can be neglected. Thus, much larger overall simulation times of the order of milliseconds to seconds can be reached.

The above mentioned concept can be applied in particular to organic solids, where particles such as electrons, holes and excitons reside in localized states and transport is determined by hopping from one localized state to another. The time scale of these hopping events is much longer than other processes, which do not change the spatial localization of the carriers.

Compared to drift-diffusion, kMC has the advantages of being able to capture the discrete nature of the problem by modeling individual particles at their molecular positions, and of providing a physically more accurate description of hopping transport without having to resort to effective continuous media models. Moreover, it can be much easier and physically more sound to treat blended materials with nanoscale morphology in kMC.

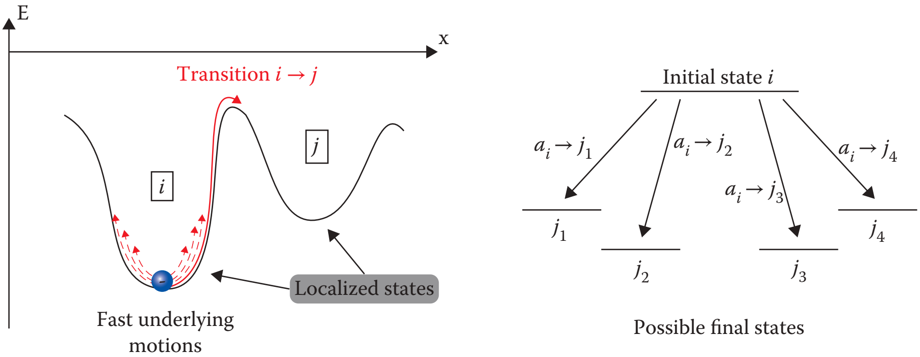


FIGURE 43.10 Transitions from long-time state to long-time state with underlying fast modes in an infrequent-event system (left). State-to-state transitions reduced to the needs for kinetic Monte Carlo (right).

43.4.2.1 kMC Algorithm

The essential concept of the kMC method is the characterization of the considered system by a set of system-specific long-time states. A state is, e.g., defined by the position of all particles in the system at a certain time. Under dynamical evolution, it is possible for the system to change its state by certain system-specific transitions, thus motion of particles from one state to another. The task for an accurate model is to identify all the important processes that take place in the system. By considering processes that happen on long time scales and neglecting those which act on very small time scales, one is able to accomplish a tradeoff between accuracy and sufficiently long simulation times. It is of importance to carefully choose the implemented processes in order not to lose essential parts of the system dynamics.

Let the system be in a certain state i. Depending on the current system setup, a set of several different states j exists to which a transition can take place. This is indicated in Figure 43.10 (right). The transitions from beginning states i to possible end states j can be characterized by rate constants a_{ij} given in units of s^{-1}. Rate constants are a measure for the probability of the corresponding transition from $i \rightarrow j$. An important property must be that the rates are memoryless, meaning that they are only dependent on the setup of the current state and are not dependent on how the system evolved into it. This property is known from the theory of Markov chains [98] and allows, if the set of rate constants is known, to determine the dynamical evolution of the system.

As known from statistical physics [99], stochastical processes characterized by discrete transitions from state to state under temporal evolution can phenomenologically be described by the master equation

$$\frac{\mathrm{d}p_i(t)}{\mathrm{d}t} = -\sum_j a_{ij}\, p_i(t) + \sum_j a_{ji}\, p_j(t) \tag{43.53}$$

The probability $p_i(t)$ to find the system in state i at a certain time t changes in a way that is determined by the rate constants a_{ij} and the probabilities of only the initial state $p_i(t)$ and the possible final states $p_j(t)$. Finding solutions of the master equation, be it by analytical or numerical methods, is often complex and not feasible. The kMC method is based on a stochastical framework and provides a numerical approach to obtain the dynamic system evolution based on the time-dependent propagations from state to state. By choosing a transition pathway through a chain of subsequent states, one possible dynamic system evolution is obtained. Such a walk through a pathway of transitions forms a Markov chain. Averaging over a large number of such Markov chains provides an equivalent system behavior as described by the master equation. Therefore, the kMC algorithm is essentially emulating the master equation.

For a detailed derivation of the associated mathematical framework, the reader is referred to the original work of Gillespie [89]. In the following, we present only the equations necessary for a possible implementation of the algorithm. The algorithm as proposed by Gillespie relies on the ability to characterize the system events by a set of transitions

$$\{\mu\} \text{ with } \mu = 1, 2, \ldots, M \tag{43.54}$$

where M is the total number of transitions. μ is equal to a transition from state i to one of the possible states j in that configuration: $i \rightarrow j_\mu$. Then a_μ is the rate value describing process μ and $\{ a_\mu \}$ is the corresponding set of rates. All transitions must describe distinct processes and their number M can vary over time, depending on which events are enabled at a certain step in the system evolution. It is essential to consider all major processes in order to be able to accurately reproduce the behavior of the system. The main task is now to develop a procedure to simulate the time-dependent evolution of the system under consideration of the given transitions. This makes it necessary to derive an algorithm that is able to choose which transition will be carried out and (for the time dependency) at what time span, all based on the drawing of random numbers.

At any given time, the set of $\{a_\mu\}$ characterizes the possible transitions. They are functions of the current system setup and will change over time, as described later.

Instead of finding a solution to the master equation, the kMC method is based on a function called the probability density function (PDF). The PDF is the suitable function to select a transition and its corresponding time step based on random numbers. The PDF is defined as

$$\begin{aligned} P(\tau,\mu)\,d\tau = {} & \text{probability at time } t \text{ for the next transition} \\ & \text{to occur in time interval } (t+\tau\,,\; t+\tau+d\tau) \\ & \textit{and} \text{ being a transition of type } R_\mu \end{aligned} \tag{43.55}$$

where $P(\tau,\mu)$ is the PDF. This function can be interpreted as follows: it is a joint PDF of the continuous time variable τ $(0 \leq \tau < \infty)$ and the integer variable μ $(\mu = 1, 2, \ldots, M)$ characterizing the transition. At a certain time t, the PDF represents the probability for transition μ to occur within an infinitesimal time span of $d\tau$ after no other event has taken place for a time interval τ. Hence, $P(\tau,\mu)$ can be seen as the independent probabilities to have no transition in the time interval $(t\,,\; t+\tau)$ *and* to have a transition of type R_μ immediately afterward in the infinitesimal interval $(t+\tau\,,\; t+\tau+d\tau)$. In common terms, the transition μ needs time τ before the next transition occurs.

Using the abbreviation

$$a = \sum_{\nu=1}^{M} a_\nu \tag{43.56}$$

as the total transition probability per unit time, the total expression for the PDF is given by

$$P(\tau,\mu) = a_\mu \cdot \exp\left(-\sum_{\nu=1}^{M} a_\nu \tau \right) = a_\mu \exp(-a\tau) \tag{43.57}$$

This function, only dependent on the set of events a_μ, is the starting point to choose a random number pair (τ,μ) that characterizes the next transition and time step in the dynamic evolution of the system. The PDF is properly normalized, meaning that every pair (τ,μ) with $\tau \in (0 \leq \tau < \infty)$ and $\mu = 1, 2, \ldots, M$ has its weighted contribution to the PDF and that it is guaranteed to obtain an associated value for the PDF by randomly choosing one pair:

$$\int_{t=0}^{\infty} d\tau \sum_{\nu=1}^{M} P(\tau,\nu) = \sum_{\nu=1}^{M} a_\nu \int_{t=0}^{\infty} d\tau \exp(-a\tau) = 1 \tag{43.58}$$

Now, the fundamental Monte Carlo step can be defined: by choosing a random pair of the variables (τ,μ) according to the PDF $P(\tau,\mu)$ after Equation 43.57, a weighting based on the magnitude of the various transitions and the exponential decrease in time is performed. For the execution of this step, two approaches were suggested by Gillespie: the *direct method* and the *first reaction method*. In this context, the former method is used. The direct method is most efficient in the way it utilizes random numbers to determine a pair of transition and a time step, while the first reaction method can save computational power at the cost of accuracy.

In the direct method, the PDF is divided into two separate PDFs, each of which is itself normalized and only dependent on either the time τ or the transition μ:

$$P(\tau, \mu) = P_1(\tau) \cdot P_2(\mu|\tau) \tag{43.59}$$

Here, the first factor represents the probability density that any of the possible transitions takes place in the time interval $(t + \tau, t + \tau + d\tau)$, irrespective of which one. Thus, $P_1(\tau)$ is the sum over all transitions. With Equation 43.57 follows

$$P_1(\tau) = \sum_{\nu=1}^{M} P(\tau, \nu) = \sum_{\nu=1}^{M} a_\nu \exp(-a\tau) = a \cdot \exp(-a\tau) \tag{43.60}$$

From this and Equation 43.59, $P_2(\mu|\tau)$ is easily derived as

$$P_2(\mu|\tau) = \frac{P(\tau, \mu)}{P_1(\tau)} = \frac{a_\mu}{a} \tag{43.61}$$

and $P_2(\mu|\tau)\, d\tau$ represents the probability that a reaction μ occurs, given the condition that a time step of τ has been chosen before. Both P_1 and P_2 are properly normalized PDFs on their own.

A pair of random numbers (τ, μ) according to the distributions (Equations 43.60 and 43.61) can be produced as follows, using (pseudo-)random number generator with uniform distribution available in all modern programming languages. Every normalized PDF $P(x)$ has an associated cumulative distribution function, defined by

$$F_1(x) = \int_{-\infty}^{x} P_1(x')\, dx' \qquad \text{for a continuous } P_1(x) \tag{43.62}$$

$$F_2(i) = \sum_{i'=1}^{i} P_2(i') \qquad \text{for a discrete } P_2(i) \tag{43.63}$$

Because $\lim_{x\to -\infty} F_1(x) = 0$ and $\lim_{x\to\infty} F_1(x) = 1$ for the continuous case, and $F_2(0) = 0$ and $F_2(M) = 1$ for the discrete case, both functions are monotone and limited. Thus, there exists an inverse function F^{-1} for each of them. It is made sure that every value from the output range has a corresponding value x or i in the domain range. The concept of the inversion method is to draw two uniform random numbers $r_1, r_2 \in (0, 1)$, and calculate $x = F_1^{-1}(r)$ and i from $F_2^{-1}(i)$. This way the obtained values for x and i are distributed according to the PDF from which F is derived.

Using this inversion method and the PDF $P_1(\tau) = a \cdot \exp(-a\tau)$, the corresponding cumulative probability function is[†]

$$F_1(\tau) = \int_{0}^{\tau} a \cdot \exp(-a\tau')\, d\tau' = 1 - \exp(-a\tau) \tag{43.64}$$

The continuous and exponentially distributed time variable τ, which describes the time passed before one transition (or, for a more vivid interpretation: the time "needed" for one transition), is then

---

[†] Note that $P_1(t)$ is only defined for $t \geq 0$.

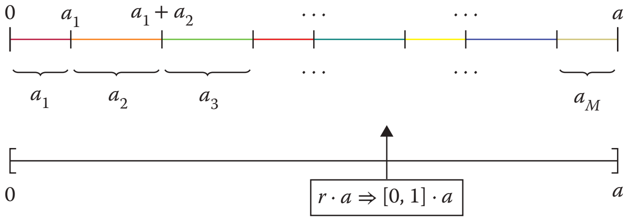


FIGURE 43.11 Drawing a discrete transition by a continuous random variable $r \in [0, 1]$. All enabled rate constants are added to yield the total rate a. Multiplication of a random number in the unit interval with the total rate corresponds to one specific rate process, weighted by its magnitude. Rates are added until the next addition will exceed $r \cdot a$. The last rate in the addition is chosen.

given by the inverse function of F_1. Hence,[†]

$$\tau = F_1^{-1}(r_1) = \frac{1}{a} \ln\left(\frac{1}{r_1}\right) \qquad \text{with } a = \sum_{\nu=1}^{M} a_\nu \text{ and } r_1 \in (0, 1) \tag{43.65}$$

To choose a transition μ from the set of events $\{a_\mu\}$ $(\mu = 1, 2, \ldots, M)$ by a random number r_2, the cumulative probability must satisfy

$$F_2(\mu - 1) < r_2 \leq F_2(\mu) \tag{43.66}$$

That is, $r_2 \in (0, 1)$ corresponds to one plateau in the discrete histogram of F_2. Substituting Equation 43.61 in Equation 43.63 yields an expression for $F_2(\tau|\mu)$:

$$F_2(\tau|\mu) = \sum_{\nu=1}^{\mu} \frac{a_\nu}{a} \tag{43.67}$$

With this, the abovementioned condition reads

$$\sum_{\nu=1}^{\mu-1} \frac{a_\nu}{a} < r_2 \leq \sum_{\nu=1}^{\mu} \frac{a_\nu}{a} \tag{43.68}$$

or, multiplying by a,

$$\sum_{\nu=1}^{\mu-1} a_\nu < r_2 \cdot a \leq \sum_{\nu=1}^{\mu} a_\nu \quad \text{with } a = \sum_{\nu=1}^{M} a_\nu \text{ and } r_2 \in (0, 1) \tag{43.69}$$

This represents a sum of all transition probabilities, up to an index for which the sum is larger than the random variable r_2. This index is set to be the chosen transition μ. The process to derive μ computationally from a random number is illustrated in Figure 43.11.

The essence of the Monte Carlo method is condensed in the two equations (Equations 43.65 and 43.69) to calculate a pair of random numbers (τ, μ), while the rest of the method is fairly straightforward and can be reduced to only the following few, simple steps, shown in a flowchart in Figure 43.12:

[‡] The term $1 - r$ has been replaced by r. Since we are dealing with probabilities, this is valid for uniform random numbers in the unit interval.

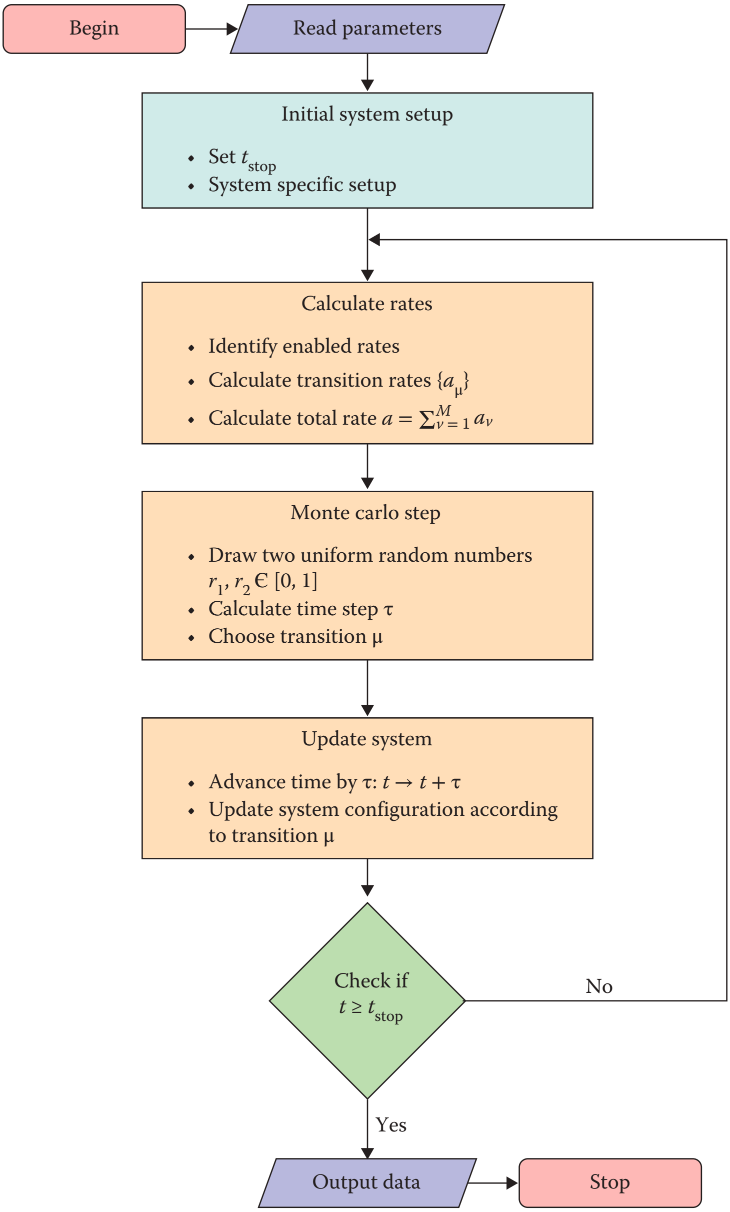

Begin
Read parameters
Initial system setup
Set t_{stop}
System specific setup
Calculate rates
Identify enabled rates
Calculate transition rates $\{a_\mu\}$
Calculate total rate $a = \sum_{\nu=1}^{M} a_\nu$
Monte carlo step
Draw two uniform random numbers $r_1, r_2 \in [0, 1]$
Calculate time step τ
Choose transition μ
Update system
Advance time by τ: $t \rightarrow t + \tau$
Update system configuration according to transition μ
Check if $t \geq t_{stop}$
No
Yes
Output data
Stop

FIGURE 43.12 Flowchart of a general kinetic Monte Carlo algorithm.

0 – Initialization

In the beginning, system-specific parameters are defined in the program. This includes all parameters that must be known in order to set up the system and to calculate the transition rates. Variables to store the occurrence of relevant processes during the simulation are set up here. The simulation time t_{stop} is defined. It determines how long the dynamic evolution should be simulated. The starting time is set to $t = 0$.

1 – Calculation of transition rates

The kMC method is reliant on physical models for all processes that can occur in the system. All physics enters the model in terms of the event rates that are calculated in this step. Depending on the current state, only some transitions may be activated. Conditions must be met in order to make sure that rates for processes that cannot occur in a certain system state are deactivated. This is realized by setting the value for a deactivated rate $a_\mu = 0$ so that it is not considered in the selection a transition after Equation 43.69. After all rates have been calculated and stored, they are summed up to provide the total rate a.

2 – Monte Carlo step

As described in Section 43.4.2.1, a random pair (τ, μ) is generated according to the PDF $P(\tau, \mu)$ from the list of available (and activated) transitions $\left\{a_\mu\right\}$. This accounts for the stochastically correct time step τ for the selection of a transition μ.

3 – Update system

To update the system the time t is advanced by τ. Based on which transition μ has been chosen, the system configuration needs to be updated accordingly. How every update is specifically depends on the physical process.

After updating the system, a check is performed whether the simulation time has reached the specified stopping time t_{stop}. If $t < t_{\text{stop}}$, the simulation will continue by jumping back to step 1 and recalculate the event rates. This way, the system update on the behavior of other processes is reflected. The steps 1–3 represent the so called *Monte Carlo loop*. They will be executed as long as termination condition is not yet fulfilled.

4 – Termination and data output

Through the successive time advance t eventually becomes larger than the stopping time and the Monte Carlo loop is terminated. At last the variables tracking selected events can be evaluated.

Overall, the kMC method offers a straightforward algorithm to simulate the time-dependent evolution of a system. It is only reliant on (1) the assumption that all physical processes can be determined by rate expressions, and (2) that a (pseudo-)random number generator with a sufficiently small correlation between two subsequent numbers and a large period is provided.

43.4.3 Example: Transport through 1D Disordered System by kMC

To illustrate an actual implementation of the kMC algorithm, we provide a short manual on how to model charge transport in a 1D, energetically disordered system, e.g., an organic polymer. We will go through the steps presented in the previous section and in Figure 43.12: the initial system setup, the simulation loop, and the evaluation of quantities.

43.4.3.1 System Setup and Starting Conditions

In the beginning, the available states that the system can occupy must be defined. The setup is illustrated in Figure 43.14. Here, we consider a linear chain of N equidistant nodes indexed by $i = 1, \dots, N$. The position of each node is quantified by $x_i = i \cdot L$ where L is the fixed distance between all nodes. Then, energy values are assigned to each node accounting for the influence of an applied electric field F, which is assumed to be constant leading to a linear energy drop across the system, and an energetic fluctuation $r(\mu, \sigma)$ drawn from a GDOS distribution with mean μ and standard deviation σ as given in Equation 43.42. This makes up for an energy landscape of the form $E_i = qFx_i + r_{\text{GDOS}}(\mu = 0, \sigma)$. The disorder is picked independently of the position. A typical configuration for E_i is depicted in Figure 43.14.

Next, initial conditions must be set. The simulation time is set to an initial time $t = 0$. It will be updated during each simulation step. The termination condition is defined by the particle position: if the particle reaches the last node N, the simulation loop stops. We allow one electron of charge $q = -e$ at a time to be in the system. In the beginning it is set to the first node x_1.

A model for the transition rates needs to be defined. We allow the particle to move only to its immediately neighboring sites to the left or to the right. The transitions are described by hopping rates according to the Miller–Abrahams formula (Equation 43.49). There are then only two distinct rates: a_l for hopping to the left and a_r for hopping to the right (Figure 43.13). They depend on the relative energy difference $\Delta E = E_j - E_i$ of the current node i of the particle and the final node $i \pm 1$. Typical values are $a_0 = 10^{11}\ \mathrm{s}^{-1}$, $\gamma = 2\mathrm{nm}$, and $r_{ij} = L = 1$ nm.

43.4.3.2 Simulation Loop

After system setup, the simulation loop is executed until the termination condition is reached. It consists of the following steps:

1. Check the energy landscape to the right and to the left of the current particle position $E_{i\pm1}$.
2. Calculate the hopping rates $a_\mathrm{l}(E_{i-1} - E_i)$ and $a_\mathrm{r}(E_{i+1} - E_i)$.
3. Pick two distinct random numbers r_1, r_2 uniformly from the interval (0, 1).
4. Calculate the total rate $a = a_\mathrm{l} + a_\mathrm{r}$. Then, determine the transition (hopping left or right): If $r_1 \cdot a \leq a_l$, pick a_l. Otherwise, if $a_\mathrm{l} < r_1 \cdot a \leq a_\mathrm{l} + a_\mathrm{r} = a$, pick a_r.
5. Calculate timestep: $\tau = -\frac{r_2}{a}$.
6. Update system according to the transition picked, i.e., change the particle position to the node on the right or on the left.
7. Advance time $t \rightarrow t + \tau$.
8. Check termination condition. If still true, go to step 1. Otherwise, exit loop.

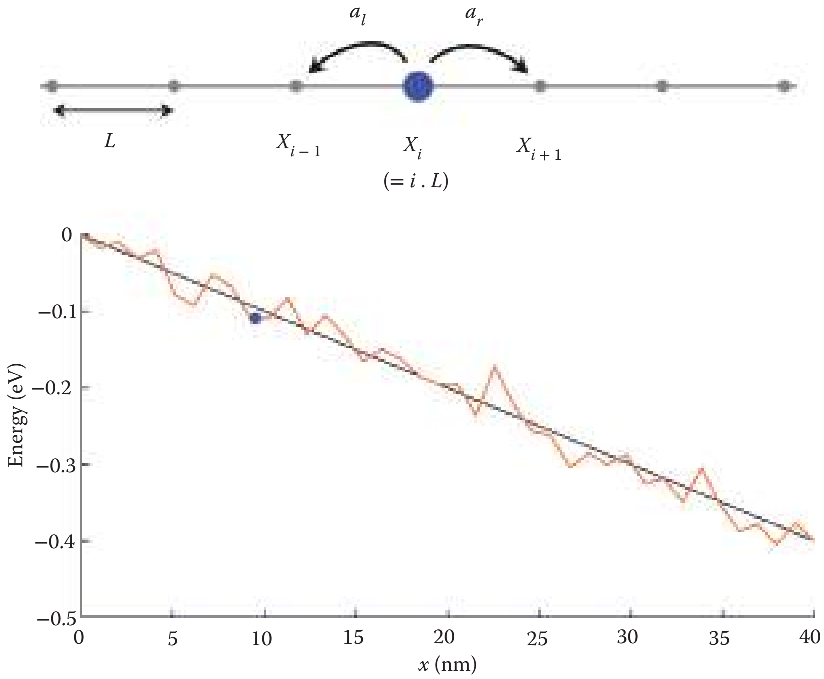


FIGURE 43.13 Setup to model hopping in a 1D system. Top: Spatial discretization of the problem in equidistant nodes. Bottom: Energetic configuration of the nodes. The setup shows the energetic landscape for $N = 40$ nodes comprised of a linear drop of the electric field $F = 10^5$ V/cm and a Gaussian energetic disorder with $\sigma = 30$ meV. The particle is allowed to hop to neighboring nodes under the influence of the energetic configuration.

43.4.3.3 Evaluation

After termination, desired quantities can be evaluated. In this case, since the time that the particle needed to hop through the disordered landscape t_{sim} is now known, the effective velocity $v = \frac{N \cdot L}{t_{sim}}$ can be calculated. From v and the electric field F, the effective transport mobility can be extracted as $\mu = \frac{v}{F}$.

43.4.3.4 Further Info

The general system setup and transition rates can be extended to 2D or 3D (see Figure 43.15). More particle types (holes, excitons) and more effects than hopping, such as injection, recombination, photogeneration of excitons can be introduced to obtain a more realistic model. The principle of choosing transition and timestep stays the same.

43.5 Specific Modeling Issues

In this section, we provide further details on the implementation of models specific to some of the emerging PV technologies, and we cover generic modeling issues.

43.5.1 Dye Solar Cells

DSCs have been extensively simulated using analytical models or numerical ones at the level of drift-diffusion. Most simulations have been performed using 1D models [100–102] and an effective material approximation (see Section 43.5.4 for an extensive description of this approximation). Only few simulations going beyond the effective medium model and trying to include the morphology [103] or to go toward 2D and 3D simulations [104,105] have been reported so far.

The main issues concerning simulations of a DSC are for sure related to the complex morphology of the mesoporous material and the chemical reactions at the interfaces. The latter include both the dye regeneration processes as well as all the recombination events where electrons in the TiO_2 are transferred to reduce ionized dyes or triiodide ions. The main issue in modeling is that all these processes are in general complex electrochemical rates where the charge transfer occurs in several steps. They depend on the effect of the electrolyte, the reorganization energy of the environment, temperature, and voltage [106]. Moreover, different dyes or electrolytes will lead to different kinetics and thus different recombination rates. This leads to recombination rates that formally are completely different from the standard formulas given in Equations 43.34 through 43.36.

Beside that, boundary conditions at the contacts also are very uncommon with respect to more conventional PV devices. While the anode with the ITO or FTO can be approximated by a standard ohmic or Schottky contact, the cathode at the platinum is an electrochemical contact which requires to take into account the formation of an ion accumulation layer. The standard boundary condition is a Butler–Volmer (BV) equation, which is generally a Robin-type boundary condition, since it connects the electron transfer rate with the electrochemical potential at the contact. A BV equation [107] describes the dependency of the electrical current at an electrode j on the electrode potential when both anodic and cathodic reactions are involved:

$$j = j_0 \cdot \left\{ \exp\left[\frac{\alpha_A z e}{k_B T}\eta\right] - \exp\left[\frac{\alpha_C z e}{k_B T}\eta\right] \right\} \tag{43.70}$$

where j_0 is the exchange current density prefactor, $\eta = E - E_{eq}$ the activation overpotential given by the difference in electrode potential and equilibrium potential, α_A and α_C are charge transfer coefficients for anode and cathode, respectively, z is the number of electrons participating in the reaction at the electrode, e

is the elementary charge and $k_B T$ the thermal energy. A Robin-type boundary condition in general is a combination of Dirichlet boundary conditions and Neumann boundary conditions. Imposed on a differential equation of function f, it reads in a general form

$$af + b\frac{\partial f}{\partial n} = g \tag{43.71}$$

where a and b are nonzero constants and g is a function defined on the boundary of the problem. Depending on whether the limiting process is the ionic diffusion of the redox pair or the charge transfer between platinum and triiodide ions, we can have different shapes of the BV equation. Finally, electron transport in the TiO_2 is mediated by trapping and detrapping events, a physical situation captured by the multitrapping model, where the mobility becomes density dependent [108,109]. For solid state DSCs, the simulation issues are more similar to OPVs as the electrolyte is replaced by an organic semiconductor.

43.5.2 Organic Photovoltaics

Modeling of OPV devices incurs several important differences when compared to inorganic devices. First of all, as mentioned already in section 43.4.1, the DOS is fundamentally different as it does not result from electronic states in a periodic potential, but from a both energetically and spatially disordered energy landscape of molecular levels as shown schematically in Figure 43.14. This has to be considered when using or implementing a drift-diffusion based simulation tool, since the standard drift-diffusion model has been derived for carriers with a parabolic DOS. As a further consequence of this energy disorder, transport cannot be described adequately by carriers moving freely in conduction and valence bands, but rather by hopping between energy levels on different molecules [86,110–112], as described in Section 43.4.1. Fortunately, on a macroscopic level, i.e. on a scale much larger than the molecular structure, transport can still be described in terms of drift and diffusion, if a suitable model for the carrier mobilities can be found. Such a model is described in [88], where the mobility has been parametrized based on numerical results of the Master equation.

Due to the different DOSs, the models for transport across interfaces and in particular for injection at the contacts also are different from their counterparts employed in standard drift-diffusion simulations of inorganic semiconductors. The properties of a metal-semiconductor contact, known as Schottky contact in the inorganic case, are controlled by the enetrgetic alignment of the metal work function with the semiconductor band edges, by the semiconductor's Fermi level, presence of interface states, and by the semiconductor DOS. The boundary condition on the particle current e.g. for the electrons at the metal-semiconductor contact can be written as $J_n = qv_R(n - n_0)$, where n_0 is the electron density at the contacts in equilibrium and v_R is a recombination velocity, with units cm/s. The latter quantity depends on the exact

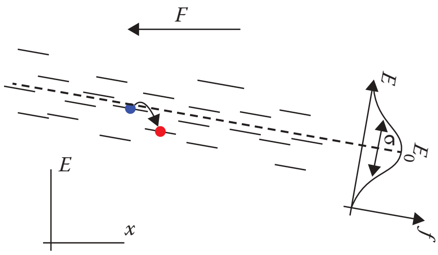


FIGURE 43.14 Schematic representation of the energetic disorder in an organic semiconductor. Each level represents a molecular energy level, which are disordered in energy and in space, following approximately a Gaussian distribution. Transport is driven by hopping between levels, described by energy difference and distance-dependent coefficients that depend also on the external electric field F.

transport mechanism and on the semiconductor DOS, and is therefore very different between inorganic and organic semiconductors [113].

A further difference of OPVs is their low permittivity of 3 ∼ 4 compared to inorganic materials with $\varepsilon_r \approx 10$. This leads to strong local Coulomb interaction which induces the formation of excitons even at room temperature. Since only electrons and holes can be collected at the contacts, the photo-generated excitons need to dissociate first. For this they need to reach the donor/acceptor interface, which is however possible only within their diffusion length of approximately 10 nm. For this reason, OPVs employ a blended donor/acceptor absorber, forming a BHJ. In combination with the hopping transport of charges in spatially and energetically disordered materials, these factors are the key differences to inorganics and make up for most of the challgenges associated with the modeling of OPVs. Since excitonic effects are so important in OPVs, several authors have proposed drift-diffusion-based models including explicitly exciton generation and dissociation [114–116].

The main disadvantage of state-of-the-art drift-diffusion models is the simplification of the BHJ morphology to a 1D effective medium. More details on the background and modeling techniques of the morphology can be found in Section 43.5.4. Since no interfaces exist in the effective medium, all interface effects between the donor/acceptor boundaries, such as exciton splitting and electron–hole recombination, are assumed to be bulk processes and evenly distributed across the photoactive layer. An absorbed photon directly creates and electron–hole pair and exciton diffusion and loss processes cannot be considered. Also, no charge accumulation effects at the donor/acceptor interface can be observed due to the lack of the blend morphology.

Most of these limitations do not apply to kMC simulations. A real 3D blend morphology, explicit treatment of exciton diffusion and splitting, and other interface processes, can easily be implemented. Also, the hopping transport in energetically and spatially disordered materials is naturally representable on a discrete lattice. However, since kMC is a statistical method, it often has to deal with unfeasible simulation times in order to achieve a reasonable accuracy for the desired quantities. This is especially critical when one has to sample very slow processes in presence of very fast events. Specifically, the evaluation of Coulomb interaction is expensive, and since particles are modeled explicitly and not by continuous quantities, the kMC algorithm is very demanding, in particular for a large number of particles. Large particle numbers are induced by a high energetic disorder, small injection barriers ("small barrier problem" in kMC [117]), or large device dimensions/thicknesses and limit the choice of parameters where kMC can be applied to certain ranges. Procedures to parallelize the algorithm and the Coulomb interaction need to be applied [118] to model larger systems and/or get better statistics in the results.

43.5.3 Perovskites

Perovskite, based devices pose two types of issues related to simulation. On the one hand, the fundamental working principle has not been completely unveiled yet, in particular the role of the perovskite in carrier transport. Therefore, it is not clear yet if the perovskite should be used merely as sensitizing absorber similar to the dye in a DSC, or if it should be used in a configuration more similar to an inorganic cell where the absorber acts also as electron and hole transporter. Partly, such questions arise because the development of perovskite PV is driven by the DSC and OPV comunities, with different views on device structure and functioning, and so also the point of view for modeling varies. From a fundamental point of view, being a crystalline material, perovskite seems to be more similar to an inorganic material with well-defined band structure [119], and it would seem that standard simulation models used for inorganic semiconductors should apply.

There are several issues encountered experimentally, where device simulation could give a strong contribution for a better understanding. One open question is whether ferroelectricity and ferroelectric domains exist in perovskite layers and whether they play a role on device functioning. While a fundamental answer to such questions based on density functional theory or MD is not easy, device-level simulation can at least provide inside into the qualitative effects that could be expected in presence of spatially varying polarization fields. This requires, however, simulation models that can include such effects. In particular, realistic domains have to be generated numerically, and the drift-diffusion model has to include the variable polarization field. Attempts in this direction have been made [120].

Another challenge for modeling is the hysteresis typically observed experimentally when measuring perovskite solar cell current–voltage characteristics [40,121]. Currently, the most probable cause of such hysteresis is believed to be migration of halide ion vacancies [122,123]. However, modeling of this effect in a standard device simulator is not easy, since the simulation model assumes transport of electrons and holes for a given fixed spatial distribution of defects. Therefore, self-consistent modeling of electronic transport and ion migration requires an extension of the standard drift-diffusion model with further mobile species, which is currently not available in standard device simulation software.

43.5.4 Modeling of Morphology and Grain Effects

Many device architectures for OPVs, (solid-state-) DSSCs and perovskite solar cells have regions with complex intermixed morphologies between two different materials where important processes are assumed to be controlled by the shape of the intermixing, especially mobilities through the morphologies and interface processes such as charge recombination and charge generation. For the sake of simplicity and low computational effort, the 3D blend is often simplified and treated as an effective medium, hiding the interface processes and 3D morphology effects in effective models for charge generation and recombination, and the mobility. An effective medium can be defined such that its band edges are given by the lower and higher values of the conduction and valence band edges of the two blended materials, respectively. Parameters like recombination rate constants and mobilities are then usually fitted to reproduce experimental data like current–voltage curves, or experimentally measured mobilities. It is yet to be investigated under which conditions an effective medium treatment is valid and which models are accurate for this description. Alternatively to the effective medium, the mobility can be derived from microscopic simulations considering the 3D blend morphology, like kMC. A comparison between the effective medium approach and a real blend kMC simulation in BHJ OPVs shows, however, that it might be difficult to include all interface effects in a blend into mobility and recombination models within an effective medium [124], especially in low-ε_r materials. A common basis in as many parameters as possible needs to be established for a comparison: the MO levels, the generation profile, and the injection processes, to name a few. By extracting electron and hole mobilities from the 3D simulations, the effect of the blend mobility can effectively be plugged into the effective medium model. It is then tried to obtain an agreement between the two models by tuning the recombination ratio. No general agreement could be obtained, which might indicate that it is not possible to merge all parameters in effective models accurately. It has been observed that, for instance, the charge density distribution within the 3D blend cannot be recovered by an effective model, since local accumulations at an interface are not present in the effective medium where both phases are assumed to be everywhere.

Thus, performing simulations on a realistic 3D blend is important to get a detailed picture about the processes in the blend. For a drift-diffusion model, complex 3D meshes are not easy to generate, increase considerably computation time and are prone to cause problems related to convergence. To avoid these problems for 3D blends, the kMC method is a suitable choice. The kMC algorithm does not have issues with convergence, even for complex morphologies. However, due to the popularity and establishment of drift-diffusion models, more effort should be made to include morphology effects in drift-diffusion-based simulation models. An example in this direction is given, e.g., in [125].

In kMC models, a blend can be created using a spin-exchange algorithm based on the work of Watkins et al. [111] which allows to control the grade of intermixing between two phases. However, the structure is based on a discrete lattice and is not suitable as basis for a DD simulation for which a high-quality finite element mesh is needed. A smoothing algorithm can be applied in order to achieve a mesh suitable for DD simulations. An illustration of a discrete BHJ lattice and a successively smoothed mesh is depicted in Figure 43.15. Further tests with an implementation of the 3D blend morphology as a mesh for DD simulations need to be performed to increase the common basis of the two models and get more information about where the two models differ and when the effective medium is appropriate.

Apart from nanometer-scale morphology effects, grain boundaries in multicrystalline materials can also affect optoelectronic device properties. Grain dimensions in such materials can range from nanometers to micrometers. A correct description of the effects of the grain boundaries, which can be related to both interface recombination and transport, may require to include the grain interfaces explicitly in the simulation model, especially if there is a limited number of interfaces along the carrier transport path such that an effective medium approximation may become questionable. Grain boundaries in organic materials have been studied related mostly to organic thin-film transistors [126,127]. Similar studies have been performed on polycrystalline silicon [128], and the effect of grain boundaries on carrier collection and recombination has been in CIGS devices has been studied in [129] and [130].

Note that also kMC does not automatically include the effect of grain boundaries, which should be included ad hoc with parameters based on *ab initio* or MD simulations [131].

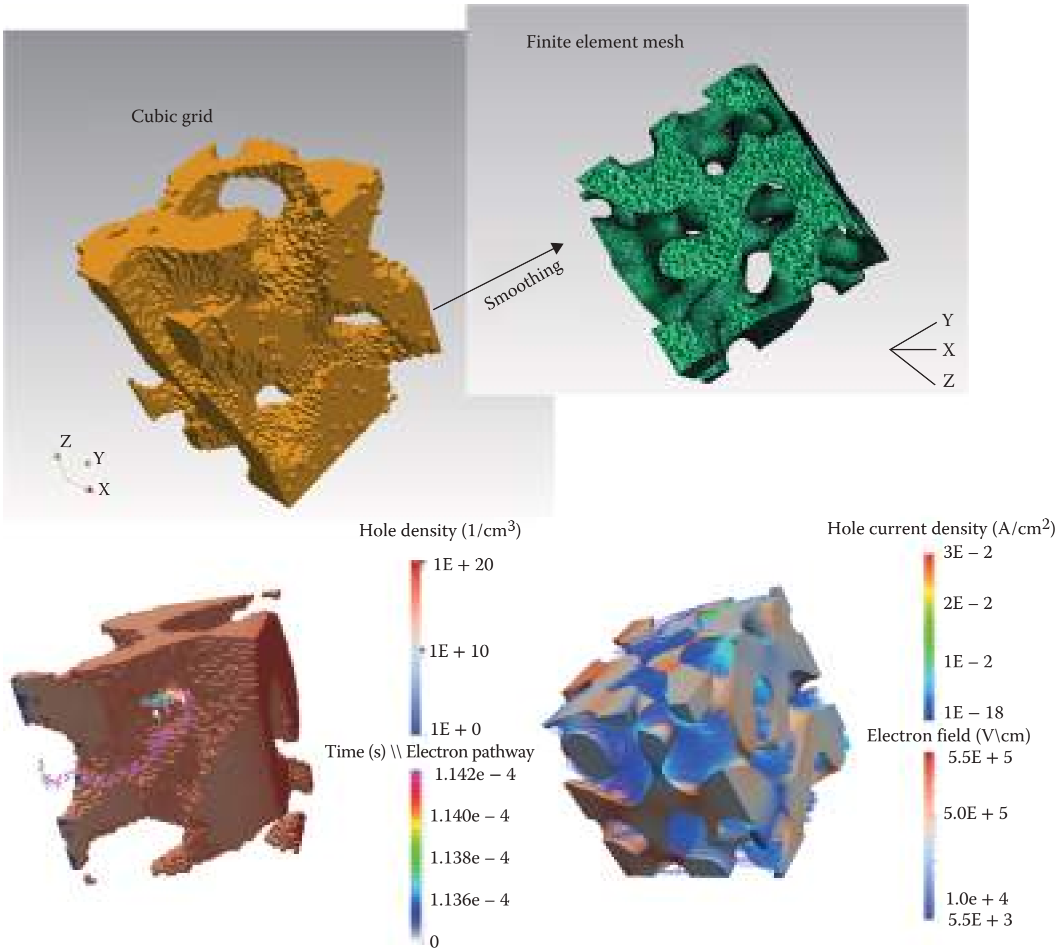

Finite element mesh
Cubic grid
Smoothing
Y
X
Z
Z
Y
X
Hole density (1/cm³)
1E + 20
1E + 10
1E + 0
Time (s) \\ Electron pathway
1.142e – 4
1.140e – 4
1.138e – 4
1.136e – 4
0
Hole current density (A/cm²)
3E – 2
2E – 2
1E – 2
1E – 18
Electron field (V\cm)
5.5E + 5
5.0E + 5
1.0e + 4
5.5E + 3

FIGURE 43.15 Smoothing a discrete lattice used for kMC simulations (top left) to obtain a suitable mesh for DD simulations (top right). Exemplary results of the kMC method (bottom left) highlighting the advantage to track single charge carrier trajectories and the DD model to extract, e.g., directed streams such as current densities through the morphology (bottom right).

References

1. J. Nelson. *The Physics of Solar Cells*. London: Imperial College Press, 2003.
2. J. Poortmans and V. Arkhipov, editors. *Thin Film Solar Cells—Fabrication, Characterization and Applications*. Hoboken, NJ: John Wiley & Sons, 2006.
3. K.L. Chopra, P.D. Paulson, and V. Dutta. Thin-film solar cells: An overview. *Progress in Photovoltaics: Research and Applications*, 12(2–3):69–92, 2004.
4. A. Jäger-Waldau. Status and perspectives of thin film photovoltaics. In A. Bosio and A. Romeo, editors, *Thin Film Solar Cells—Current Status and Future Trends*. Hauppauge, NY: Nova Science Publishers, 2011.
5. T.M. Brown, A. Reale, and A. Di Carlo. Organic and hybrid solar cells. In A. Bosio and A. Romeo, editors, *Thin Film Solar Cells—Current Status and Future Trends*. Hauppauge, NY: Nova Science Publishers, 2011.
6. Fraunhofer Institute for Solar Energy (ISE). *Photovoltaics Report*, 2015. Available at http://www.ise.fraunhofer.de/.
7. http://www.nrel.gov/ncpv.
8. M.A. Green. Consolidation of thin-film photovoltaic technology: The coming decade of opportunity. *Progress Photovoltaics: Research and Applications*, 14:383–392, 2006.
9. L. Dou, J. You, Z. Hong, Z. Xu, G. Li, R.A. Street, Y. Yang, L. Dou, J. You, Z. Hong, Z. Xu, G. Li, and Y. Yang. 25th Anniversary Article: A decade of organic/polymeric photovoltaic research. *Advanced Materials*, 25:6642–6671, 2013.
10. W. Cao and J. Xue. Recent progress in organic photovoltaics: Device architecture and optical design. *Energy & Environmental Science*, 7:2123–2144, 2014.
11. S.D. Stranks and H.J. Snaith. Metal-halide perovskites for photovoltaic and lightemitting devices. *Nature Nanotechnology*, 10(5):391–402, 2015.
12. A. Bosio and A. Romeo, editors. *Thin Film Solar Cells—Current Status and Future Trends*. Hauppauge, NY: Nova Science Publishers, 2011.
13. P. Delli Veneri and L.V. Mercaldo. Thin film silicon solar cells. In A. Bosio and A. Romeo, editors, *Thin Film Solar Cells—Current Status and Future Trends*. Hauppauge, NY: Nova Science Publishers, 2011.
14. M. Zeman. Advanced amorphous silicon solar cell technologies. In J. Poortmans and V. Arkhipov, editors, *Thin Film Solar Cells—Fabrication, Characterization and Applications*. Hoboken, NJ: John Wiley & Sons, 2006.
15. M.A. Green, K. Emery, Y. Hishikawa, W. Warta, and E.D. Dunlop. Solar cell efficiency tables (version 46). *Progress in Photovoltaics: Research and Applications*, 23(7):805–812, 2015.
16. H. Fritzsche. Development in understanding and controlling the Staebler-Wronski effect in a-Si:H. *Annual Review of Materials Science*, 31:47–79, 2001.
17. E.A. Schiff, S. Hegedus, and X. Deng. Amorphous silicon-based solar cells. In A. Luque and S. Hegedus, editors, *Handbook of Photovoltaic Science and Engineering*, pp. 487–545. Hoboken, NJ: John Wiley and Sons, 2011.
18. A. Bosio, A. Romeo, D. Menossi, S. Mazzamuto, and N. Romeo. Review: The secondgeneration of CdTe and CuInGaSe2 thin film PV modules. *Crystal Research and Technology*, 46(8):857–864, 2011.
19. S. Girish Kumar and K.S.R. Koteswara Rao. Physics and chemistry of CdTe/CdS thin film heterojunction photovoltaic devices: Fundamental and critical aspects. *Energy and Environmental Science*, 7:45–102, 2014.
20. R. Klenk and M.C. Lux-Steiner. Chalcopyrite based solar cells. In J. Poortmans and V. Arkhipov, editors, *Thin Film Solar Cells—Fabrication, Characterization and Applications*. Hoboken, NJ: John Wiley & Sons, 2006.

21. P. Reinhard, A. Chirila, P. Blosch, F. Pianezzi, S. Nishiwaki, S. Buechelers, and A.N. Tiwari. Review of progress toward 20% efficiency exible cigs solar cells and manufacturing issues of solar modules. *IEEE Journal of Photovoltaics,* 3(1):572–580, 2013.
22. T. Unold and C.A. Kaufmann. *Chalcopyrite Thin-Film Materials and Solar Cells*, Volume 1. Amsterdam: Elsevier, 2012.
23. B. O'Regan and M. Grätzel. A low-cost, high-efficiency solar cell based on dye-sensitized colloidal TiO2 films. *Nature*, 353:737–740, 1991.
24. K. Kalyanasundaram, editor. *Dye Sensitized Solar Cells*. Lausanne: EPFL Press, 2010.
25. S. Mathew, A. Yella, P. Gao, R. Humphry-Baker, B.F.E. Curchod, N. Ashari-Astani, I. Tavernelli, U. Rothlisberger, M.K. Nazeeruddin, and M. Grtzel. Dye-sensitized solar cells with 13% efficiency achieved through the molecular engineering of porphyrin sensitizers. *Nature Chemistry*, 6(3):242–247, 2014.
26. U. Bach, D. Lupo, P. Comte, J.E. Moser, F. Weissrtel, J. Salbeck, H. Spreitzer, and M. Grtzel. Solid-state dye-sensitized mesoporous TiO_2 solar cells with high photon-to-electron conversion efficiencies. *Nature*, 395(6702):583–585, 1998.
27. F. Matteocci, G. Mincuzzi, F. Giordano, A. Capasso, E. Artuso, C. Barolo, G. Viscardi, T.M. Brown, A. Reale, and A. Di Carlo. Blocking layer optimisation of poly(3-hexylthiopene) based solid state dye sensitized solar cells. *Organic Electronics*, 14(7):1882–1890, 2013.
28. A. Hagfeldt, G. Boschloo, L. Sun, L. Kloo, and H. Pettersson. Dye-sensitized solar cells. *Chemical Reviews*, 110:6595–6663, 2010.
29. A. Hinsch. Status of the dye solar cell technology (DSC) as a guideline for further research. In *28th European PV Solar Energy Conference and Exhibition*, Paris, France, 2013.
30. N.S. Sariciftci, L. Smilowitz, A.J. Heeger, and F. Wudl. Photoinduced electron transfer from a conducting polymer to buckminsterfullerene. *Science*, 258(5087):1474–1476, 1992.
31. C.W. Tang. Two-layer organic photovoltaic cell. *Applied Physics Letters*, 48(2):183–185, 1986.
32. G. Yu, J. Gao, J.C. Hummelen, F. Wudl, and A.J. Heeger. Polymer photovoltaic cells: Enhanced efficiencies via a network of internal donor-acceptor heterojunctions. *Science*, 270(5243):1789–1790, 1995.
33. Y. Liu, J. Zhao, Z. Li, C. Mu, W. Ma, H. Hu, K. Jiang, H. Lin, H. Ade, and H. Yan. Aggregation and morphology control enables multiple cases of high-efficiency polymer solar cells. *Nature Communications*, 5:5293, 2014.
34. C. Deibel and V. Dyakonov. Polymer-fullerene bulk heterojunction solar cells. *Reports on Progress in Physics*, 73(9):096401, 2010.
35. G. Li, R. Zhu, and Y. Yang. Polymer solar cells. *Nature Photonics*, 6(3):153–161, 2012.
36. H.-S. Kim, C.-R. Lee, J.-H. Im, K.-B. Lee, T. Moehl, A. Marchioro, S.-J. Moon, R. Humphry-Baker, J.-H. Yum, J.E. Moser, M. Grtzel, and N.-G. Park. Lead iodide perovskite sensitized all-solid-state submicron thin film mesoscopic solar cell with efficiency exceeding 9%. *Scientific Reports*, 2:591, 2012.
37. H.S. Jung and N.-G. Park. to Perovskite solar cells: From materials devices. *Small*, 11(1):10–25, 2015.
38. P. Gao, M. Grätzel, and M.K. Nazeeruddin. Organohalide lead perovskites for photovoltaic applications. *Energy and Environmental Science*, 7(8):2448–2463, 2014.
39. D. Wang, M. Wright, N.K. Elumalai, and A. Uddin. Stability of perovskite solar cells. *Solar Energy Materials and Solar Cells*, 147:255–275, 2016.
40. M. Grätzel. The light and shade of perovskite solar cells. *Nature Materials*, 13(9):838–842, 2014.
41. P.P. Boix, K. Nonomura, N. Mathews, and S.G. Mhaisalkar. Current progress and future perspectives for organic/inorganic perovskite solar cells. *Materials Today*, 17(1):16–23, 2014.
42. P. Würfel. *Physics of Solar Cells*. Weinheim: Wiley-VCH Verlag, 2009.
43. G.P. Smestad. *Optoelectronics of Solar Cells*. SPIE Press, 2002.

44. N.-K. Persson and O. Inganäs. Simulations of optical processes in organic photovoltaic devices. In S.-S. Sung and N.S. Sariciftci, editors, *Organic Photovoltaics—Mechanisms, Materials and Devices*. Hoboken, NJ: Taylor & Francis, 2005.
45. S.L. Chuang. *Physics of Optoelectronic Devices. Wiley Series in Pure and Applied Optics*, 1st edition. Hoboken, NJ: John Wiley & Sons, 1995.
46. A. Taove and S.C. Hagness. *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Norwood, MA: Artech House Publishers, 2005.
47. J. Jin. *The Finite Element Method in Electromagnetics*. Hoboken, NJ: Wiley, 2002.
48. I. Semenikhin, M. Zanuccoli, M. Benzi, V. Vyurkov, E. Sangiorgi, and C. Fiegna. Computational efficient RCWA method for simulation of thin film solar cells. *Optical and Quantum Electronics*, 44(3):149–154, 2012.
49. M. Born and E. Wolf. *Principles of Optics*. New York, NY: Cambridge University Press, 1999.
50. B. Harbecke. Coherent and incoherent reflection and transmission of multilayer structures. *Applied Physics B Photophysics and Laser Chemistry*, 39(3):165–170, 1986.
51. J.J.M. Halls, K. Pichler, R.H. Friend, S.C. Moratti, and A.B. Holmes. Exciton diffusion and dissociation in a poly(*p*-phenylenevinylene)/C_{60} heterojunction photovoltaic cell. *Applied Physics Letters*, 68(22):3120–3122, 1996.
52. J.J.M. Halls. *Photoconductive Properties of Conjugated Polymers*. PhD dissertation, Cambridge: St. John's College, 1997.
53. L.A.A. Pettersson, L.S. Roman, and O. Inganäs. Modeling photocurrent action spectra of photovoltaic devices based on organic thin films. *Journal of Applied Physics*, 86(1):487–496, 1999.
54. L.A.A. Pettersson, L.S. Roman, and O. Inganäs. Quantum efficiency of exciton-to-charge generation in organic photovoltaic devices. *Journal of Applied Physics*, 89(10):5564–5569, 2001.
55. N.-K. Persson, M. Schubert, and O. Inganäs. Optical modelling of a layered photovoltaic device with a polyuorene derivative/fullerene as the active layer. *Solar Energy Materials and Solar Cells*, 83(23):169–186, 2004.
56. H. Hoppe, N. Arnold, N.S. Sariciftci, and D. Meissner. Modeling the optical absorption within conjugated polymer/fullerene-based bulk-heterojunction organic solar cells. *Solar Energy Materials and Solar Cells*, 80(1):105–113, 2003.
57. D.P. Gruber, G. Meinhardt, and W. Papousek. Spatial distribution of light absorption in organic photovoltaic devices. *Solar Energy*, 79(6):697–704, 2005.
58. F. Monestier, J.-J. Simon, P. Torchio, L. Escoubas, F Flory, S. Bailly, R. de Bettignies, S. Guillerez, and C. Defranoux. Modeling the short-circuit current density of polymer solar cells based on p3HT:PCBM blend. *Solar Energy Materials and Solar Cells*, 91(5):405–410, 2007.
59. R. Husermann, E. Knapp, M. Moos, N.A. Reinke, T. Flatz, and B. Ruhstaller. Coupled optoelectronic simulation of organic bulk-heterojunction solar cells: Parameter extraction and sensitivity analysis. *Journal of Applied Physics*, 106(10):104507, 2009.
60. G.F. Burkhard, E.T. Hoke, and M.D. McGehee. Accounting for interference, scattering, and electrode absorption to make accurate internal quantum efficiency measurements in organic and other thin solar cells. *Advanced Materials*, 22(30):3293–3297, 2010.
61. L.D. Landau, E.M. Lifshitz, and L.P. Pitaevskii. *Electrodynamics of Continuous Media*. Amsterdam: Elsevier, 1993.
62. H.-Y.D. Yang. A spectral recursive transformation method for electromagnetic waves in generalized anisotropic layered media. *IEEE Transactions on Antennas and Propagation,* 45(3):520–526, 1997.
63. J. Ning and E.L. Tan. Hybrid matrix method for stable analysis of electromagnetic waves in stratified bianisotropic media. *IEEE Microwave and Wireless Components Letters,* 18(10):653–655, 2008.
64. A.H. Fallahpour, A. Gagliardi, F. Santoni, D. Gentilini, A. Zampetti, M. Auf der Maur, and A. Di Carlo. Modeling and simulation of energetically disordered organic solar cells. *Journal of Applied Physics*, 116(18), 2014.

65. M. Zeman, O. Isabella, K. Jger, R. Santbergen, S. Solntsev, M. Topic, and J. Krc. Advanced light management approaches for thin-film silicon solar cells. *Energy Procedia*, 15:189–199, 2012.
66. S. In, D.R. Mason, H. Lee, M. Jung, C. Lee, and N. Park. Enhanced light trapping and power conversion efficiency in ultrathin plasmonic organic solar cells: A coupled optical-electrical multiphysics study on the effect of nanoparticle geometry. *ACS Photonics*, 2(1):78–85, 2015.
67. M. Peters, B. Bläsi, S.W. Glunz, A.G. Aberle, J. Luther, and C. Battaglia. Optical simulation of silicon thin-film solar cells. *Energy Procedia*, 15:212–219, 2012.
68. A.H. Fallahpour, G. Ulisse, M. Auf der Maur, A. Di Carlo, and F. Brunetti. 3-D simulation and optimization of organic solar cell with periodic back contact grating electrode. *IEEE Journal of Photovoltaics*, 5(2):591–596, 2015.
69. M. Zeman, O. Isabella, S. Solntsev, and K. Jäger. Modelling of thin-film silicon solar cells. *Solar Energy Materials and Solar Cells*, 119:94–111, 2013.
70. N. Tucher, J. Eisenlohr, P. Kiefel, O. Höhn, H. Hauser, M. Peters, C. Müller, J.C. Goldschmidt, and B. Bläsi. 3D optical simulation formalism optos for textured silicon solar cells. *Optics Express*, 23(24):A1720–A1734, 2015.
71. S. Selberherr. *Analysis and Simulation of Semiconductor Devices*, 1st edition. New York, NY: Springer-Verlag Wien, 1984.
72. P.A. Markowich, C.A. Ringhofer, and C. Schmeiser. *Semiconductor Equations*, 1st edition. New York, NY: Springer-Verlag Wien, 1990.
73. W.V. van Roosbroeck. Theory of the flow of electrons and holes in germanium and other semiconductors. *Bell System Technical Journal*, 29:560–607, 1955.
74. S.M. Sze. *Semiconductor Devices: Physics and Technology*. New York, NY: John Wiley & Sons, 1985.
75. A. Schenk. An improved approach to the Shockley-Read-Hall recombination in inhomogeneous fields of spacecharge regions. *Journal of Applied Physics*, 71(7):3339–3349, 1992.
76. T. Goudon, V. Miljanovic, and C. Schmeiser. On the Shockley-Read-Hall model: Generation-recombination in semiconductors. *SIAM Journal on Applied Mathematics*, 67(4):1183–1201, 2007.
77. E. O'Reilly. *Quantum Theory of Solids*. London: Taylor & Francis, 2002.
78. S. John, C. Soukoulis, M.H. Cohen, and E.N. Economou. Theory of electron band tails and the urbach optical-absorption edge. *Physical Review Letters*, 57:1777–1780, 1986.
79. H. Bässler. Charge transport in disordered organic photoconductors: A Monte Carlo simulation study. *Physica Status Solidi (B)*, 175(1):15–56, 1993.
80. L.J.A. Koster, V.D. Mihailetchi, R. Ramaker, and P.W.M. Blom. Light intensity dependence of open-circuit voltage of polymer: Fullerene solar cells. *Applied Physics Letters*, 86(12):123509, 2005.
81. G. Paasch and S. Scheinert. Charge carrier density of organics with Gaussian density of states: Analytical approximation for the Gauss-Fermi integral. *Journal of Applied Physics*, 107(10):104501, 2010.
82. A. Zampetti, A.H. Fallahpour, M. Dianetti, L. Salamandra, F. Santoni, A. Gagliardi, M. Auf der Maur, F. Brunetti, A. Reale, T.M. Brown, and A. Di Carlo. Influence of the interface material layers and semiconductor energetic disorder on the open circuit voltage in polymer solar cells. *Journal of Polymer Science Part B: Polymer Physics*, 53(10):690–699, 2015.
83. G.K. Wachutka. Rigorous thermodynamic treatment of heat generation and conduction in semiconductor device modeling. *IEEE Transactions on Computer-Aided Design*, 11:1141–1149, 1990.
84. N.A. Zakhleniuk. Nonequilibrium drift-diffusion transport in semiconductors in presence of strong inhomogeneous electric fields. *Applied Physics Letters*, 89(25):252112, 2006.
85. N.F. Mott. On the transition to metallic conduction in semiconductors. *Canadian Journal of Physics*, 34(12A):1356–1368, 1956.
86. A. Miller and E. Abrahams. Impurity conduction at low concentrations. *Physical Review*, 120(3):745, 1960.
87. R.A. Marcus. Chemical and electrochemical electron-transfer theory. *Annual Review of Physical Chemistry*, 15(1):155–196, 1964.

88. W.F. Pasveer, J. Cottaar, C. Tanase, R. Coehoorn, P.A. Bobbert, P.W.M. Blom, D.M. De Leeuw, and M.A.J. Michels. Unified description of charge-carrier mobilities in disordered semiconducting polymers. *Physical Review Letters*, 94(20):206601, 2005.
89. D.T. Gillespie. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22(4):403–434, 1976.
90. A.F. Voter. Introduction to the kinetic Monte Carlo Method. *Radiation Effects*, 235:1–23, 2005.
91. H.L. Anderson. Scientific uses of the maniac. *Journal of Statistical Physics*, 43(5–6):731–748, 1986.
92. N. Metropolis, A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087, 1953.
93. J.R. Beeler (Jr). Displacement spikes in cubic metals. i. α-Iron, copper, and tungsten. *Physical Review*, 150(2):470, 1966.
94. E.S. Hood, B.H. Toby, and W.H. Weinberg. Precursor-mediated molecular chemisorption and thermal desorption: The interrelationships among energetics, kinetics, and adsorbate lattice structure. *Physical Review Letters*, 55(22):2437, 1985.
95. A.B. Bortz, M.H. Kalos, and J.L. Lebowitz. A new algorithm for Monte Carlo simulation of Ising spin systems. *Journal of Computational Physics*, 17(1):10–18, 1975.
96. C. Jacoboni and P. Lugli. *The Monte Carlo Method for Semiconductor Device Simulation. Computational Microelectronics*. Vienna: Springer, 1989.
97. D.T. Gillespie. Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361, 1977.
98. J.R. Norris. *Markov Chains*. Number 2008. Cambridge: Cambridge University Press, 1998.
99. F. Schwabl. *Statistical Mechanics. Advanced Texts in Physics*. Berlin: Springer, 2006.
100. J. Ferber, R. Stangl, and J. Luther. An electrical model of the dye sensitized solar cell. *Solar Energy Materials and Solar Cells*, 53:29–54, 1998.
101. P.R.F. Barnes and B.C. O'Regan. Electron recombination kinetics and the analysis of collection efficiency and diffusion length measurements in dye sensitized solar cells. *The Journal of Physical Chemistry C*, 114(44):19134–19140, 2010.
102. A. Gagliardi, S. Mastroianni, D. Gentilini, F. Giordano, A. Reale, T.M. Brown, and A.D. Carlo. Multiscale modeling of dye solar cells and comparison with experimental data. *IEEE Journal of Selected Topics in Quantum Electronics,* 16(6):1611–1618, 2010.
103. A. Gagliardi, M. Auf der Maur, D. Gentilini, F. Di Fonzo, A. Abrusci, H.J. Snaith, G. Divitini, C. Ducati, and A. Di Carlo. The real TiO_2/HTM interface of solid-state dye solar cells: Role of trapped states from a multiscale modelling perspective. *Nanoscale*, 7(3):1136–1144, 2015.
104. A. Gagliardi, M. Auf der Maur, D. Gentilini, and A. Di Carlo. Simulation of dye solar cells: Through and beyond one dimension. *Journal of Computational Electronics*, 10(4):424–436, 2011.
105. A. Gagliardi, M. Auf der Maur, and A. Di Carlo. Theoretical investigation of a dye solar cell wrapped around an optical fiber. *IEEE Journal of Quantum Electronics*, 47(9):1214–1221, 2011.
106. J. Nissfolk, K. Fredin, A. Hagfeldt, and G. Boschloo. Recombination and transport processes in dye-sensitized solar cells investigated under working conditions. *The Journal of Physical Chemistry B*, 110(36):17715–17718, 2006.
107. U. Landau. Current distribution in electrochemical cells: Analytical and numerical modeling. In *Modern Aspects of Electrochemistry No. 44: Modelling and Numerical Simulations II*, pp. 451–501. New York, NY: Springer New York, 2009.
108. J. Bisquert and V.S. Vikhrenko. Interpretation of the time constants measured by kinetic techniques in nanostructured semiconductor electrodes and dye-sensitized solar cells. *Journal of Physical Chemistry B*, 108(7):2313–2322, 2004.
109. D. Gentilini, A. Gagliardi, M. Auf der Maur, L. Vesce, D. D'Ercole, T.M. Brown, A. Reale, and A. Di Carlo. Correlation between cell performance and physical transport parameters in dye solar cells. *Journal of Physical Chemistry C*, 116(1):1151–1157, 2012.

110. S.D. Baranovskii. Theoretical description of charge transport in disordered organic semiconductors. *Physica Status Solidi (B)*, 251(3):487–525, 2014.
111. P.K. Watkins, A.B. Walker, and G.L.B. Verschoor. Dynamical Monte Carlo modelling of organic solar cells: The dependence of internal quantum efficiency on morphology. *Nano Letters*, 5(9):1814–1818, 2005.
112. M. Casalegno, G. Raos, and R. Po. Methodological assessment of kinetic Monte Carlo simulations of organic photovoltaic devices: The treatment of electrostatic interactions. *Journal of Chemical Physics*, 132(9):094705, 2010.
113. F. Santoni, A. Gagliardi, M. Auf der Maur, and A. Di Carlo. The relevance of correct injection model to simulate electrical properties of organic semiconductors. *Organic Electronics*, 15(7):1557–1570, 2014.
114. I. Kamohara, M. Townsend, and B. Cottle. Simulation of heterojunction organic thin film devices and exciton diffusion analysis in stacked-hetero device. *Journal of Applied Physics*, 97(1):014501, 2005.
115. Il-S. Park, S.-R. Park, D.-Y. Shin, J.-S. Oh, W.-J. Song, and J.-H. Yoon. Modeling and simulation of electronic and excitonic emission properties in organic host/guest systems. *Organic Electronics*, 11(2):218–226, 2010.
116. B. Ruhstaller, T. Beierlein, H. Riel, S. Karg, J.C. Scott, and W. Riess. Simulating electronic and optical processes in multilayer organic light-emitting devices. *IEEE Journal of Selected Topics in Quantum Electronics*, 9(3):723–731, 2003.
117. D. Kipp and V. Ganesan. A kinetic Monte Carlo model with improved charge injection model for the photocurrent characteristics of organic solar cells. *Journal of Applied Physics*, 113(23):234502, 2013.
118. N.J. van der Kaap and L.J.A. Koster. Massively parallel kinetic monte carlo simulations of charge carrier transport in organic semiconductors. *Journal of Computational Physics*, 307:321–332, 2016.
119. L. Pedesseau, J.-M. Jancu, A. Rolland, E. Deleporte, C. Katan, and J. Even. Electronic properties of 2D and 3D hybrid organic/inorganic perovskites for optoelectronic and photovoltaic applications. *Optical and Quantum Electronics*, 46(10):1225–1232, 2014.
120. A. Pecchia, D. Gentilini, D. Rossi, M. Auf der Maur, and A. Di Carlo. Role of ferroelectric nanodomains in the transport properties of perovskite solar cells. *Nano Letters*, 16(2):988–992, 2016.
121. D. Liu, M.K. Gangishetty, and T.L. Kelly. Effect of CH3NH3PbI3 thickness on device efficiency in planar heterojunction perovskite solar cells. *Journal of Materials Chemistry A*, 2:19873–19881, 2014.
122. S. Meloni, T. Moehl, W. Tress, M. Franckeviius, M. Saliba, Y.H. Lee, P. Gao, M.K. Nazeeruddin, S.M. Zakeeruddin, U. Rothlisberger, and M. Graetzel. Ionic polarization-induced current-voltage hysteresis in $CH_3NH_3PbX_3$ perovskite solar cells. *Nature Communications*, 7, 2016, Article number: 10334. doi:10.1038/ncomms10334.
123. J.M. Azpiroz, E. Mosconi, J. Bisquert, and F. De Angelis. Defect migration in methylammonium lead iodide and its role in perovskite solar cell operation. *Energy and Environmental Science*, 8(7):2118–2127, 2015.
124. T. Albes, P. Lugli, and A. Gagliardi. Investigation of the blend morphology in bulk-heterojunction organic solar cells. *IEEE Transactions on Nanotechnology*, 15(2):281–288, 2016.
125. H.K. Kodali and B. Ganapathysubramanian. A computational framework to investigate charge transport in heterogeneous organic photovoltaic devices. *Computer Methods in Applied Mechanics and Engineering*, 247:113–129, 2012.
126. A. Bolognesi, M. Berliocchi, M. Manenti, A. Di Carlo, P. Lugli, K. Lmimouni, and C. Dufour. Effects of grain boundaries, field-dependent mobility, and interface trap states on the electrical characteristics of pentacene TFT. *IEEE Transactions on Electron Devices*, 51(12):1997–2003, 2004.
127. Y. Hu, L. Wang, Q. Qi, D. Li, and C. Jiang. Charge transport model based on single-layered grains and grain boundaries for polycrystalline pentacene thin-film transistors. *Journal of Physical Chemistry C*, 115(47):23568–23573, 2011.

128. M. Kimura, S. Inoue, T. Shimoda, and T. Sameshima. Device simulation of carrier transport through grain boundaries in lightly doped polysilicon films and dependence on dopant density. *Japanese Journal of Applied Physics, Part 1: Regular Papers and Short Notes and Review Papers*, 40(9A):5237–5243, 2001.
129. W.K. Metzger and M. Gloeckler. The impact of charged grain boundaries on thin-film solar cells and characterization. *Journal of Applied Physics*, 98(6):063701, 2005.
130. K. Taretto and U. Rau. Numerical simulation of carrier collection and recombination at grain boundaries in Cu(In,Ga)Se_2 solar cells. *Journal of Applied Physics*, 103(9), 2008.
131. H. Kobayashi and Y. Tokita. Modeling of hole transport across grain boundaries in organic semiconductors for mesoscale simulations. *Applied Physics Express*, 8(5):051602, 2015.

IX

Novel Applications

44 Electroluminescent Refrigerators *Kuan-Chen Lee and Shun-Tung Yen* 541
Introduction • Theoretical Methods • Results and Discussion • Summary

45 Photonic Crystal Laser Diodes *Maciej Dems* 561
Photonic Crystals Basics • Fully Vectorial Method for Modeling VCSELs with Subwavelength Structures • VCSELs with Photonic Crystals for Light Confinement • Subwavelength Gratings in VCSELs • Summary

46 Single-Photon Sources *Niels Gregersen, Dara P. S. McCutcheon, and Jesper Mørk* 585
Introduction • Modeling of Spontaneous Emission • Cavity QED Approach • Waveguide QED Approach • Planar Dielectric Antenna Approach • Decoherence Effects • Summary

47 Nanoplasmonic Lasers and Spasers *A. Freddie Page and Ortwin Hess* 609
Introduction • Models for Metals and Gain • Transfer Matrix Method • Evolutionary Algorithm • Finite-Difference Time-Domain Method • Loss Compensation • Lasing Dynamics • Stopped-Light Lasing • Summary

48 Quantum-Dot Nanolasers *Christopher Gies, Michael Lorke, Frank Jahnke, and Weng W. Chow* 627
Going Nano: A New Era of Small Lasers • Connecting to Earlier Work: Laser Rate Equations and the β Factor • Key Aspects of Nanolasers • Essential Physics I: Quantum-Optical Semiconductor Laser Theory • Essential Physics II: Carrier Kinetics in QD Systems • Connecting to Experiments: Transition to Lasing and the β Factor • Interemitter Coupling Effects and Superradiance in Nanolasers • Nonresonant Coupling and Lasing from Multi-Exciton States in Few-QD Systems • Outlook

49 Nonlinear Dynamics in Quantum Photonic Structures *Gabriela Slavcheva and Mirella Koleva* 661
Introduction • Maxwell-Pseudospin Method • Simulation of Nonlinear Dynamics in Optical Waveguides and Semiconductor Microcavities • Optical Orientation in Quantum Wells and QDs • Modeling Coherent Spin Generation, Manipulation, and Readout in Charged QDs • Quantum Stochastic Formalism for Modeling Cavity-Emitter Systems • Modeling Optical Rotation in Chiral Carbon Nanotubes with Master Maxwell-Pseudospin Equations • Conclusion

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

44

Electroluminescent Refrigerators*

44.1 Introduction 541
44.2 Theoretical Methods 542
Fundamentals of EL Refrigeration • Device Structure and Main Equations • Recombination and Generation Processes • Boundary Conditions • Output Quantities and Models for Steady-State Analysis
44.3 Results and Discussion 549
Influence of Device Structures on EL Refrigeration • Influence of Optical Processes on EL Refrigeration
44.4 Summary 557

Kuan-Chen Lee
and
Shun-Tung Yen

44.1 Introduction

Solid-state optical refrigeration, which is featured by carrying the internal thermal energy away through light emission, has progressed rapidly after the first demonstration of photoluminescent (PL) refrigeration in 1995 (Epstein et al., 1995). To date, it has been shown that an Yb-doped $LiYF_4$ cooler can reach a temperature of 114 K and give a cooling power of 750 mW at room temperature (Melgaard et al., 2014). Further improving this kind of rare-earth-doped coolers to an even lower cooling temperature will be a difficult task. This is because the cooling process involves the thermalization of dopants and becomes inefficient when the thermal energy is comparable to the energy difference between discrete dopant levels in the ground-state manifold (Sheik-Bahae and Epstein, 2007). Different to the doped coolers, a semiconductor luminescent refrigerator is not restricted by such limitation and is expected to have lower operation temperature and higher cooling power. Moreover, it can be directly integrated with other semiconductor devices. These attractive features have stimulated extensive research on semiconductor luminescent refrigeration (Sheik-Bahae and Epstein, 2007, 2009). However, the required external efficiency for cooling is as high as nearly unity such that the realization of semiconductor PL refrigeration was not achieved until recently by Zhang et al. (2013) on CdS mircorod.

Semiconductor luminescent refrigeration using electrical pumping scheme, on the other hand, receives growing attention because of its looser requirement on the external efficiency compared to semiconductor PL refrigeration. Despite several works on the feasibility of electroluminescent (EL) refrigeration several decades ago (Berdahl, 1985; Dousmanis et al., 1964), the experimental verification of EL cooling has been reported just recently. In 2012, Santhanam et al. (2012) biased a light-emitting diode (LED) under several tens of microvolts and showed that the output emission power can exceed twice of the applied electrical input power. This positive difference of the output over the input powers implies that the LED is capable of extracting thermal energy from the environments, as a refrigerator, to keep heat flow balanced. However, the cooling power is smaller than nanowatts so that the cooling phenomenon is not directly observed. It has been known for a while that the cooling power density of an EL cooler can be significantly high if the cooler is biased at a voltage of $\sim E_g/q$, where E_g is the bandgap energy of the active layer and q the elementary charge. Extensive studies have been made to investigate the cooling feasibility and capability of EL refrigeration in this bias region. Mal'shukov and Chao studied the influence of the Auger recombination on the cooling capability of a GaAs-based refrigerator. They showed that a net cooling power density of several watts per square centimeter is achievable (Mal'shukov and Chao, 2001). Wang et al. performed a detailed self-consistent calculation with various recombination mechanisms accounted. Their analysis showed an impressive cooling efficiency of 35% for a cooler of p-GaAs active layer (Wang et al., 2005). Oksanen and Tulkki (2010) proposed a thermal heat pump, which consists of an EL cooler integrated directly with an absorbing diode, to deal with the problem of low light extraction efficiency encountered in semiconductor luminescent refrigeration. Other theoretical works, such as on heat transfer and cooling temperature (Han et al., 2007; Heikkilä et al., 2010; Lin et al., 2012; Piprek and Li, 2016), the design of refrigerator structure (Oksanen and Tulkki, 2010; Yen and Lee, 2010), the influence of the active materials (Yu et al., 2007), light extraction efficiency (Yen and Lee, 2010), and the photon recycling effect (Heikkilä et al., 2009; Lee and Yen, 2012), has also been conducted to provide a deep insight into EL refrigeration. Experimental works on the feasibility (Xue et al., 2015) and the application (Liu et al., 2013) of EL refrigeration are also being conducted. Although the high-cooling-power EL refrigeration has not been achieved, a thorough understanding of its operation mechanism still plays a necessary role in its realization.

---

* This chapter is reproduced with permission from Yen and Lee (2010), and Lee and Yen (2012).

In the following sections, we begin with the fundamental concept of EL refrigeration. The self-consistent calculation generally used for analyzing EL refrigeration and the operation principle of EL refrigeration is introduced in Section 44.2. Calculated results together with the analysis of various structural and optical effects on EL refrigeration are given in Section 44.3. Finally, a conclusion is given in Section 44.4.

44.2 Theoretical Methods

44.2.1 Fundamentals of EL Refrigeration

In EL refrigeration, electrons and holes are mediated as refrigerants and are intended to be injected into the active region for radiative recombination, which generates photons with thermal energy carried away. Recently, studies indicate that the extracted thermal energy comes from the thermalization of carriers during transport, similar to the Peltier effect in thermoelectric (TE) cooling (Santhanam et al., 2012; Oksanen and Tulkki, 2015; Xue et al., 2015). This process can be seen during the cooling cycle of EL refrigeration as illustrated in Figure 44.1a. At terminal contacts, carriers are injected into the device under an external bias V and flow all the way to the active layer. In the process of carrier transport, local cooling occurs in the place where the external bias is not large enough for carriers to overcome the potential barriers. This situation can be seen, for instance, in the cladding layer near the active layer. In such regions, only carriers of sufficiently high energy can surmount the barrier and move forward as shown in the inset circle. This causes high-energy empty states that are then occupied by carriers through thermalization, i.e., drawing phonons from the lattice matrix. The redistribution of carriers progresses rapidly and hence the quasi-thermal equilibrium is achieved as a result of much short phonon relaxation time. In the active layer, the injected electron–hole pairs recombine through various mechanisms. A heating energy of qV is generated if the electron–hole pair recombines nonradiatively. Contrary, for each photon generated from radiative recombination escaping out of the device, the internal energy of the whole system is reduced by $\hbar\omega - qV$. Here, $\hbar\omega$ is the photon energy, which comes from the combination of the thermal energy from the lattice matrix and the electric potential energy from the external bias. These three processes, i.e., electrical excitation from external bias, thermalization during transport, and relaxation through radiative recombination, complete the cooling cycle of EL refrigeration. Figure 44.1b shows the system view of an ideal EL cooler

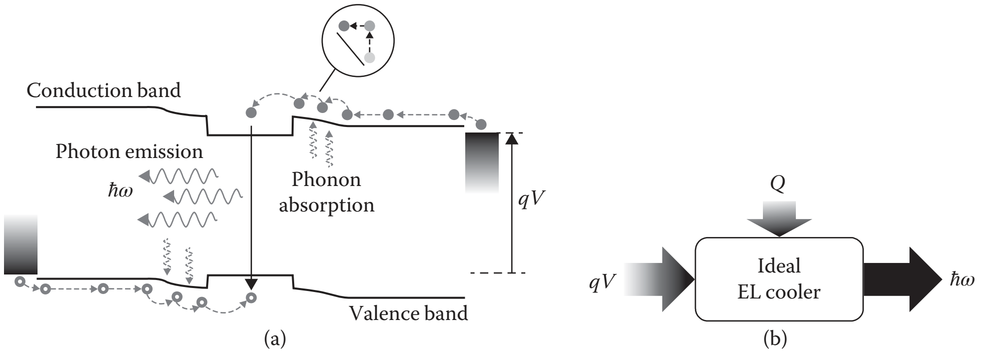


FIGURE 44.1 (a) The illustration of cooling cycling for EL refrigeration. The inset circle shows the thermalization process of carriers, which happens in the cladding layer near the active region. This is similar to the Peltier effect in thermoelectric cooling and attributed to EL refrigeration. (b) The system view of an ideal EL cooler. Net cooling occurs when $Q > 0$; otherwise, the device generates heat.

for a single cooling cycle, in which the parasitic effects are omitted. In this figure, an energy flow per cycle $Q = \hbar\omega - qV$ is required to be drawn from the ambient environment for compensating the internal energy loss of the system. If $Q > 0$, heat flows into the device and net cooling occurs; otherwise, the device generates heat. In this chapter, the cooling power, one of the most important parameters for coolers, is obtained in a similar way as Q described earlier, for which the input and the output powers are calculated from a self-consistent calculation based on one-dimensional drift-diffusion model with various recombination mechanisms accounted. In such a picture, the cooler is assumed to be in direct contact with a heat reservoir and the thermal resistance inside the cooler is neglected so that the temperature of the whole device is kept at a constant value of 300 K. In reality, a temperature gradient between the cooler and its environment exists due to a finite thermal conductance. The temperature distribution may also be nonuniform inside the cooler because of a nonuniform distribution of TE heat exchange (Yu et al., 2007). To account these effects, the heat balance equation between the cooler and its environment and the heat transfer equation inside the cooler have to be considered. Interested readers can refer to the works of Ashcroft and Mermin (1976), Han et al. (2007), Heikkilä et al. (2010), Lin et al. (2012), and Piprek and Li (2016) for more information. The geometric effects, such as current spreading and the detailed optical processes, are not accounted either. The consideration of these effects relies on the detailed information about the device geometry and a full three-dimensional simulation, which is beyond the scope of this chapter. The goal of this chapter is to provide a comprehensive view on how various structural and optical components affect the performance of an EL cooler without losing accuracy.

44.2.2 Device Structure and Main Equations

The device structure used in this work is composed of seven layers, as shown in Figure 44.2. They are, in order, a 100-nm p-doped $\mathrm{Al_{0.25}Ga_{0.75}As}$ cladding layer (with an acceptor concentration of $N_a = 1 \times 10^{18}$ cm^{-3}), a 50-nm p-doped $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ carrier blocking layer ($N_a = 1 \times 10^{18}$ cm^{-3}), a 50-nm undoped $\mathrm{Al_{0.25}Ga_{0.75}As}$ spacer, an undoped GaAs active layer of thickness $L = 100$ nm, a 50-nm undoped $\mathrm{Al_{0.25}Ga_{0.75}As}$ spacer layer, a 50-nm n-doped $\mathrm{Al}_x\mathrm{Ga}_{1-x}\mathrm{As}$ carrier blocking layer (with a donor concentration of $N_\mathrm{d} = 1 \times 10^{18}$ cm^{-3}), and a 100-nm n-doped $\mathrm{Al_{0.25}Ga_{0.75}As}$ cladding layer ($N_\mathrm{d} = 1 \times 10^{18}$ cm^{-3}). Figure 44.2 also demonstrates the calculated energy band profiles, including the band edges of the conduction band E_c and the valence band E_v, the electron quasi-Fermi level E_Fc, and the hole quasi-Fermi level E_Fv, along the growth direction of z under a forward bias of $V = 1.4$ V. The Al content x is 0.4 in the carrier blocking layer. For the drift-diffusion model, the central problem is to find out the electrostatic potential Φ

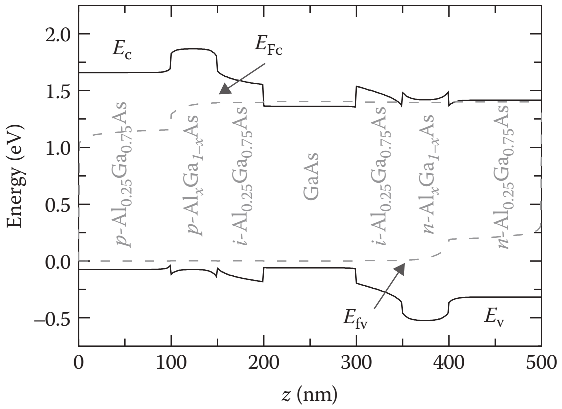


FIGURE 44.2 The energy band profiles, E_c and E_v, and the quasi-Fermi level profiles, E_{Fc} and E_{Fv}, along the growth direction z of the conduction and the valence bands, respectively, for the heterostructure under consideration. The profiles are obtained under a forward bias of 1.4 V.

and the quasi-Fermi levels for electrons E_{Fc}, and for holes E_{Fv} by solving three coupled equations (Horio and Yanai, 1990; Yen and Lee, 2010). These are the one-dimensional Poisson equation in the steady state:

$$\frac{d}{dz}\epsilon\frac{d}{dz}\Phi = -q\left(p - n + N_d^+ - N_a^-\right), \tag{44.1}$$

and the one-dimensional continuity equations for the electron and the hole currents:

$$\begin{aligned}\frac{d}{dz}J_n &= q(R - G),\\ \frac{d}{dz}J_p &= -q(R - G),\end{aligned} \tag{44.2}$$

where ϵ is the electric permittivity; p, n, N_d^+, and N_a^- are the concentrations of the holes, electrons, ionized donors, and ionized acceptors, respectively; J_n (J_p) is the current densities due to the electron (hole) flow; R is the electron–hole recombination rate minus the intrinsic generation rate due to thermal excitation; and G is the electron–hole generation rate due to the photon recycling effect to be described. It is noted that all of the aforementioned p, n, N_d^+, N_a^-, J_n, J_p, R, and G are functions of Φ, E_{Fc}, and E_{Fv}. The first six ones can be expressed by the following formulas:

$$n = N_c F_{1/2}\left(\eta_n\right), \eta_n \equiv \left(E_{Fc} - E_c\right)/k_B T, \tag{44.3}$$

$$p = N_v F_{1/2}\left(\eta_p\right), \eta_p \equiv \left(E_v - E_{Fv}\right)/k_B T, \tag{44.4}$$

$$N_d^+ = \frac{N_d}{\left\{1 + 2\exp\left[\left(E_{Fc} - E_d\right)/k_B T\right]\right\}}, \tag{44.5}$$

$$N_a^- = \frac{N_a}{\left\{1 + 4\exp\left[\left(E_a - E_{Fv}\right)/k_B T\right]\right\}}, \tag{44.6}$$

$$J_n = -q\mu_n n d\Phi/dz + qD_n dn/dz, \tag{44.7}$$

$$J_p = -q\mu_p p d\Phi/dz - qD_p dp/dz, \tag{44.8}$$

where N_c (N_v) is the effective density of states for the conduction (valence) band, F_j the Fermi–Dirac integral of order j, k_B the Boltzmann constant, T the temperature, E_d (E_a) the donor (acceptor) level, μ_n (μ_p) the electron (hole) mobility, and D_n (D_p) the diffusivity for the electrons and the holes.

44.2.3 Recombination and Generation Processes

For the recombination processes in the device, the Shockley–Read–Hall (SRH), the radiative, and the Auger recombinations are considered in the high-quality GaAs active layer, while only the SRH recombination process is assumed to be dominant in the other layers containing aluminum. Therefore, the recombination rate is

$$R = \begin{cases} R_{\mathrm{SRH}} + R_{\mathrm{rad}} + R_{\mathrm{Aug}} & z \text{ in the active layer,} \\ R_{\mathrm{SRH}} & \text{otherwise,} \end{cases} \tag{44.9}$$

where R_{SRH}, R_{rad}, and R_{Aug} are the recombination rates due to the SRH, the radiative, and the Auger recombination processes, respectively. In the calculation of R_{rad}, we consider the spontaneous emission contributed from the transition from the conduction band (c) to both the heavy-hole (hh) and the light-hole (lh) bands and express R_{rad} as

$$R_{\mathrm{rad}} = \int r(\hbar\omega)\, d\hbar\omega, \quad r = r_{\mathrm{hh}} + r_{\mathrm{lh}}, \tag{44.10}$$

where $\hbar$ is the reduced Planck constant, ω the optical angular frequency, and r the total spontaneous emission rate that consists of r_{hh} and r_{lh} attributed to the transitions of c→hh and c→lh, respectively. With the parabolic band approximation, r_{hh} can be expressed as (Chuang, 1995)

$$r_{\mathrm{hh}}(\hbar\omega) = C_{\mathrm{hh}} \frac{\hbar\omega\sqrt{\hbar\omega - E_g}\,\Theta(\hbar\omega - E_g)}{\left[1 + e^{-\eta_n}\exp\left(\frac{m_r}{m_c}\frac{\hbar\omega - E_g}{k_B T}\right)\right]\left[1 + e^{-\eta_p}\exp\left(\frac{m_r}{m_{\mathrm{hh}}}\frac{\hbar\omega - E_g}{k_B T}\right)\right]}, \tag{44.11}$$

where

$$C_{\mathrm{hh}} = \frac{q^2 n_r m_r^{3/2} E_P}{3\sqrt{2}\pi^3 \epsilon_0 m_0 \hbar^5 c^3}. \tag{44.12}$$

ϵ_0, m_0, and c are the electric permittivity, the electron mass, and the light speed, *in vacuo*, respectively; n_r is the refractive index; E_P is the optical matrix parameter; $m_r = m_c m_{\mathrm{hh}}/(m_c + m_{\mathrm{hh}})$ is the reduced mass; and Θ is the unit step function. For r_{lh}, it is simply obtained by replacing the role of m_{hh} in Equations 44.11 and 44.12 with the effective mass of the light hole band, m_{lh}.

For the Auger recombination rate, the CHHS and CHCC processes are considered. Therefore, R_{Aug} is

$$R_{\mathrm{Aug}} = \left(C_p p + C_n n\right)\left(np - n_i^2\right), \tag{44.13}$$

where C_p (C_n) is the Auger coefficient corresponding to the CHHS (CHCC) process and n_i the intrinsic carrier density. For the SRH recombination, we use

$$R_{\mathrm{SRH}} = \frac{pn - n_i^2}{\tau_n\left(p + n_i\right) + \tau_p\left(n + n_i\right)}, \tag{44.14}$$

where τ_n (τ_p) is the minority carrier lifetime when $p \gg n, n_i$ ($n \gg p, n_i$).

For the generation rate G, the carriers generated by photon recycling are accounted in the active layer. Photon recycling is a process involving the absorption of spontaneously emitted photons by the active material itself and hence strongly related to the distribution of optical field in the device. A detailed consideration of the recycling process relies on the information about the device geometry, which is out of the scope for the present work. In this work, the generation rate G is directly connected to R_{rad} by

$$G = \frac{1}{L}\eta_{\mathrm{pr}}\left(1-\eta_{\mathrm{xp}}\right)\int\limits_{z\in\ \text{active layer}} R_{\mathrm{rad}}dz, \tag{44.15}$$

where η_{xp} is the light extraction efficiency, the probability for the spontaneously generated photons emitting out of the device, and η_{pr} is the photon recycling efficiency, the ratio of trapped photons eventually being absorbed by the active layer. Therefore, Equation 44.15 implies that the photon recycling process generates carriers uniformly in the active layer, which results from the absorption of a portion of spontaneously generated photons trapped in the device. The assumption of a uniform carrier generation rate is justifiable for an active layer thickness no longer than the inverse of the absorption coefficient.

44.2.4 Boundary Conditions

To obtain an unambiguous distribution of Φ, E_{Fc}, and E_{Fv}, the boundary conditions for Equations 44.1 and 44.2 have to be specified. For the boundary condition of Φ, it is simply assumed that the electrostatic potential Φ and the displacement field $-\epsilon \mathrm{d}\Phi/\mathrm{d}z$ are continuous at interfaces. The boundary conditions of E_{Fc} and E_{Fv} are, however, complicated and classified into two cases according to the magnitude of band offsets. For the case of a band offset smaller than $2k_{\mathrm{B}}T$, the quasi-Fermi levels are determined from the continuity of drift-diffusion currents at the both sides of the interface under the assumption of a continuous quasi-Fermi level. For the other case of a larger band offset, the drift-diffusion model is no longer applicable for carriers' transport between layers. In this case, the thermionic process has to be taken into account, which allows quasi-Fermi levels to split at the interface. The thermionic emission current density due to the electron flow at the interface z_{i} with Fermi–Dirac distribution accounted can be expressed as[†]

$$J_{\mathrm{n}}\left(z_{\mathrm{i}}\right) = A_{\mathrm{n}}\left(z_{\mathrm{i}}^{+}\right)T^{2}F_{1}\left[\xi_{\mathrm{n}}\left(z_{\mathrm{i}}^{+}\right)\right] - A_{\mathrm{n}}\left(z_{\mathrm{i}}^{-}\right)T^{2}F_{1}\left[\xi_{\mathrm{n}}\left(z_{\mathrm{i}}^{-}\right)\right], \tag{44.16}$$

where $z_{\mathrm{i}}^{\pm} = \lim_{\varepsilon\to 0}\left(z_{\mathrm{i}}\pm\varepsilon\right)$, $A_{\mathrm{n}}\left(z_{\mathrm{i}}^{\pm}\right)$ is the effective Richardson constant of electrons at $z_{\mathrm{i}}^{\pm}$, and $\xi_{\mathrm{n}}\left(z_{\mathrm{i}}^{\pm}\right)$ is the energy difference, normalized by $k_{\mathrm{B}}T$, from the quasi-Fermi levels at $z_{\mathrm{i}}^{\pm}$ to the larger one of the conduction band edges at the interface. By matching the drift-diffusion current to the thermionic current at interfaces, we can write the boundary condition for E_{Fc} as

$$\begin{aligned} J_{\mathrm{n}}\left(z_{\mathrm{i}}\right) &= J_{\mathrm{n}}\left(z_{\mathrm{i}}^{\pm}\right), && \text{for} \left|E_{\mathrm{c}}\left(z_{\mathrm{i}}^{+}\right)-E_{\mathrm{c}}\left(z_{\mathrm{i}}^{-}\right)\right| \geq 2k_{\mathrm{B}}T, \\ J_{\mathrm{n}}\left(z_{\mathrm{i}}^{+}\right) &= J_{\mathrm{n}}\left(z_{\mathrm{i}}^{-}\right), && \text{otherwise,} \end{aligned} \tag{44.17}$$

with $J_{\mathrm{n}}\left(z_{\mathrm{i}}^{\pm}\right)$ being the drift-diffusion current at $z_{\mathrm{i}}^{\pm}$. The boundary condition for E_{Fv} is determined by the continuity of J_{p} at the interface, similar to that for E_{Fc} just described above.

At the end of surfaces on which the metal contacts are to be made, the Fermi levels are assumed to be pinned at a fixed position from the band edges such that the carrier densities equal to the values of those in thermal equilibrium. The Fermi level positions at the surface are thus determined by the charge neutral condition, $p + N_{\mathrm{d}}^{+} = n + N_{\mathrm{A}}^{-}$, with $E_{\mathrm{Fc}} = E_{\mathrm{Fv}}$.

[†] It is noted that, at a low temperature, the contribution of the tunneling current through an abrupt heterojunction may not be neglected. In such a situation, Equation 44.16 has to be modified. Interested readers can refer to Yang et al. (1993) for detailed expression.

With the boundary conditions being specified, the differential Equations 44.1 and 44.2 can be solved steadily by the Newton–Raphson iteration scheme for each applied voltage V. The external electrical work qV done on each of injected carriers gives the Fermi level difference between two surfaces $E_{\mathrm{Fc}}\big|_{z \text{ at cathode}} - E_{\mathrm{Fc}}\big|_{z \text{ at anode}} = qV$. Parameters used in this work are listed in Table 44.1, except for mobility, which is obtained from an empirical model (Sotoodeh et al., 2000). The band offset ratio ($\Delta E_{\mathrm{c}}/\Delta E_{\mathrm{v}}$) is set as 0.6:0.4 for GaAs/Al$_x$Ga$_{1-x}$As interfaces. The binding energies of donors and acceptors are set at $k_{\mathrm{B}}T$ for simplicity.

44.2.5 Output Quantities and Models for Steady-State Analysis

Once the solutions Φ, E_{Fc}, and E_{Fv} are obtained, the quantities of interest, including various current components, power densities, and efficiencies, can be calculated and used for further analysis. Figure 44.3 shows the working principle of an EL refrigerator. In the steady state, electrons and holes are injected

TABLE 44.1 Material Parameters of GaAs and AlGaAs at 300 K Used for Calculation

Parameter	Symbol	Unit	GaAs	Al$_x$Ga$_{1-x}$As ($x < 0.45$)
Bandgap	E_{g}	eV	$1.424 + 1.247x$	
Effective mass	m_{c}	m_0	$0.063 + 0.083x$	
	m_{hh}		$0.51 + 0.25x$	
	m_{lh}		$0.082 + 0.068x$	
Effective density of states	N_{c}	cm^{-3}	$2.5 \times 10^{19}(m_{\mathrm{c}}/m_0)^{3/2}$	
	N_{v}		$2.5 \times 10^{19}[(m_{\mathrm{hh}}/m_0)^{3/2} + (m_{\mathrm{lh}}/m_0)^{3/2}]$	
Effective Richardson constant	A_{n}	A/cm^2/K^2	$120 \times (m_{\mathrm{c}}/m_0)$	
	A_{p}		$120 \times (m_{\mathrm{hh}} + m_{\mathrm{lh}})/m_0$	
Electric permittivity	ϵ	ϵ_0	$12.9 - 2.84x$	
Refractive index	n_{r}	–	3.6	–
Optical matrix parameter	E_{P}	eV	25.7	–
SRH recombination lifetime	τ_{n}	s	1.3×10^{-6}	1×10^{-8}
	τ_{p}		1.2×10^{-6}	1×10^{-8}
Auger coefficient	C_{n}	cm^6/s	1.0×10^{-31}	–
	C_{p}		1.2×10^{-30}	–

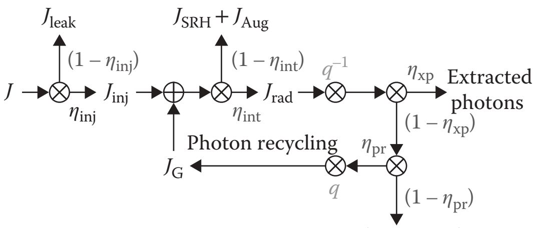


FIGURE 44.3 A diagram illustrating of the operation of an electroluminescent refrigerator with photon recycling as a feedback. It indicates the relationships among various currents. Symbol + stands for an addition operation that sums its inputs, indicated by inward arrows with the tips pointing at the symbol, and gives its output, indicated by an outward arrow with the tip pointing away from the symbol. Symbol × stands for multiplication operations, which give their outputs, indicated by outward arrows, by multiplying their inputs, indicated by inward arrows, by the factors along with the individual outward arrows.

from the electrodes under a forward bias, leading to a total current J in the device. Most of the electrons and holes are intended to be injected into the active region, recombine therein, and result in a major current component, called the injection current J_{inj}. The injection efficiency η_{inj} is defined as the ratio J_{inj}/J. The other part of the small current, called the leakage current J_{leak}, arises from the SRH recombination outside the active region $J'_{\text{SRH}} = \int_{z \notin \text{active layer}} R_{\text{SRH}} dz$ and the recombination current at the device surfaces $J_{\text{sr}} = J_{\text{n}}\big|_{z \text{ at anode}} + J_{\text{p}}\Big|_{z \text{ at cathode}}$. In this figure, we see that there are two means of injecting carriers into the active region. One is by the external electric injection and the other is by the photon recycling effect, which result in two effective current components J_{inj} and J_{G}, respectively. As a result, the sum $J_{\text{inj}} + J_{\text{G}}$ accounts for the total effective injection rate of electrons and holes to the active region. In the steady-state condition, the injection rate is balanced by the recombination rate in the active layer. Hence,

$$J_{\text{inj}} + J_{\text{G}} = J_{\text{SRH}} + J_{\text{rad}} + J_{\text{Aug}}, \tag{44.18}$$

where J_{SRH}, J_{rad}, and J_{Aug} are the current components in the active region due to the SRH, the radiative, and the Auger recombinations, respectively, with $J_i = q \int_{z \in \text{active region}} R_i dz$, for i = SRH, rad, and Aug. The internal quantum efficiency η_{int} is naturally defined as the probability for electron–hole pairs recombining through the radiative recombination:

$$\eta_{\text{int}} = \frac{J_{\text{rad}}}{J_{\text{SRH}} + J_{\text{rad}} + J_{\text{Aug}}}. \tag{44.19}$$

Owing to the radiative recombination, photons are generated with a rate J_{rad}/q. Since only a ratio η_{xp} of generated photons can escape out of the device, the emission flux is $\eta_{\text{xp}} J_{\text{rad}}/q$ and the corresponding emission power is:

$$P_{\text{rad}} = \eta_{\text{xp}} \int\int r\hbar\omega d\hbar\omega dz \equiv \eta_{\text{xp}} \langle \hbar\omega \rangle J_{\text{rad}}/q, \tag{44.20}$$

where $\langle \hbar\omega \rangle$ is the average photon energy. In addition to these emitted photons, the others are trapped in the device, either being parasitically absorbed, causing heat, or being recycled, reincarnating into electron–hole pairs in the active region. Figure 44.3 shows that the photon recycling effect is a feedback process, supplying a portion of electron–hole pairs required for recombinations. From Figure 44.3, we can simply express J_{G} as $J_{\text{G}} = \eta_{\text{pr}} \left(1 - \eta_{\text{xp}}\right) J_{\text{rad}}$, which is equivalent to Equation 44.15 with the equality $J_{\text{G}} = qGL$.

In conventional optoelectronic devices, the external quantum efficiency η_{ext} is defined as the ratio of the photon flux emitted out of the devices to the carrier flux injected into the devices, i.e., $\eta_{\text{ext}} = \eta_{\text{xp}} J_{\text{rad}}/J$. Using Equation 44.18 and the definitions of various efficiencies, η_{ext} with photon recycling effect accounted can be expressed in terms of the other efficiencies:

$$\eta_{\text{ext}} = \frac{\eta_{\text{xp}} \eta_{\text{int}} \eta_{\text{inj}}}{1 - \eta_{\text{int}} \eta_{\text{pr}} \left(1 - \eta_{\text{xp}}\right)}. \tag{44.21}$$

In the absence of photon recycling ($\eta_{\text{pr}} = 0$), η_{ext} returns to the conventional form of $\eta_{\text{ext}} = \eta_{\text{xp}} \eta_{\text{int}} \eta_{\text{inj}}$.

For the purpose of refrigeration, the cooling power and the cooling efficiency are two important parameters of interests. Following the concepts shown in Figure 44.1b, we define the cooling power as the power difference of the output and the input powers, which is required to keep the cooler operating steadily at 300 K:

$$P_{\text{c}} = P_{\text{rad}} - JV = \eta_{\text{c}} JV, \tag{44.22}$$

where $\eta_c \equiv P_c/JV$ is known as the cooling efficiency. The notion of EL refrigeration is to set the devices in the cooling mode, which requires a positive cooling power or cooling efficiency. Using Equation 44.20 and the definition of η_{ext}, one can rewrite the cooling efficiency as

$$\eta_c = \frac{\eta_{ext}}{\eta_{ext,cr}} - 1, \tag{44.23}$$

where

$$\eta_{ext,cr} \equiv \frac{qV}{\langle \hbar\omega \rangle}, \tag{44.24}$$

is the critical external efficiency for the devices to act in the cooling mode, which is consistent with the result from Figure 44.1b if the external efficiency of the cooler is considered. From Equation 44.21, the critical extraction efficiency for cooling is:

$$\eta_{xp,cr} = \frac{1 - \eta_{pr}\eta_{int}}{\eta_{int}\left(\eta_{inj}/\eta_{ext,cr} - \eta_{pr}\right)}, \tag{44.25}$$

for $\eta_{inj}/\eta_{ext,cr} > \eta_{pr}$.

44.3 Results and Discussion

In this section, we present the results simulated by the aforementioned self-consistent calculation to investigate various structural and optical effects on the cooling feasibility and capability of GaAs/AlGaAs EL refrigerators. In the discussion of the structural effects, we have assumed an unity light extraction efficiency ($\eta_{xp} = 1$).

44.3.1 Influence of Device Structures on EL Refrigeration

44.3.1.1 Leakage Current and Carrier Blocking Layer

We first study the cooling feasibility of a simple double heterostructure (DH) EL cooler by plotting the internal, the external, and the critical efficiencies shown in Figure 44.4 as functions of the applied voltage

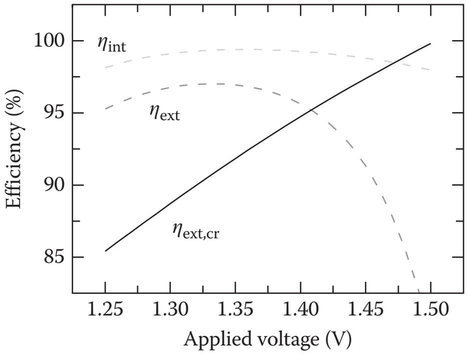
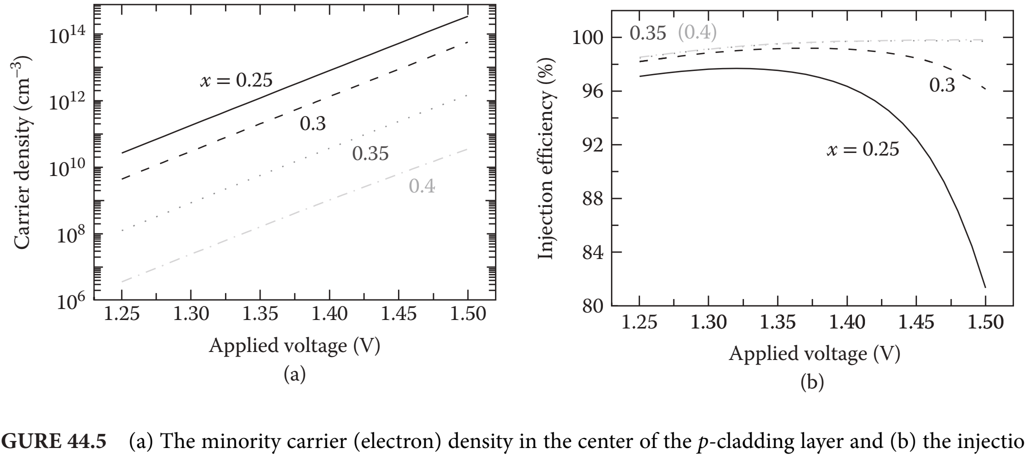


FIGURE 44.4 The internal efficiency η_{int}, the external efficiency η_{ext}, and the critical external efficiency $\eta_{ext,cr}$ as functions of the applied voltage V for the heterostructure with an active layer of width $L = 0.1$ μm and carrier blocking layers of Al composition $x = 0.25$, which has no effect on blocking carriers.

V for $\eta_{xp} = 1$. The Al content x in the blocking layer is 0.25, which is identical with that in the cladding and the spacer layers so that there is no blocking effect. As can be seen, while the critical external efficiency $\eta_{ext,cr}$ increases linearly with the applied voltage V, the internal efficiency η_{int} first increases to a maximum at $V = 1.36$ V and then decreases. The variation of η_{int} on V reveals a more important role played by the SRH and the Auger recombinations at a lower and a higher voltage, respectively. Similar to η_{int}, the external efficiency η_{ext} reaches to a maximum at $V = 1.33$ V and then droops more severely than η_{int} as V further increases, crossing with $\eta_{ext,cr}$ at the critical voltage $V_{cr} = 1.41$ V, where $\eta_{ext} = \eta_{ext,cr}$. We see that the DH EL cooler can act in the cooling mode for $V < 1.41$ V.

It is not difficult to find that the deterioration of η_{ext} is caused by a rapid reduction in the injection efficiency η_{inj} since $\eta_{inj} = \eta_{ext}/\eta_{int}$ for $\eta_{xp} = 1$. This implies that V_{cr} can be improved if the leakage current can be alleviated. In our calculation, it is shown that the strong surface recombination dominates in the leakage current as a result of a high minority carrier density near the terminal contacts. This situation becomes worse in particular for the devices of shorter cladding layers as in this work. The high minority carrier density can be reduced significantly by carrier blocking layers. In Figure 44.5a, we show the electron density in the center of the p-cladding layer as a function of V for $x = 0.25$, 0.3, 0.35, and 0.4. As can be seen, the electron density decreases monotonically with increasing x. This indicates that the energy barrier provided by the blocking layer can effectively block the minority carriers, preventing them from flowing over the carrier blocking layer. This leads to a lower minority carrier density in the cladding layer and hence a lower surface recombination current. The low surface recombination current is confirmed by the injection efficiency as plotted in Figure 44.5b. Evidently, the injection efficiency is improved significantly by increasing the barrier height when $x \leq 0.3$. The improvement however becomes insignificant as x increases from 0.35 to 0.4, in contrast to the result from Figure 44.5a This is because the leakage current across the blocking layer is almost eliminated when $x = 0.35$. In this case, $\eta_{inj} \approx 1$ and the residue of the leakage current is mostly due to the SRH recombination in the space layer. Contrarily, for $x \leq 0.3$, the injection efficiency decreases rapidly at high bias as a result of J_{sr}.

In Figure 44.6, we show the cooling power as a function of the applied voltage for EL cooler of various Al content in the blocking layer. As expected, the critical voltage increases from 1.41 V for $x = 0.25$ to 1.47 V for $x = 0.35$ and 0.4. Moreover, the cooling power also increases from 7.5 to 20 W/cm^2 for $x = 0.25$ to 0.35. The improvement of cooling power comes from an improved η_{inj} and hence η_{ext} as indicated by Equations 44.22 and 44.23, i.e., $P_c = JV\left(\eta_{ext}/\eta_{ext,cr} - 1\right)$. Further increasing the Al content to 0.4 does not help in improving the cooling power, which is consistent with the variation of the injection efficiency as shown in Figure 44.5b.



(a)

(b)

FIGURE 44.5 (a) The minority carrier (electron) density in the center of the p-cladding layer and (b) the injection efficiency as functions of the applied voltage V for various aluminum content $x = 0.25$, 0.3, 0.35, and 0.4.

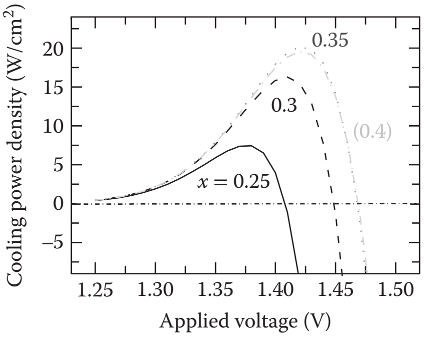


FIGURE 44.6 The ideal cooling power density P_c^0 as a function of the applied voltage V for various Al compositions of $x = 0.25$, 0.3, 0.35, and 0.4 in the carrier blocking layer. The extraction efficiency is set at unity.

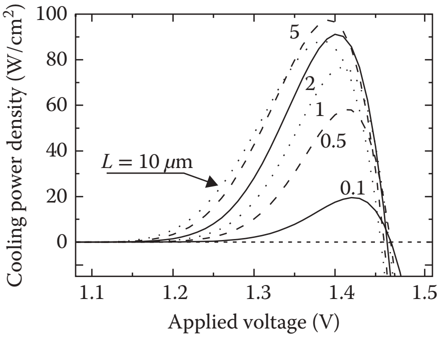


FIGURE 44.7 The ideal cooling power density P_c^0 as a function of V for $L = 0.1$, 0.5, 1, 2, 5, and 10 μm.

44.3.1.2 Active Layer Thickness

The cooling power density of 20 W/cm^2 seems to be the limiting value for the device of active layer thickness $L = 0.1$ μm. To further enhance it, one can increase the active layer thickness. Intuitively, a wider active layer can contain more electron–hole pairs to recombine therein, leading to an enhanced J_{rad} and hence more effective cooling cycles. Figure 44.7 shows the cooling power density P_c as a function of V for the structures with the same Al content ($x = 0.4$) in the carrier blocking layer but various active layer thickness $L = 0.1$, 0.5, 1, 2, 5, and 10 μm; here, $\eta_{\text{xp}} = 1$. As can be seen, the maximum cooling power $P_{\text{c,max}}$ can be improved about five times up to as high as 97 W/cm^2 by increasing L from 0.1 to 5 μm. However, this improvement is only prominent for $L \le 5$ μm. Further increasing the active layer thickness does not enhance but degrades the maximum cooling power due to the remarked nonuniformity of carrier distribution, as L is comparable to or greater than the carrier diffusion length. From our calculation, we find that the average carrier density in the active layer is lower for a larger L; this difference in the average carrier density between different Ls is particularly remarkable when V is high. This renders the cooling power for $L = 10$ μm larger than that for $L = 5$ μm at a low V but smaller at a high V.

In Figure 44.8, we show the dependences of various efficiencies on the active layer thickness by plotting the injection efficiency η_{inj} and the internal efficiency η_{int} in Figure 44.8a, the product $\eta_{\text{inj}}\eta_{\text{int}}$ and the critical external efficiency $\eta_{\text{ext,cr}}$ in Figure 44.8b, and the cooling efficiency η_c in Figure 44.8c as functions of the applied voltage V for the active layer thickness $L = 0.1$, 1, and 10 μm and $\eta_{\text{xp}} = 1$. The Al content of the blocking layer is $x = 0.4$. Since J_{inj} is enhanced by increasing L, the η_{inj} is also enhanced. For $L \ge 1$ μm, the injection efficiency is nearly 100%, almost independent of the applied voltage. As been described, the internal efficiency η_{int} reaches to a maximum at about 1.37 V. It declines more profoundly as V decreases

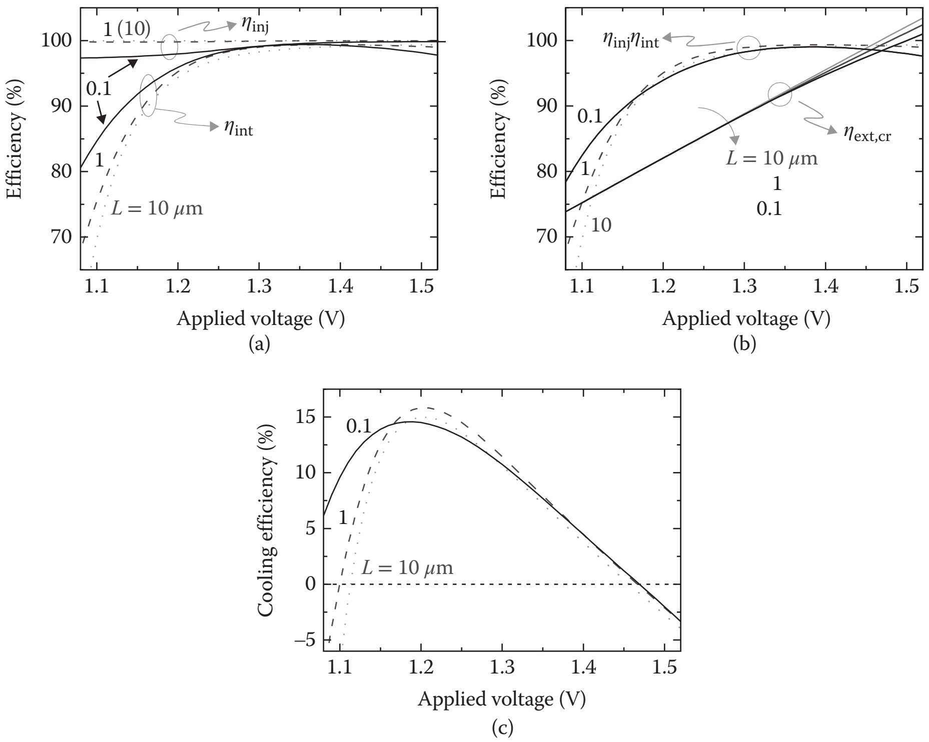


FIGURE 44.8 (a) The injection efficiency η_{inj} and the internal efficiency η_{int}, (b) the production of $\eta_{inj}\eta_{int}$ and the critical extraction efficiency $\eta_{ext,cr}$, and (c) the ideal cooling efficiency η_c^0 as functions of the applied voltage for $L = 0.1$, 1, and 10 μm.

than increases, implying a more important SRH recombination at low V than the Auger recombination at high V. Such a lowering of η_{int} at low V is more prominent for a larger L because of a lower carrier density in the active layer. On the other side, at high V, the Auger recombination is less important for a larger L due also to a lower carrier density. The ratio of $\eta_{ext,cr}/(\eta_{inj}\eta_{int})$ is the critical extraction efficiency $\eta_{xp,cr}$, which gives the design margin for η_{xp}. For instance, at $V = 1.2$ V, $\eta_{inj}\eta_{int} = 94\%$ and $\eta_{ext,cr} = 82\%$, as can be found in Figure 44.8b, give $\eta_{xp,cr} = 87\%$. Evidently, the margin becomes tighter when bias V deviates from 1.2 V. If an EL cooler is improperly biased at a voltage where $\eta_{xp,cr} \geq 1$, the ideal cooling efficiency will become negative (Figure 44.8c). In this case, it is impossible for the device to act in the cooling mode. At $V = 1.39$ V where the maximum cooling power occurs (see Figure 44.7), $\eta_{xp,cr} = 96\%$ for $L = 10$ μm. In such a bias condition, the corresponding cooling efficiency is only about 4%. The cooling efficiency can be improved by biasing the cooler at $V = 1.2$ V at which it can be as high as 15% at the cost of the cooling power density.

44.3.1.3 Doping in Active Layer

In Figure 44.9, we show the influences of the doping types and concentrations on EL refrigeration by plotting the ideal cooling power as a function of the applied voltage V for the active regions which are undoped, doped with $N_d = 1 \times 10^{18}\ \text{cm}^{-3}$, with $N_a = 1 \times 10^{18}\ \text{cm}^{-3}$, and with $N_a = 5 \times 10^{18}\ \text{cm}^{-3}$. Here, the thickness of the active layer is $L = 1$ μm and the Al content of the blocking layers is $x = 0.4$. We find that intentionally doping into the active region does not improve the cooling power. The degradation caused by p-type doping is mainly due to the enhanced Auger recombination. The n-type doping suppresses the radiative recombination and also the Auger recombination, leading to a slightly enhanced cooling power at high V but a reduced cooling power at low V.

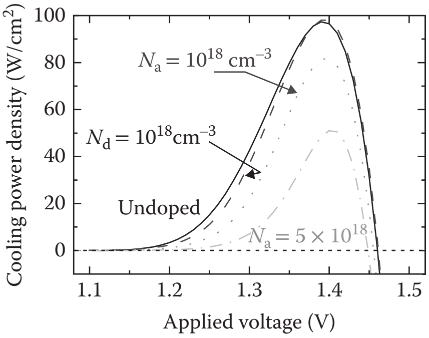


FIGURE 44.9 The ideal cooling power density P_c^0 as a function of the applied voltage for active layers with different doping types and concentrations, including undoped, doped with $N_d = 1 \times 10^{18}$ cm^{-3}, with $N_a = 1 \times 10^{18}$ cm^{-3}, and with $N_a = 5 \times 10^{18}$ cm^{-3}. Here, the active layer thickness is $L = 1$ μm, the Al content of the blocking layer is $x = 0.4$, and the light extraction efficiency is set at unity.

44.3.2 Influence of Optical Processes on EL Refrigeration

44.3.2.1 Influence of Light Extraction

Light extraction efficiency influences significantly the performance of an EL refrigerator in the absence of photon recycling. This is because the photons trapped within the device turn out to be totally absorbed and transformed into thermal energy. In this subsection, we study the influence of the light extraction efficiency on EL refrigeration for the worst case of zero photon recycling efficiency ($\eta_{pr} = 0$). Here, the active layer thickness L is set at 1 μm and the Al content $x = 0.4$ is used for the carrier blocking layer. Figure 44.10a shows the curves of the cooling power P_c as a function of the applied voltage V for five values of extraction efficiency $\eta_{xp} = 0.8$, 0.85, 0.9, 0.95, and 1. The curves exhibit a peak for $\eta_{xp} = 0.9$, 0.95, and 1, but no peak for $\eta_{xp} = 0.8$ and 0.85. As expected, the height of the peak diminishes drastically from 79 to 2 W/cm^2 as η_{xp} reduces slightly from unity to 0.9. The voltage at which the curve peaks moves from 1.41 V for $\eta_{xp} = 1$ to 1.28 V for $\eta_{xp} = 0.9$. The cooling power falls rapidly from the peak value to negative as the voltage increases.

These behaviors can be understood from Equation 44.23, i.e., $\eta_c = \eta_{ext}/\eta_{ext,cr} - 1$, with the help of the relationships between the curves of the corresponding external efficiencies η_{ext} and the critical external efficiencies $\eta_{ext,cr}$ shown in Figure 44.10b. The curves of $\eta_{ext,cr}$ for the different η_{xp} values merge into a single line, since $\eta_{ext,cr} = qV/\langle\hbar\omega\rangle$ is nearly proportional to V and independent of η_{xp}. The curves of η_{ext} for the different η_{xp} values behave similarly and separate almost equidistantly from their neighbors. As V increases, the η_{ext} value increases and then almost reaches saturation at a level $\eta_{ext} \approx \eta_{xp}$, with a slight decline in the high-voltage region. These features can be explained by the simple relation $\eta_{ext} = \eta_{xp}\eta_{int}\eta_{inj}$, where the injection efficiency η_{inj} is almost unity for the device containing carrier blocking layers. Therefore, the variation of η_{ext} with V for various η_{xp} simply follows the variation of the internal efficiency η_{int} multiplied by a scaling factor of η_{xp}. In Figure 44.10b, we see that there exist operation regions of refrigeration for $\eta_{xp} = 0.9$, 0.95, and 1, but no operation region for $\eta_{xp} = 0.8$ and 0.85, where the operation region of refrigeration is a range of voltage in which the device operates in the cooling modes ($\eta_c > 0$ or $\eta_{ext} > \eta_{ext,cr}$). This explains the feature shown in Figure 44.10a that only the three curves for $\eta_{xp} = 0.9$, 0.95, and 1 exhibit a peak of positive P_c. Figure 44.10c shows the cooling efficiency η_c as a function of V for $\eta_{xp} = 0.85$, 0.9, 0.95, and 1. We see that the peaks for η_c-V curves fall at $V = 1.2$ V. The cooling efficiencies decline when the operation voltage deviates from 1.2 V. Different to the peak voltage for η_c-V curves, the peaks of P_c-Vcurves in Figure 44.10a occur at voltages near the critical voltages, for which the cooling efficiency is small. This is why a slight increase of V may cause the drastic reduction of P_c from the peak value and turn the operation of the device from the cooling mode to the heating mode.

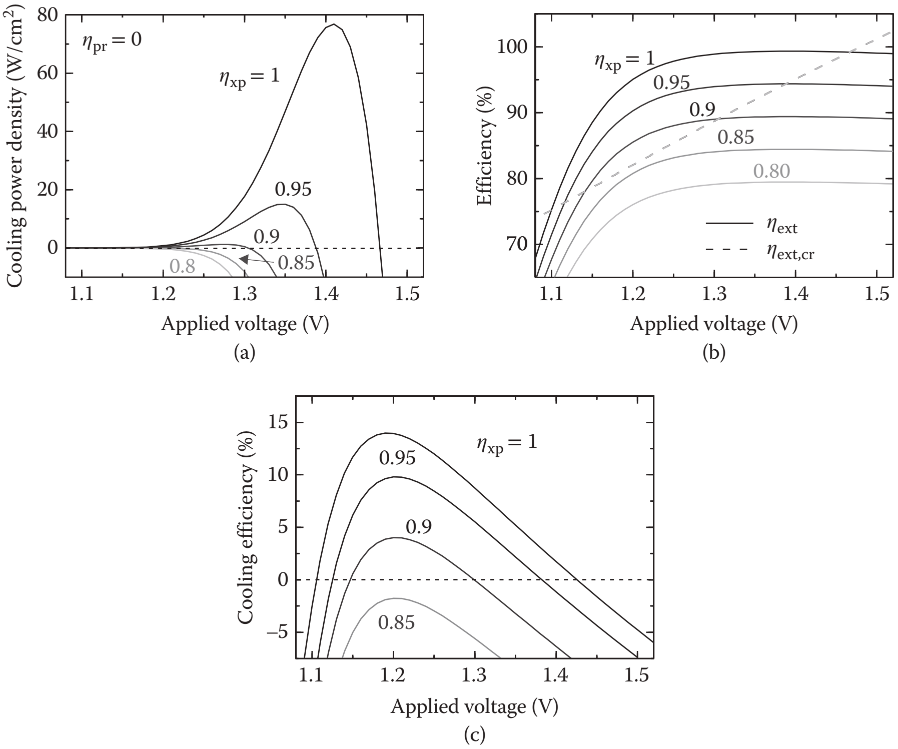


FIGURE 44.10 (a) The cooling power density P_c, (b) the external quantum efficiency η_{ext} and the critical extraction efficiency $\eta_{ext,cr}$, and (c) the cooling efficiency η_c as functions of the applied voltage for various extraction efficiency $\eta_{xp} = 0.8, 0.85, 0.9, 0.95$, and 1. The photon recycling efficiency is $\eta_{pr} = 0$.

44.3.2.2 Influence of Photon Recycling

Actually, most of the trapped photons can be reabsorbed by the high-quality GaAs active region and reincarnated into electron–hole pairs rather than into the thermal energy if the device structure is well designed. The photon recycling, as will be described, alleviates considerably the requirement of extraction efficiency for EL refrigeration. In this subsection, we demonstrate the photon recycling effect by examining another extreme case of unity recycling efficiency $\eta_{pr} = 1$, i.e., all of the trapped photons are recycled, turning into electron–hole pairs in the active region.

Figure 44.11a shows the cooling power density P_c as a function of V for five values of the extraction efficiency $\eta_{xp} = 0.2, 0.4, 0.6, 0.8$, and 1. All the curves in the figure behave as an asymmetrical peak of positive P_c, even for the low extraction efficiency of 0.2. The height of the peak now diminishes gently as η_{xp} reduces. These peaks occur almost at the same voltage ($V = 1.41$ V) except for $\eta_{xp} = 0.2$, for which the curve peaks at 1.39 V. Similarly, these behaviors can also be understood from Equation 44.23 and the relations between the corresponding η_{ext} and $\eta_{ext,cr}$ shown in Figure 44.11b. Comparing with those in Figure 44.10b, the external efficiencies η_{ext} shown in Figure 44.11b still vary with V in the manner that they first rise, reach a maximum, and then slightly decline. The difference is that η_{ext} is now much less sensitive to η_{xp}, especially in the voltage range where the internal efficiency η_{int} is near unity, as plotted in Figure 44.11c together with η_{inj} for various η_{xp}. In the voltage range between 1.3 and 1.5 V, where η_{int} is close to unity, the external efficiency η_{ext} can be approximated by

$$\eta_{ext} \approx \eta_{inj}\left[1 - \frac{1}{\eta_{xp}}\left(1 - \eta_{int}\right)\right]. \tag{44.26}$$

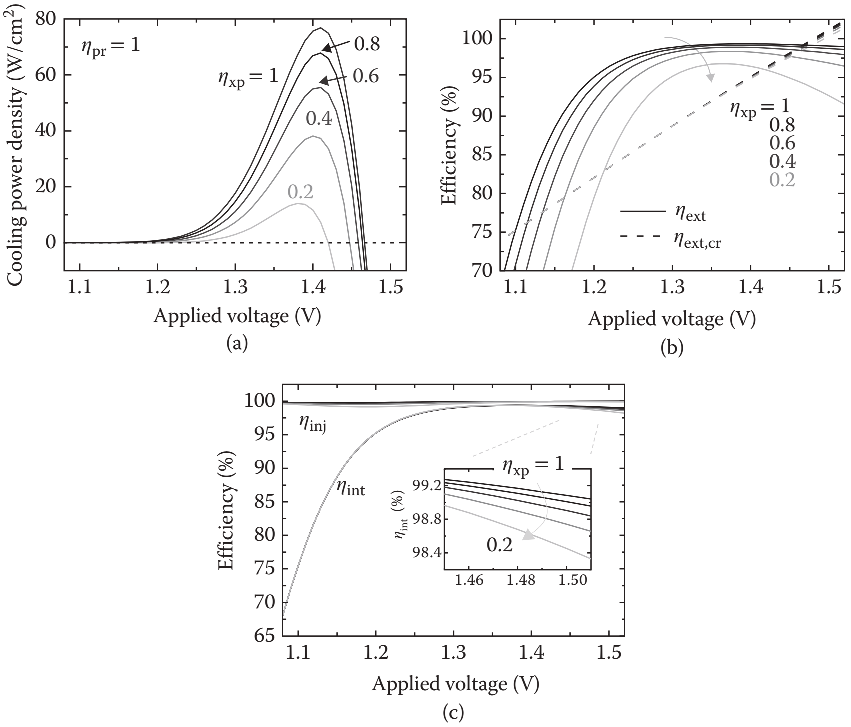


FIGURE 44.11 (a) The cooling power density P_c, (b) the external quantum efficiency η_{ext} and the critical external efficiency $\eta_{ext,cr}$, and (c) the injection efficiency η_{inj} and the internal efficiency η_{int} as functions of V for various extraction efficiency $\eta_{xp} = 0.2, 0.4, 0.6, 0.8$, and 1. The inset of (c) shows η_{int} at high V. The photon recycling efficiency is set at unity.

Equation 44.26 explains that, as $1 - \eta_{int}$ is small, η_{ext} is insensitive to η_{xp} and has a value close to η_{inj}, which is nearly unity.

In principle, the variation of η_{ext} basically follows the variation of η_{int} modified by photon recycling. This is because the only important loss is the nonradiative recombination in the active region when we set $\eta_{pr} = 1$ and $\eta_{inj} \approx 1$. For $\eta_{xp} = 1$, all the generated photons are emitted out of the device and there is no photon recycling, regardless of the value of η_{pr}, as can also be seen from Equation 44.21. The photon recycling effect manifests when η_{xp} deviates from unity. It causes an effective feedback current J_G and boosts the carrier concentration in the active region. The increasing carrier concentration degrades η_{int} and hence η_{ext} in the high-voltage region. This effect is particularly important for a low η_{xp} (see Figure 44.11b and the inset in Figure 44.11c). The lines of $\eta_{ext,cr}$ shown in Figure 44.11b do not merge as completely as those for the case of $\eta_{pr} = 0$ shown in Figure 44.10b. This is a consequence of the photon recycling effect, which causes a higher average photon energy $\langle \hbar\omega \rangle$ and, hence, a lower $\eta_{ext,cr}$ for a lower η_{xp}, due to a higher carrier concentration. The condition that $P_c = 0$ occurs at a critical voltage of about 1.47 V for $\eta_{xp} = 0.4$, 0.6, 0.8, and 1. The voltages are close to each other because of the crowding of the η_{ext}-Vcurves in the high-voltage region, which cross the $\eta_{ext,cr}$-Vcurves in the vicinity of 1.47 V. This implies that the P_c-V curves peak almost at the same voltage. The voltages for the peaks of the η_c-V curves, which can be estimated from Figure 44.11b, move from 1.2 to 1.3 V as η_{xp} decreases from 1 to 0.2.

Under the condition that $\eta_{pr} = 1$ and $\eta_{inj} \approx 1$, the critical extraction efficiency $\eta_{xp,cr}$ for the device to operate in the cooling mode can be written as

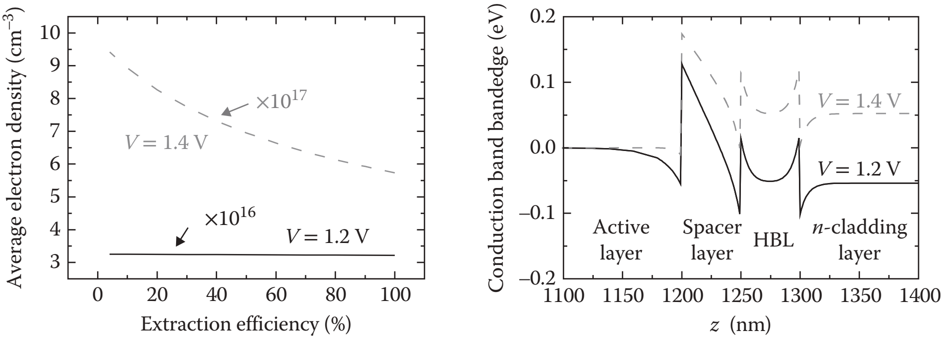
$$\eta_{xp,cr} \approx \frac{(1/\eta_{int}) - 1}{(1/\eta_{ext,cr}) - 1}. \tag{44.27}$$

The above expression implies that $\eta_{xp,cr}$ can be very small when η_{int} approaches unity, while $\eta_{ext,cr}$ does not. An EL refrigerator having a very low η_{xp} could operate in the cooling mode if there were no optical parasitic loss ($\eta_{pr} = 1$) and no current leakage ($\eta_{inj} = 1$).

The aforementioned discussion reveals that a high internal efficiency is essential to the photon recycling process. This implies that the quality of the materials making up the device is critical for refrigeration. If the material quality is not as high as in the study and gives $\eta_{int} < \eta_{ext,cr}$, it is impossible for the device to operate in the cooling mode as a result of the inequality $\eta_{ext} < \eta_{int}$ (see Equation 44.21), regardless of the values of the other efficiencies.

We now know that the internal efficiency η_{int} is an important factor in determining η_{ext} and, hence, to the performance of the device. It is influenced by photon recycling through the feedback generation of electron–hole pairs in the active region. To further gain insight into this effect, we set $\eta_{pr} = 1$ and investigate the average carrier density in the active region (Figure 44.12) as a function of η_{xp} when the device is biased at $V = 1.2$ and 1.4 V. As shown in Figure 44.12a, the average carrier density is low ($\approx 3 \times 10^{16}$ cm^{-3}) and nearly independent of η_{xp} for the device biased at 1.2 V, but is more than 5×10^{17} cm^{-3} and increases as η_{xp} reduces for the device biased at 1.4 V. This difference arises from the difference in the electric potential energy of the electrons (holes) in the active region relative to that in the *n*-type (*p*-type) cladding layer, from which the electrons (holes) are injected. This situation can be seen in Figure 44.12b, which shows the conduction band profiles at the *n*-doping side as functions of *z* for $V = 1.2$ and 1.4 V. When the device is biased at a low voltage, such as $V = 1.2$ V, the electric potential energy of the electrons is higher in the active layer than in the *n*-type cladding layer. This drop in electric potential energy causes the excess carriers, due to optical generation, to leak out of the active region, leading to an effective current J_G opposite the injection current J_{inj}. Consequently, accumulation of carriers in the active region is negligible, and the carrier density is basically independent of the photon recycling (and, hence, η_{int}). The drop in electric potential energy diminishes as the bias increases and, eventually, the situation reverses. At $V = 1.4$ V, the electric potential energy of the electrons is lower in the active region than in the n-type cladding layer. The potential profile now forms a potential well for the carriers and prohibits most of the excess carriers in the active region from leaking out. As a result, the optically generated carriers are mostly accumulated within the active region, recombined therein, and contribute an additional current, known as J_G. The carrier accumulation depends on the photon recycling and becomes particularly important for low η_{xp}, as shown in Figure 44.12a.

The variation of recombination current densities (J_{SRH}, J_{rad}, and J_{Aug}) on the extraction efficiency also follows the variation of carrier density such that they are invariant on η_{xp} at $V = 1.2$ V but increase considerably at $V = 1.4$ V for a decreasing η_{xp} (Lee and Yen, 2012). The feedback current J_G, which compensates J_{inj} for the total recombination in the active region, increases with decreasing η_{xp} for both



(a)

(b)

FIGURE 44.12 (a) The average electron densities in the active region for $V = 1.2$ and 1.4 V as functions of the extraction efficiency and (b) the conduction band profiles near the *n*-doping side as functions of *z* for $V = 1.2$ and 1.4 V. (HBL, hole-blocking layer.)

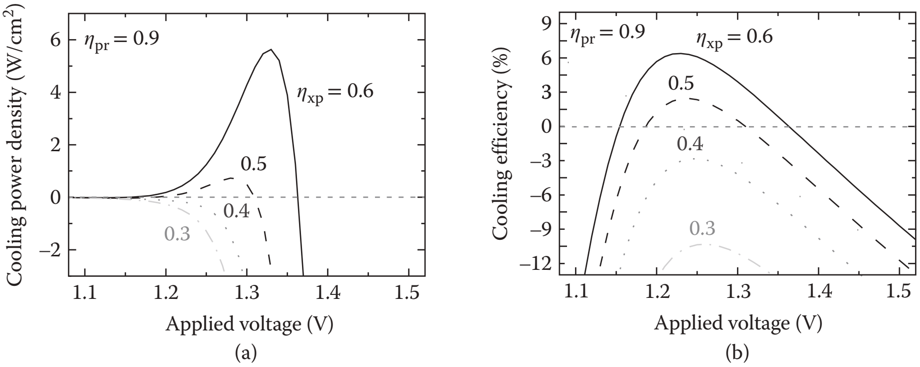


FIGURE 44.13 (a) The cooling power density P_c and (b) the cooling efficiency η_c as functions of applied voltage for various light extraction efficiencies $\eta_{xp} = 0.3$, 0.4, 0.5, and 0.6. The recycling efficiency is set at a practical value of 0.9.

applied voltages because $J_G = \eta_{pr}(1 - \eta_{xp})J_{rad}$. This leads to a significant reduction in J_{inj} and hence improves external efficiency for a low η_{xp}.

For a more practical case, we show the cooling power P_c in Figure 44.13a and the cooling efficiency η_c in Figure 44.13b as functions of the applied voltage V. Since $L = 1$ μm, η_{xp} is at most 60%. We therefore consider $\eta_{xp} = 0.3$, 0.4, 0.5, and 0.6 and set $\eta_{pr} = 0.9$, which is potentially achievable for a high-quality GaAs material. The resulting P_c-V relationship changes with η_{xp} to a degree milder than for $\eta_{pr} = 0$ (Figure 44.10a) but more severe than for $\eta_{pr} = 1$ (Figure 44.11a). The peak cooling power is now 5.6 W/cm^2 at $V = 1.33$ V for $\eta_{xp} = 0.6$, which is only 10% for the case of $\eta_{pr} = 1$ (Figure 44.11a). At the voltage of 1.33 V, the cooling efficiency η_c is small (2%) for $\eta_{xp} = 0.6$ and falls negative as η_{xp} reduces to 0.5, associated with the severe degradation of P_c from 5.6 to −3 W/cm^2. The situation can be cured by biasing the device at a lower voltage for an improved η_c. For example, at $V = 1.24$ V, η_c is 6.4% for $\eta_{xp} = 0.6$ and remains positive (about 2%) for $\eta_{xp} = 0.5$. The device can operate in the cooling mode for $\eta_{xp} > 0.45$. However, the cooling power is sacrificed. It is now only 0.9 W/cm^2 for $\eta_{xp} = 0.6$ and 0.3 W/cm^2 for $\eta_{xp} = 0.5$.

In the end of this chapter, we remark on the choices of active materials and operation temperatures for EL refrigeration. From Equation 44.24, the critical external efficiency of an EL cooler operating at E_g/q for maximum cooling power can be approximated by $\eta_{ext,cr} \approx E_g/\left(E_g + 2k_BT\right)$. Evidently, the external efficiency required for cooling is alleviated with decreasing bandgap energy and an increasing operating temperature. Indeed, recent research has shown that operating a 450-nm InGaN LED at 615 K is capable of producing a negative Peltier heating power inside the device (Xue et al., 2015). Moreover, theoretical study also confirms that elevating the operating temperature can improve the cooling power and efficiency for a GaAs EL cooler (Heikkilä et al., 2010). These improvements, as can be seen from the approximation of $\eta_{ext,cr}$, are strongly related to the ratio $E_g/2k_BT$. A smaller value of $E_g/2k_BT$ implies a fewer number of effective cooling cycles required for compensating a single parasitic loss. This means that the EL refrigeration will be more likely achieved for a cooler with an active material of smaller bandgap energy and operating in a higher temperature if the internal efficiency is high.

44.4 Summary

We discussed the influences of device structures and optical processes on the feasibility and capability of GaAs-based EL refrigerators by self-consistent calculation. For EL devices with short cladding layers, the leakage current, which arises from the minority carriers flowing toward the opposite electrodes, deteriorates the injection efficiency significantly at high bias. The degradation of injection efficiency can be cured by employing carrier blocking layers of $Al_xGa_{1-x}As$ ($x \geq 0.35$), leading to an injection efficiency of nearly unity. The maximum cooling power density generally increases with the active layer thickness because of an increasing total radiative recombination rate. A GaAs active layer with thickness of 5 μm can give a high internal efficiency of nearly 100% and a limiting cooling power density of 97 $\mathrm{W/cm^2}$. Further increasing the active layer thickness causes a nonuniform distribution of carriers, reducing the radiative power and also the maximum cooling power. Doping in the active layer does not help significantly in improving the cooling power density. Contrarily, a heavily p-doped active layer has much lower cooling power density than an undoped one owing to the enhanced Auger recombination.

The cooling capability depends strongly on the external efficiency, and hence the extraction efficiency, due to a small ratio of the extracted thermal energy to the work done on each injected electron–hole pair. For a GaAs EL cooler to act in the cooling mode, an external efficiency of 82% is needed at 1.2 V. This corresponds to an extraction efficiency of 87% if the trapped photons are all absorbed parasitically. This harsh requirement of extraction efficiency can be alleviated by photon recycling process provided that the internal efficiency is sufficiently high. The photon recycling behaves as an additional means of generating carriers, improving the external efficiency by reducing the driving currents. Therefore, for an ideal case of unity recycling efficiency, an EL device can act in the cooling mode even if the extraction efficiency is only 20%. For a more practical case of 90% recycling efficiency, our calculation shows that a maximum cooling power of 5.6 $\mathrm{W/cm^2}$ and a maximum cooling efficiency of 6.4% can be achieved for an EL cooler of 1-μm GaAs active layer. The limiting extraction efficiency for this cooler is found to be 45%, which is achievable for the state-of-the-art technology. This reveals a good possibility of experimentally realizing EL refrigeration for high-cooling-power application.

References

Ashcroft NW, Mermin ND (1976) *Solid State Physics*. Fort Worth, TX: Harcourt Brace College.

Berdahl P (1985) Radiant refrigeration by semiconductor diodes. *J. Appl. Phys.* 58:1369–1374.

Chuang SL (1995) *Physics of Optoelectronic Devices*. New York, NY: Wiley.

Dousmanis GC, Mueller CW, Nelson H, Petzinger KG (1964) Evidence of refrigerating action by means of photon emission in semiconductor diodes. *Phys. Rev.* 133:A316–A318.

Epstein RI, Buchwald MI, Edwards BC, Gosnell TR, Mungan CE (1995) Observation of laser-induced fluorescent cooling of a solid. *Nature* 377:500–503.

Han P, Jin KJ, Ren SR, Zhou YL, Lu HB (2007) Numerical analysis of optothermionic refrigeration in semiconductor triple-well structure. *J. Appl. Phys.* 102:114501.

Heikkilä O, Oksanen J, Tulkki J (2009) Ultimate limit and temperature dependency of light-emitting diode efficiency. *J. Appl. Phys.* 105:093119.

Heikkilä O, Oksanen J, Tulkki J (2010) The challenge of unity wall plug efficiency: The effects of internal heating on the efficiency of light emitting diodes. *J. Appl. Phys.* 107:033105.

Horio K, Yanai H (1990) Numeral modeling of heterojunctions including the thermionic emission mechanism at heterojunction interface. *IEEE Tran. Electron. Devices* 37:1093–1098.

Lee KC, Yen ST (2012) Photon recycling effect on electroluminescent refrigeration. *J. Appl. Phys.* 111:014511.

Lin FR, Lee KC, Yen ST (2012) The ultimate cooling temperature of semiconductor electroluminescent refrigeration. In *the 12th International Conference on Numerical Simulation of Optoelectronic Devices (NUSOD 2012)*, Shanghai, China, ThPD2.

Liu X, Zhao G, Zhang Y, Deppe DG (2013) Semiconductor laser monolithically pumped with a light emitting diode operating in the thermoelectrophotonic regime. *Appl. Phys. Lett.* 102:081116.

Mal'shukov AG, Chao KA (2001) Opto-thermionic refrigeration in semiconductor heterostructures. *Phys. Rev. Lett.* 86:5570–5573.

Melgaard S, Seletskiy D, Polyak V, Asmerom Y, Shiek-Bahae M (2014) Identification of parasitic loss in Yb:YLF and prospects for optical refrigeration down to 80 K. *Opt. Express* 22:7756–7764.

Oksanen J, Tulkki J (2010) Thermophotonic heat pump—A theoretical model and numerical simulations. *J. Appl. Phys.* 107:093106.

Oksanen J, Tulkki J (2015) LEDs feed on waste heat. *Nat. Photonics* 9:782–783.

Piprek J, Li Z-M (2016) Electroluminescent cooling mechanisms in InGaN/GaN light-emitting diodes. *Opt. Quant. Electron*. 48:472.

Santhanam P, Gray DJ Jr., Ram RJ (2012) Thermoelectrically pumped light-emitting diodes operating above unity efficiency. *Phys. Rev. Lett.* 108:097403.

Sheik-Bahae M, Epstein RI (2007) Optical refrigeration. *Nat. Photonics* 4:693–699.

Sheik-Bahae M, Epstein RI (2009) Laser cooling of solids. *Laser Photon. Rev.* 3:67–83.

Sotoodeh M, Khalid AH, Rezazadeh AA (2000) Empirical low-field mobility model for III-V compounds applicable in device simulation codes. *J. Appl. Phys.* 87:2890–2900.

Wang JB, Ding D, Yu SQ, Johnson SR, Zhang YH (2005) Electroluminescence cooling in semiconductors. In 2005 *Conference on Lasers and Electro-Optics/Quantum Electronics and Laser Science and Photonic Applications Systems Technologies, QELS'05*, Baltimore, MD, 655–657.

Xue J, Zhao Y, Oh SH, Herrington WF, Speck JS, Denbaars ST, Nakamura S, Ram RJ (2015) Thermally enhanced blue light-emitting diode. *Appl. Phys. Lett.* 107:121109.

Yang K, East JR, Haddad GI (1993) Numerical modeling of abrupt heterojunctions using a thermionic-field emission boundary condition. *Solid State Electron.* 36:321–330

Yen ST, Lee KC (2010) Analysis of heterostructures for electroluminescent refrigeration and light emitting without heat generation. *J. Appl. Phys.* 107:054513.

Yu SQ, Wang JB, Ding D, Johnson SR, Valsileska D, Zhang YH (2007) Impact of electronic density of states on electroluminescence refrigeration. *Solid State Electron.* 51:1387–1390.

Zhang J, Li D, Chen R, Xiong Q (2013) Laser cooling of a semiconductor by 40 Kelvin. *Nature* 493:504–508.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

45

Photonic Crystal Laser Diodes

45.1 Photonic Crystals Basics 561
Optical Periodic Structures • Photonic Bandgap • Photonic Crystals in Lasers
45.2 Fully Vectorial Method for Modeling VCSELs with Subwavelength Structures 565
Plane-Wave Admittance Method • Plane-Wave Reflection Transfer
45.3 VCSELs with Photonic Crystals for Light Confinement.......... 570
45.4 Subwavelength Gratings in VCSELs.................................... 574
High-Contrast Subwavelength Grating Theory • Modeling of Subwavelength Grating Reflectivity • Cavities with Subwavelength Grating Mirrors • Manufacturing Sensitivity of High-Contrast Grating Mirrors
45.5 Summary.. 582

Maciej Dems

45.1 Photonic Crystals Basics

45.1.1 Optical Periodic Structures

Photonic crystals are periodic structures designed to affect electromagnetic waves in a similar manner to how solid-state crystals affect electrons. The simplest, one-dimensional form of a photonic crystal is a periodic multilayer film (Bragg mirror). Such films were already being studied 120 years ago, in 1887, by Rayleigh [1]. They continued to be the subject of intensive research, resulting in the development of thin-film optics: the science and technology of fabricating dielectric multilayer mirrors, filters, polarizers, antireflection coatings, and so on. However, it took a whole century before the technology was expanded into two and three dimensions by Yablonovitch [2] and John [3] in 1987. Since then, the term *photonic crystal* has become widely used and the subject is studied all over the world.

The physics of photonic crystals has much in common with solid-state physics, as both concern the solution of wave equations in a periodic medium. Hence, concepts known from solid-state physics can also be used in photonics. In particular, for any regular lattice, it is possible to define a *reciprocal lattice* in the Fourier space, i.e., the space of the wavevectors. Such reciprocal lattices are an important tool for the analysis of light propagation in photonic crystals. Because of the Bloch theorem, any eigenmode in periodic medium can be represented as

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}_{\mathbf{k}}(\mathbf{r}) \exp(-i\,\mathbf{k} \cdot \mathbf{r}), \tag{45.1}$$

where $\mathbf{k}$ is a wavevector and $\mathbf{E}_{\mathbf{k}}$ is a periodic function with the same periodicity as the refractive index, i.e.,

$$n_{\mathrm{R}}(\mathbf{r} + \mathbf{a}_i) = n_{\mathrm{R}}(\mathbf{r}), \tag{45.2}$$

$$\mathbf{E}_{\mathbf{k}}(\mathbf{r} + \mathbf{a}_i) = \mathbf{E}_{\mathbf{k}}(\mathbf{r}). \tag{45.3}$$

Vectors $\mathbf{a}_i$ are referred to as the base vectors of the periodic lattice and their number is equal to the number of dimensions (in Equations 45.2 and 45.3).

A reciprocal lattice has vectors $\mathbf{b}_j$ defined in the space of wavevectors. They fulfill the condition

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij},$$

where δ_{ij} is Kronecker's delta. For three 3D lattices, $\mathbf{b}_j$ can be expressed directly as

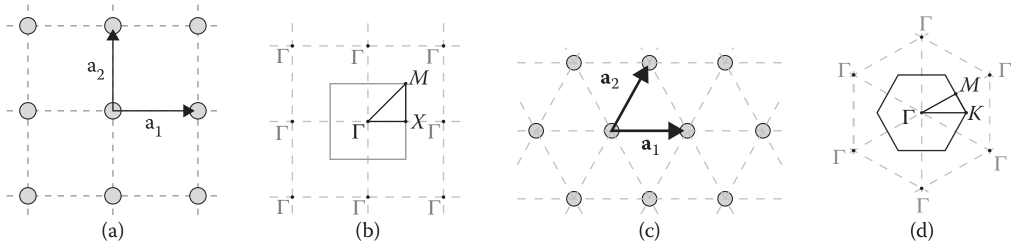
$$\mathbf{b}_1 = 2\pi\frac{\mathbf{a}_2 \times \mathbf{a}_3}{a_1 a_2 a_3}, \qquad \mathbf{b}_2 = 2\pi\frac{\mathbf{a}_3 \times \mathbf{a}_1}{a_1 a_2 a_3}, \text{ and} \qquad \mathbf{b}_3 = 2\pi\frac{\mathbf{a}_1 \times \mathbf{a}_2}{a_1 a_2 a_3},$$

and for 2D lattices as

$$\mathbf{b}_1 = 2\pi\frac{\mathbf{a}_2 \times \hat{\mathbf{n}}}{a_1 a_2}, \qquad \mathbf{b}_2 = 2\pi\frac{\hat{\mathbf{n}} \times \mathbf{a}_1}{a_1 a_2},$$

where $\hat{\mathbf{n}}$ is the unit vector perpendicular to the lattice plane. These vectors define the periodicity of the wavevector $\mathbf{k}$ from Equation 45.1. This periodicity enables the modal analysis of PCs with the vectors $\mathbf{k}$ located within a limited area, called the first Brillouin zone [4].

In the case of two-dimensional photonic crystals, two lattice configurations are particularly common. These are the square lattice and the hexagonal lattice, shown in Figure 45.1a and c. Their reciprocal lattices have the same properties as corresponding physical lattices and are depicted in Figure 45.1b and d. Each of these lattices possesses several high symmetry points, which are important for determination of the photonic band structure. The Γ point is always defined for wavevector $\mathbf{k} = \mathbf{0}$. Others are located at the edges of the first Brillouin zone. For the square lattice, these are X for $\mathbf{k} = \frac{\pi}{a}\hat{\mathbf{x}}$ and M for $\mathbf{k} = \frac{\pi}{a}\hat{\mathbf{x}} + \frac{\pi}{a}\hat{\mathbf{y}}$[†]. In the case of the hexagonal lattice, the high symmetry points are M ($\mathbf{k} = \frac{\pi}{a}\hat{\mathbf{x}} + \frac{\pi}{a\sqrt{3}}\hat{\mathbf{y}}$) and K ($\mathbf{k} = \frac{4\pi}{3a}\hat{\mathbf{x}}$). These positions are given for the points marked in Figure 45.1, although there are other points of each kind on the other edges and corners. In practice, this changes little, as the eigenmodes may be subjected to any symmetry transformation that does not change the distribution of the refractive index. Hence, the triangles made by the adjacent symmetry points are the *irreducible zone*, into which all the possible wavevectors can



(a) (b) (c) (d)

FIGURE 45.1 Arrangements of periodic cells and base vectors $\mathbf{a}_1$ and $\mathbf{a}_2$ of two-dimensional photonic crystals for (a) rectangular and (c) triangular lattices and reciprocal lattices of (b) rectangular and (d) triangular two-dimensional photonic crystals. The important high-symmetry points in the middle and at the edges of the first Brillouin zone (marked with solid lines) are shown. The periodicity of this lattice allows each point outside of the first Brillouin zone to be mapped within it. All the lattice nodes correspond to the Γ point. The triangles formed by the high-symmetry points and shown in the figures are irreducible zones. Each point outside this zone can be mapped to one inside it by periodicity and symmetry transformations.

[†] $\hat{\mathbf{x}}$ and $\hat{\mathbf{y}}$ are unit tensors of the Cartesian coordinate system.

be transformed using periodicity and symmetry transformations. As a consequence, analysis of the Bloch modes of a photonic crystal can be limited to this irreducible zone. In order to determine the bandgap, it is even sufficient to compute the frequencies for the wavevectors at the edges of that zone.

45.1.2 Photonic Bandgap

One of the most important properties of photonic crystals is the *photonic bandgap* (PBG), i.e., a range of frequencies in which light cannot propagate through the crystal [5]. If a light wave of such frequency were to propagate into the photonic crystal, it would be almost 100% reflected. This effect is well known in one-dimensional periodic stacks and is commonly used in distributed Bragg reflectors (DBRs, see Chapter 34). Figure 45.2a shows the reflectivity spectrum of a typical quarter-wavelength GaAs/AlAs DBR. The clearly visible high-reflectivity range between 927 nm and 1039 nm is the bandgap. The same effect in two dimensions can be seen in Figure 45.2b, which shows the optical bands as a horizontal wavevector to normalized frequency relation and the density of optical states within an infinite crystal. The bandgap is the region with zero density of states. In this frequency range, there is no band for any wavevector.

Two- and three-dimensional photonic crystals can be considered a multi-dimensional equivalent of DBRs. Hence, one of their major applications is to confine light. This is what they are most commonly used for in vertical-cavity surface-emitting lasers (VCSELs).

In order to confine light using a photonic crystal, one must create a *defect*: a localized perturbation of the periodic structure. Such perturbations can be either linear, in which case we have a waveguide, or isolated, in which case it makes a cavity. Examples of two-dimensional photonic crystal waveguides are shown in Figure 45.3. Because of the reflective properties of the photonic crystal within the bandgap, the waveguides can be bended with sharp angles and still effectively transmit light. In lasers, such waveguides can be coupled with edge-emitting lasers. In the case of surface-emitting lasers, the light is confined within a two-dimensional photonic crystal cavity. The most common configurations of such cavities are shown in Figure 45.4.

45.1.3 Photonic Crystals in Lasers

Photonic crystals are seldom used in edge-emitting lasers. One design of such a laser is presented in [6]. It is a W1 or W3 linear waveguide filled with an active material. Such lasers are reported to be intrinsically

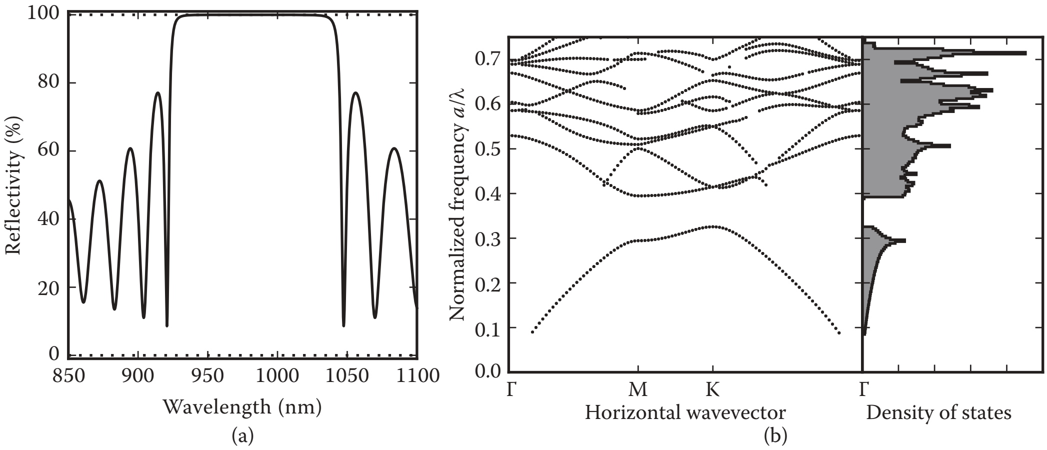


FIGURE 45.2 (a) Reflectivity of a quarter-wavelength one-dimensional stack and (b) band diagram and density of states (right histogram) of a two-dimensional photonic crystal. A photonic bandgap is visible in both cases as the frequency/wavelength region with very high reflectivity or vanishing density of states.

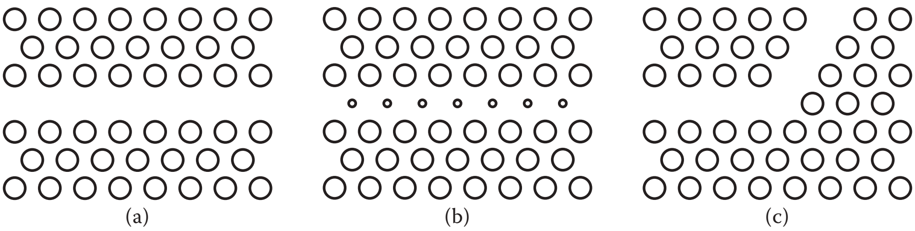

(a) (b) (c)

FIGURE 45.3 Examples of photonic crystal waveguides: (a) simple W1 waveguide with one row of missing holes, (b) linear waveguide with smaller holes, and (c) bent waveguide.

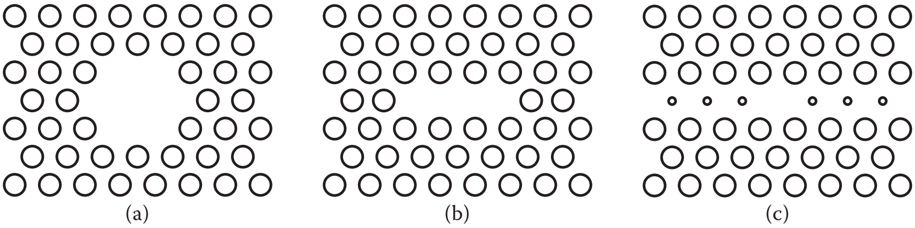

(a) (b) (c)

FIGURE 45.4 Examples of photonic crystal cavities: (a) cavity formed by removing seven holes, (b) L3 cavity with shifted edge holes, and (c) birefringent cavity formed by one missing hole and a row of smaller holes.

single-mode, in contrast to classical first-order DFB lasers. Depending on the configuration, the emitting mode has its wavevector either at the Γ or K point of the hexagonal photonic crystal lattice. The emission spectrum has two peaks, corresponding to the edges of the so-called mini-stop-band, which appears when two modes cross (as is the case at both Γ and K points). If emission takes place at frequencies that do not fall within the bandgap of the photonic crystal, only one of these modes is confined to the waveguide. The other spreads infinitely into the photonic crystal, and therefore suffers significantly larger losses. This is the main cause of single-mode operation.

In contrast to edge-emitting lasers, photonic crystals are much more commonly used in surface-emitting lasers. The first experimental designs were published at the end of the nineties. Similarly to photonic-crystal edge-emitting lasers (PC-EELs), such lasers were at first pumped with an optical beam and their design was not very complicated. The simplest surface-emitting photonic crystal laser is a PC membrane with a defect inside. The greatest advantage of such a design is its simplicity and the ease of fabricating a photonic crystal with a small lattice constant. For a typical hexagonal lattice of air holes, etched in a semiconductor material, the pitch must be around 0.3 λ if the frequency of the lasing mode c/λ is to be inside the bandgap. Because of the Purcell effect, the small size of the cavity helps to achieve a high spontaneous emission rate and a high Q-factor that can reach up to 10^4 [7].

In most designs of photonic crystal surface-emitting lasers [8–12], the light is not confined by the bandgap, but through the effective index of the defect area, which is higher than that of the surrounding photonic crystal. The role of the photonic crystal is to provide single-mode emission on either a higher order [9] or a fundamental mode [12]. Single-mode emission is one of the most promising applications of photonic crystals in the area of VCSEL design. This is especially true for high-power lasers [13], where classical designs fail because of the necessity of using broad area cavities.

The origin of a single-mode emission in PC-VCSELs can be attributed to one of the two factors. The first is a significant increase of modal losses for the higher-order modes, caused by light scattering. By carefully choosing the etching depth of the holes in the top DBR and their diameter, one can maximize this difference while keeping optical losses for the fundamental mode reasonably low [14,15]. The other factor is a cutoff similar to that found in so-called *endlessly single-mode* fibers [16], i.e., fibers that remain single-mode for any particular wavelength. The origin of this phenomena lies in the dependence of the PC effective index on the wavelength. The number of modes in a classical fiber depends on the parameter V_{eff}, which is defined as

$$V_{\mathrm{eff}} = \frac{2\pi r}{\lambda}\left(n_{\mathrm{core}}^2 - n_{\mathrm{clad}}^2\right)^{1/2}, \tag{45.4}$$

where r is the radius of the core, λ is the wavelength, and n_{core} and n_{clad} are the effective indexes of the fiber core and the cladding, respectively. The fiber is single-mode if $V_{\mathrm{eff}} < 2.405$. It is possible to design a photonic crystal fiber in which the V_{eff} does not exceed 2.405 even for $\lambda \to 0$. Because the VCSEL cavity can be approximated as a waveguide (although many phenomena are disregarded in such an approximation), Equation 45.4 can be used in this case. In practice, this means that single-mode operation is achieved if the diameters and depths of the holes in the photonic crystal are sufficiently small.

45.2 Fully Vectorial Method for Modeling VCSELs with Subwavelength Structures

Unlike classical oxide-confined VCSELs, discussed in Chapter 34, VCSELs incorporating subwavelength structures cannot be modeled using either the effective frequency method or any other approximate scalar method. The reason for this is the fact that such methods do not adequately consider light scattering from the subwavelength structures. Hence, in order to correctly model the optical properties of such devices, one needs to apply more exact vector methods, such as the finite-difference time-domain method [17], plane-wave expansion [18], the finite element method (FEM) [19], the coupled mode model (CMM) [20], the plane-wave admittance method (PWRT) [21], or plane-wave reflection transfer [22]. The greatest drawback of all these models is the increased computation time and computer memory usage. This cost is, however, necessary to bear, as only the rigorous solution of Maxwell equations provided by these methods allows light scattering from subwavelength structures to be included in the model.

In this chapter, we illustrate the fundamental optical properties of surface-emitting lasers incorporating subwavelength periodic structures, called photonic crystals. First, we need to sketch the principles of the rigorous vector methods used to obtain the results presented later in the chapter. Instead of describing the details of all popular vector models (some of which are discussed elsewhere in this book), we focus on the two that have proved to be the most useful for our purposes: the plane-wave admittance method and PWRT.

45.2.1 Plane-Wave Admittance Method

The plane-wave admittance method is an extension of the vector approach presented in Chapter 34. It uses the same procedure—admittance transfer—to find the eigenmodes of a resonant cavity. However, in this method, the electromagnetic field is expanded in the plane-wave Fourier basis. This basis is particularly useful for periodic structures, as it has natural periodic boundary conditions. Hence, modeling of infinite two-dimensional photonic crystals is straightforward and requires little effort. However, to analyze localized modes in lasers and correctly consider emitted radiation, one needs to introduce absorbing perfectly matched layers at the edges of the computational domain.

As the exact computation procedure was described in Chapter 34, it need not be repeated here. Instead, this section presents a derivation of the matrix form of Maxwell equations using the plane-wave expansion approach. Such matrices can be entered straightforwardly into computer memory for effective computations.

To start the derivation, consider an anisotropic linear dielectric or semiconductor material without any free charges ($\rho = 0$) or currents ($\mathbf{j} = 0$). In such a material, the time-independent Maxwell equations read

$$\nabla \times \mathbf{E} = -i\omega\mu\mu_0\mathbf{H}, \tag{45.5}$$

$$\nabla \times \mathbf{H} = \quad i\omega\varepsilon\varepsilon_0\mathbf{E}, \tag{45.6}$$

where $i^2 = -1$, ω is the angular frequency of light, and physical observables are the real parts of **E** and **H**.

Now consider an anisotropic material in Cartesian coordinates. Assuming that both permittivity and permeability of the material can be represented as diagonal tensors, the Equations 45.5 and 45.6 can be expanded as

$$-\partial_z E_y + \partial_y E_z = -i\omega\mu_x\mu_0 H_x, \tag{45.7}$$

$$\partial_x E_z - \partial_z E_x = -i\omega\mu_y\mu_0 H_y, \tag{45.8}$$

$$-\partial_y E_x + \partial_x E_y = -i\omega\mu_z\mu_0 H_z, \tag{45.9}$$

$$-\partial_z H_y + \partial_y H_z = i\omega\varepsilon_x\varepsilon_0 E_x, \tag{45.10}$$

$$\partial_x H_z - \partial_z H_x = i\omega\varepsilon_y\varepsilon_0 E_y, \tag{45.11}$$

$$-\partial_y H_x + \partial_x H_y = i\omega\varepsilon_z\varepsilon_0 E_z, \tag{45.12}$$

where ∂_ξ means a partial derivative in the ξ direction. These equations can be transformed into the form already introduced in Chapter 34:

$$\partial_z \begin{bmatrix} -E_y \\ E_x \end{bmatrix} = -i\frac{\eta_0}{k_0} \begin{bmatrix} \partial_y \varepsilon_z^{-1} \partial_y + \mu_x k_0^2 & -\partial_y \varepsilon_z^{-1} \partial_x \\ -\partial_x \varepsilon_z^{-1} \partial_y & \partial_x \varepsilon_z^{-1} \partial_x + \mu_y k_0^2 \end{bmatrix} \begin{bmatrix} H_x \\ H_y \end{bmatrix}, \tag{45.13}$$

$$\partial_z \begin{bmatrix} H_x \\ H_y \end{bmatrix} = -i\frac{1}{\eta_0 k_0} \begin{bmatrix} \partial_x \mu_z^{-1} \partial_x + \varepsilon_y k_0^2 & \partial_x \mu_z^{-1} \partial_y \\ \partial_y \mu_z^{-1} \partial_x & \partial_y \mu_z^{-1} \partial_y + \varepsilon_x k_0^2 \end{bmatrix} \begin{bmatrix} -E_y \\ E_x \end{bmatrix}, \tag{45.14}$$

where $k_0 = \omega/c = \omega\left(\mu_0\varepsilon_0\right)^{1/2}$.

Now, consider an electromagnetic field in a structure exhibiting two-dimensional periodicity, i.e.,

$$\varepsilon(\mathbf{r}) = \varepsilon(\mathbf{r} + \mathbf{a}_i) \qquad (i = 1, 2), \tag{45.15}$$

$$\mu(\mathbf{r}) = \mu(\mathbf{r} + \mathbf{a}_i) \qquad (i = 1, 2), \tag{45.16}$$

with $\mathbf{a}_1$ and $\mathbf{a}_2$ being elementary lattice vectors in the xy plane. Because of their periodicity, the solutions of the Maxwell equations obey the Bloch theorem, i.e., both the electric and magnetic fields can be represented as

$$\mathbf{E}(\mathbf{r}) = \bar{\mathbf{E}}(\mathbf{r}) \exp(-i\,\mathbf{k}_{xy} \cdot \mathbf{r}), \tag{45.17}$$

$$\mathbf{H}(\mathbf{r}) = \bar{\mathbf{H}}(\mathbf{r}) \exp(-i\,\mathbf{k}_{xy} \cdot \mathbf{r}), \tag{45.18}$$

where $\bar{\mathbf{E}}(\mathbf{r})$, $\bar{\mathbf{H}}(\mathbf{r})$ are periodic functions and $\mathbf{k}_{xy}$ is a projection of the mode wavevector into the xy plane. Furthermore, the functions $\bar{\mathbf{E}}(\mathbf{r})$ and $\bar{\mathbf{H}}(\mathbf{r})$ can be expanded into the two-dimensional

Fourier (plane-wave) series as

$$\bar{\mathbf{E}}(\mathbf{r}) = \mathbf{E}^{\mathbf{g}} \exp(i\,\mathbf{g} \cdot \mathbf{r}) = \mathbf{E}^{\mathbf{g}} \left|\varphi_{\mathbf{g}}\right\rangle, \tag{45.19}$$

$$\bar{\mathbf{H}}(\mathbf{r}) = \mathbf{H}^{\mathbf{g}} \exp(i\,\mathbf{g} \cdot \mathbf{r}) = \mathbf{H}^{\mathbf{g}} \left|\varphi_{\mathbf{g}}\right\rangle, \tag{45.20}$$

where the Einstein summation convention[†] is used and **g** is the reciprocal lattice vector:

$$\mathbf{g} = l_1\mathbf{b}_1 + l_2\mathbf{b}_2, \tag{45.21}$$

$$\mathbf{a}_i \cdot \mathbf{b}_j = 2\pi\delta_{ij}, \tag{45.22}$$

with l_1 and l_2 being arbitrary integers and δ_{ij} the Kronecker's delta. As we perform Fourier expansion only in the xy plane, the coefficients $\mathbf{E}^{\mathbf{g}}$ and $\mathbf{H}^{\mathbf{g}}$ are functions of z.

Using the Bloch theorem together with Equations 45.19 and 45.20, we can represent both fields as

$$\mathbf{E}(\mathbf{r}) = \mathbf{E}^{\mathbf{g}} \left|\varphi_{\mathbf{g}-\mathbf{k}_{xy}}\right\rangle, \tag{45.23}$$

$$\mathbf{H}(\mathbf{r}) = \mathbf{H}^{\mathbf{g}} \left|\varphi_{\mathbf{g}-\mathbf{k}_{xy}}\right\rangle. \tag{45.24}$$

Now consider Equations 45.13 and 45.14. As the Fourier basis is orthonormal, i.e.,

$$\langle\varphi_{\mathbf{g}}|\varphi_{\mathbf{g}'}\rangle = \delta_{\mathbf{g}\mathbf{g}'}, \tag{45.25}$$

we can introduce Equations 45.23 and 45.24 into Equations 45.13 and 45.14 and left-multiply them by $\langle\varphi_{\mathbf{g}-\mathbf{k}_{xy}}|$ to obtain

$$\partial_z \begin{bmatrix} E_y^{\mathbf{g}} \\ E_x^{\mathbf{g}} \end{bmatrix} = -i\frac{\eta_0}{k_0} \left\langle \varphi_{\mathbf{g}-\mathbf{k}_{xy}} \left| \begin{matrix} \partial_y\varepsilon_z^{-1}\partial_y + \mu_x k_0^2 & -\partial_y\varepsilon_z^{-1}\partial_x \\ -\partial_x\varepsilon_z^{-1}\partial_y & \partial_x\varepsilon_z^{-1}\partial_x + \mu_y k_0^2 \end{matrix} \right| \varphi_{\mathbf{g}'-\mathbf{k}_{xy}} \right\rangle \begin{bmatrix} H_x^{\mathbf{g}'} \\ H_y^{\mathbf{g}'} \end{bmatrix}, \tag{45.26}$$

$$\partial_z \begin{bmatrix} H_x^{\mathbf{g}'} \\ H_y^{\mathbf{g}'} \end{bmatrix} = -\frac{i}{\eta_0 k_0} \left\langle \varphi_{\mathbf{g}-\mathbf{k}_{xy}} \left| \begin{matrix} \partial_x\mu_z^{-1}\partial_x + \varepsilon_y k_0^2 & \partial_x\mu_z^{-1}\partial_y \\ \partial_y\mu_z^{-1}\partial_x & \partial_y\mu_z^{-1}\partial_y + \varepsilon_x k_0^2 \end{matrix} \right| \varphi_{\mathbf{g}'-\mathbf{k}_{xy}} \right\rangle \begin{bmatrix} E_y^{\mathbf{g}} \\ E_x^{\mathbf{g}} \end{bmatrix}, \tag{45.27}$$

where, in order to simplify notation, we represent $E_y(\mathbf{r})$ as

$$E_y(\mathbf{r}) = -E_y^{\mathbf{g}} \left|\varphi_{\mathbf{g}-\mathbf{k}_{xy}}\right\rangle \tag{45.28}$$

instead of directly using Equation 45.23.

The matrices in the above equations can be computed easily when both permittivity and permeability are expanded in the Fourier basis. For this purpose, the analyzed layer is assumed to be invariant in the

---

[†] By this convention, if an index is repeated twice in any product, we perform summation over this index. We use this convention in this whole chapter, unless explicitly stated otherwise.

z-direction. Then, we can write

$$\varepsilon_i(\mathbf{r}) = \varepsilon_i^{\mathbf{g}} \left|\varphi_{\mathbf{g}}\right\rangle \quad (i = x, y), \tag{45.29}$$

$$\mu_i(\mathbf{r}) = \mu_i^{\mathbf{g}} \left|\varphi_{\mathbf{g}}\right\rangle \quad (i = x, y), \tag{45.30}$$

$$\varepsilon_z^{-1}(\mathbf{r}) = \kappa^{\mathbf{g}} \left|\varphi_{\mathbf{g}}\right\rangle, \tag{45.31}$$

$$\mu_z^{-1}(\mathbf{r}) = \gamma^{\mathbf{g}} \left|\varphi_{\mathbf{g}}\right\rangle. \tag{45.32}$$

Substituting this into Equations 45.26 and 45.27, we finally obtain the matrix form of the equations in the plane-wave basis:

$$\partial_z \begin{bmatrix} E_y^{\mathbf{g}} \\ E_x^{\mathbf{g}} \end{bmatrix} = -i\frac{\eta_0}{k_0} \begin{bmatrix} -(g_y - k_y)(g_y' - k_y)\kappa^{\mathbf{g}-\mathbf{g}'} + k_0^2\mu_x^{\mathbf{g}-\mathbf{g}'} & (g_y - k_y)(g_x' - k_x)\kappa^{\mathbf{g}-\mathbf{g}'} \\ (g_x - k_x)(g_y' - k_y)\kappa^{\mathbf{g}-\mathbf{g}'} & -(g_x - k_x)(g_x' - k_x)\kappa^{\mathbf{g}-\mathbf{g}'} + k_0^2\mu_y^{\mathbf{g}-\mathbf{g}'} \end{bmatrix} \begin{bmatrix} H_x^{\mathbf{g}'} \\ H_y^{\mathbf{g}'} \end{bmatrix}, \tag{45.33}$$

$$\partial_z \begin{bmatrix} H_x^{\mathbf{g}} \\ H_y^{\mathbf{g}} \end{bmatrix} = -\frac{i}{\eta_0 k_0} \begin{bmatrix} -(g_x - k_x)(g_x' - k_x)\gamma^{\mathbf{g}-\mathbf{g}'} + k_0^2\varepsilon_y^{\mathbf{g}-\mathbf{g}'} & -(g_x' - k_x)(g_y' - k_y)\gamma^{\mathbf{g}-\mathbf{g}'} \\ -(g_y - k_y)(g_x' - k_x)\gamma^{\mathbf{g}-\mathbf{g}'} & -(g_y - k_y)(g_y' - k_y)\gamma^{\mathbf{g}-\mathbf{g}'} + k_0^2\varepsilon_x^{\mathbf{g}-\mathbf{g}'} \end{bmatrix} \begin{bmatrix} E_y^{\mathbf{g}'} \\ E_x^{\mathbf{g}'} \end{bmatrix}, \tag{45.34}$$

where k_x, k_y, g_x, and g_y are the corresponding components of the mode wavevector and the reciprocal lattice vector **g**, respectively. These equations can be introduced into computer memory for further computations by truncating the Fourier basis $\{\mathbf{g}\}$ at some point. The introduced error will depend on two factors—the distribution of the electromagnetic field and the distribution of the material parameters. The former, as a result of simulation, cannot be influenced directly, although the latter can be convoluted with a Gaussian window to improve convergence.

The vertical components of electric and magnetic fields can also be expanded in the Fourier basis as

$$E_z^{\mathbf{g}} = \frac{\eta_0}{k_0}\kappa^{\mathbf{g}-\mathbf{g}'} \left[-(g_y' - k_y)H_x^{\mathbf{g}'} + (g_x' - k_x)H_y^{\mathbf{g}'}\right], \tag{45.35}$$

$$H_z^{\mathbf{g}} = \frac{1}{k_0\eta_0}\gamma^{\mathbf{g}-\mathbf{g}'} \left[(g_x' - k_x)E_y^{\mathbf{g}'} + (g_y' - k_y)E_x^{\mathbf{g}'}\right]. \tag{45.36}$$

Once Equations 45.33 and 45.34 have been derived and represented in their numerical forms, we can use them to determine the eigenmodes of the analyzed structure. For this purpose, we first show their solution for a single z-invariant layer and then present a numerically stable technique for combining fields in all layers and deriving an eigenmode condition for the whole structure. The procedure described below was originally used with the method of lines [23] and is also applied to the plane-wave admittance method.

Further in this section, we will use the short representation of the vectors and matrices introduced in Equations 45.33 and 45.34. In particular, we will name

$$\hat{\mathbf{E}} = \begin{bmatrix} E_y^{\mathbf{g}} \\ E_x^{\mathbf{g}} \end{bmatrix},$$

$$\hat{\mathbf{H}} = \begin{bmatrix} H_x^{\mathbf{g}} \\ H_y^{\mathbf{g}} \end{bmatrix},$$

$$\mathbf{R_E} = \frac{1}{\eta_0 k_0}\begin{bmatrix} -(g_x - k_x)(g'_x - k_x)\gamma^{\mathbf{g}-\mathbf{g}'} + k_0^2\varepsilon_y^{\mathbf{g}-\mathbf{g}'} & -(g'_x - k_x)(g'_y - k_y)\gamma^{\mathbf{g}-\mathbf{g}'} \\ -(g_y - k_y)(g'_x - k_x)\gamma^{\mathbf{g}-\mathbf{g}'} & -(g_y - k_y)(g'_y - k_y)\gamma^{\mathbf{g}-\mathbf{g}'} + k_0^2\varepsilon_x^{\mathbf{g}-\mathbf{g}'} \end{bmatrix},$$

$$\mathbf{R_H} = \frac{\eta_0}{k_0}\begin{bmatrix} -(g_y - k_y)(g'_y - k_y)\kappa^{\mathbf{g}-\mathbf{g}'} + k_0^2\mu_x^{\mathbf{g}-\mathbf{g}'} & (g_y - k_y)(g'_x - k_x)\kappa^{\mathbf{g}-\mathbf{g}'} \\ (g_x - k_x)(g'_y - k_y)\kappa^{\mathbf{g}-\mathbf{g}'} & -(g_x - k_x)(g'_x - k_x)\kappa^{\mathbf{g}-\mathbf{g}'} + k_0^2\mu_y^{\mathbf{g}-\mathbf{g}'} \end{bmatrix}.$$

It is worth noting that the whole procedure described below can be applied with any correct representation of the vectors and matrices above. Thus, it is applicable not only to plane-wave expansion, but also to any orthogonal or non-orthogonal basis. In particular, the representation of $\hat{\mathbf{E}}$, $\hat{\mathbf{H}}$, $\mathbf{R_E}$, and $\mathbf{R_H}$ by finite-difference approximation is used in the method of lines.

Take Equations 45.33 and 45.34 and represent them in their short form:

$$\partial_z \hat{\mathbf{E}} = -i\mathbf{R_H}\hat{\mathbf{H}}, \tag{45.37}$$

$$\partial_z \hat{\mathbf{H}} = -i\mathbf{R_E}\hat{\mathbf{E}}. \tag{45.38}$$

By taking the z-derivative of one of these equations and substituting the other into it, one can write two second-order equations with the fields decoupled:

$$\partial_z^2 \hat{\mathbf{E}} = -\mathbf{Q_E}\hat{\mathbf{E}}, \tag{45.39}$$

$$\partial_z^2 \hat{\mathbf{H}} = -\mathbf{Q_H}\hat{\mathbf{H}}, \tag{45.40}$$

where $\mathbf{Q_E} = \mathbf{R_H}\mathbf{R_E}$ and $\mathbf{Q_H} = \mathbf{R_E}\mathbf{R_H}$.

The eigenmodes can be found using the procedure described in Chapter 34 by constructing and solving equation $\mathbf{Y}\bar{\mathbf{E}} = 0$.

45.2.2 Plane-Wave Reflection Transfer

PWRT differs from the plane-wave admittance method in the way the diagonalized Equations 45.33 and 45.34 are solved. In this method, the electric field in each layer uniform in z-direction is chosen to have the form

$$\tilde{\mathbf{E}} = \exp(-i\boldsymbol{\Gamma}z)\tilde{\mathbf{F}} + \exp(i\boldsymbol{\Gamma}z)\tilde{\mathbf{B}}, \tag{45.41}$$

where $\tilde{\mathbf{F}}$ denotes the amplitude of the forward propagating wave and $\tilde{\mathbf{B}}$ the amplitude of the backward propagating wave. The reflection transfer means an iterative determination of the reflection matrix $\mathbf{R}$, which is defined in each layer as a relation between $\tilde{\mathbf{F}}$ and $\tilde{\mathbf{B}}$ such that

$$\tilde{\mathbf{F}} = \mathbf{R}\,\tilde{\mathbf{B}}. \tag{45.42}$$

We assume that in the first layer, which has infinite thickness, there is no in-going field (we will introduce any optional incident field in the last layer). Thus, $\tilde{\mathbf{F}} \equiv 0$ and so $\mathbf{R} = \mathbf{0}$. The reflection matrices in the subsequent layers can be determined in an iterative process using the requirement that the electric and magnetic fields are continuous at layer interfaces.

By careful selection of units in the magnetic field, it is possible to show that in the diagonalized domain it takes the form of

$$\tilde{\mathbf{H}} = \exp(-i\boldsymbol{\Gamma}z)\tilde{\mathbf{F}} - \exp(i\boldsymbol{\Gamma}z)\tilde{\mathbf{B}}, \tag{45.43}$$

which can be derived from Equation 45.41. The requirement of field continuity between the n-th and $(n+1)$th layers can be expressed as

$$\mathbf{T_E}^n \left\{ \boldsymbol{\varphi}^n \tilde{\mathbf{F}}^n + \left(\boldsymbol{\varphi}^n\right)^{-1} \tilde{\mathbf{B}}^n \right\} = \mathbf{T_E}^{n+1} \left\{ \tilde{\mathbf{F}}^{n+1} + \tilde{\mathbf{B}}^{n+1} \right\}, \tag{45.44}$$

$$\mathbf{T_H}^n \left\{ \boldsymbol{\varphi}^n \tilde{\mathbf{F}}^n - \left(\boldsymbol{\varphi}^n\right)^{-1} \tilde{\mathbf{B}}^n \right\} = \mathbf{T_H}^{n+1} \left\{ \tilde{\mathbf{F}}^{n+1} - \tilde{\mathbf{B}}^{n+1} \right\}, \tag{45.45}$$

where $\boldsymbol{\varphi}^n = \exp\left(-i\boldsymbol{\Gamma} d_n\right)$ and d_n is the thickness of nth layer. Using the relation (Equation 45.42), we obtain

$$\tilde{\mathbf{F}}^{n+1} + \tilde{\mathbf{B}}^{n+1} = \mathbb{E}^n \left(\boldsymbol{\varphi}^n\right)^{-1} \tilde{\mathbf{B}}^n, \tag{45.46}$$

$$\tilde{\mathbf{F}}^{n+1} - \tilde{\mathbf{B}}^{n+1} = \mathbb{H}^n \left(\boldsymbol{\varphi}^n\right)^{-1} \tilde{\mathbf{B}}^n, \tag{45.47}$$

where

$$\mathbb{E}^n = \left(\mathbf{T_E}^{n+1}\right)^{-1} \mathbf{T_E}^n \left\{ \boldsymbol{\varphi}^n \mathbf{R}^n \boldsymbol{\varphi}^n + \mathbf{I} \right\}, \tag{45.48}$$

$$\mathbb{H}^n = \left(\mathbf{T_H}^{n+1}\right)^{-1} \mathbf{T_H}^n \left\{ \boldsymbol{\varphi}^n \mathbf{R}^n \boldsymbol{\varphi}^n - \mathbf{I} \right\}. \tag{45.49}$$

This, after simple algebra, gives the reflection matrix for the $(n+1)$th layer in the form

$$\mathbf{R}^{n+1} = \left[\mathbb{E}^n + \mathbb{H}^n\right] \left[\mathbb{E}^n - \mathbb{H}^n\right]^{-1}. \tag{45.50}$$

In contrast to plane-wave admittance transfer, we need not assume the electric field at the boundaries is equal to zero and, hence, there is no need to use absorbing perfectly matched layers as vertical boundary conditions. Instead, at the edge of the structure, we can assume the $\mathbf{R}^0$ matrix to be equal to 0. This means that in the outermost layer, there is no forward propagating field. This field corresponds to the wave coming into the structure. As the analyzed device is a laser resonant cavity, it only emits light; hence, $\tilde{\mathbf{F}}^0 = \mathbf{0}$ and $\tilde{\mathbf{B}}^0 \neq \mathbf{0}$. Then, similarly to PWAM, we can perform iterative progressions of the reflection matrix from both top and bottom boundaries, up to a selected layer. In this, the following relations hold:

$$\tilde{\mathbf{F}}^{\text{top}} = \tilde{\mathbf{B}}^{\text{bottom}}, \tag{45.51}$$

$$\tilde{\mathbf{B}}^{\text{top}} = \tilde{\mathbf{F}}^{\text{bottom}}. \tag{45.52}$$

Using this condition, it is possible to construct a matching matrix $\mathbf{M}$ such that

$$\mathbf{M}\tilde{\mathbf{F}}^{\text{bottom}} = \left[\mathbf{R}^{\text{top}} - \left(\mathbf{R}^{\text{bottom}}\right)^{-1}\right] \tilde{\mathbf{F}}^{\text{bottom}} = \mathbf{0}.$$

In other words, this matrix must be singular, which is the condition for finding the eigenmodes.

45.3 VCSELs with Photonic Crystals for Light Confinement

In VCSELs, photonic crystals are used to provide effective light confinement. By etching a regular lattice of holes in either the DBRs or the VCSEL cavity (or both), together with intentional lattice defects, it is possible to mold the optical mode distribution. In particular, one can laterally confine the fundamental laser mode within a photonic crystal defect without adding any more lowindex oxidation layers. There are several advantages of this approach: First, it allows the fabrication of VCSELs with well-defined cavities in material systems where selective oxidation is either impossible or difficult to achieve. Second, nonuniform patterns in the periodic lattice can be tuned to interact differently with fundamental and high-order modes, providing increased mode selectivity and allowing single-mode VCSEL operation at higher currents (see Section 45.1.3).

To illustrate the exclusive optical properties of PC-VCSELs, consider a basic optical-only VCSEL structure made of GaAs, as shown in Figure 45.5a. It contains 24 GaAs/AlAs pairs of top DBRs, 29 pairs of GaAs/AlAs bottom DBRs, and a very short single-wavelength GaAs cavity. The exact layer structure is listed in Table 45.1. The photonic crystal is based on a triangular lattice (of pitch L) and takes the form of three hexagons surrounding a small cavity in the form of a missing PC hole. Holes with diameter d are etched either in the top DBRs (with varying etching depths), throughout the whole structure, or only inside the cavity.

We now investigate the impact of photonic crystal parameters on the properties of the fundamental mode in the VCSEL in question. Due to its Gaussian-like shape, this mode is the most preferred in most applications. Furthermore, VCSELs with incorporated photonic crystals are known for their high mode selectivity [24], which makes it possible to tune them into a single-mode source with high power emission.

Our results are depicted in Figures 45.6 through 45.9. They express typical tendencies observed in PC-VCSELs. Figure 45.6 shows that increasing the size of the holes results in a decrease in the resonant wavelength, which can be explained by two factors. The first is the decrease in the area of the photonic crystal defect, which forms the cavity (corresponding to a decrease in the aperture of a classical VCSEL).

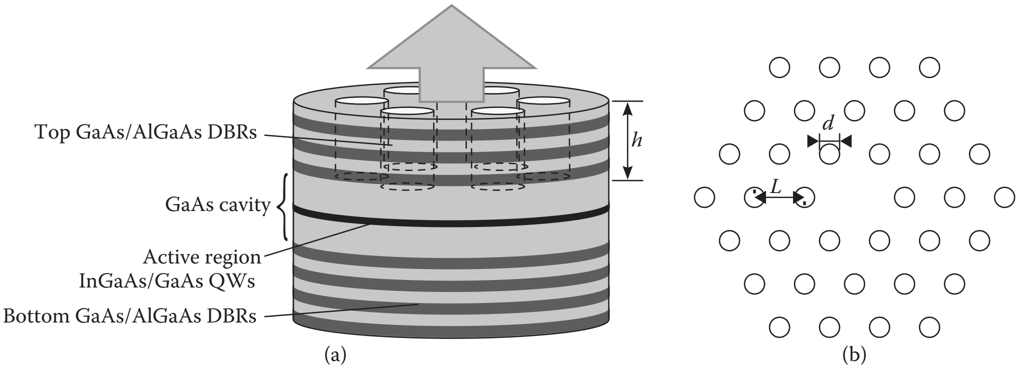


FIGURE 45.5 Analyzed VCSEL structure: (a) schematic three-dimensional view, (b) PC structure. (DBRs, distributed Bragg reflectors; QWs, quantum wells.)

TABLE 45.1 Details of the Layer Structure of the Analyzed VCSEL.

	Thickness [nm]	Material	Refractive index
		Air	1.00
Top DBR	69.40	GaAs	3.53
24 pairs	79.55	AlGaAs	3.08
	121.71	GaAs	3.53
Cavity	$3 \times 8.00 + 2 \times 5.00$	QW/GaAs	$\begin{cases} 3.56 + j\,n_g & \text{for } r < a \\ 3.56 - 0.01j & \text{for } r \geq a \end{cases} \Big/ \ 3.53$
	121.71	GaAs	3.53
Bottom DBR	79.55	AlGaAs	3.08
29 pairs	69.40	GaAs	3.53
Substrate		GaAs	3.53

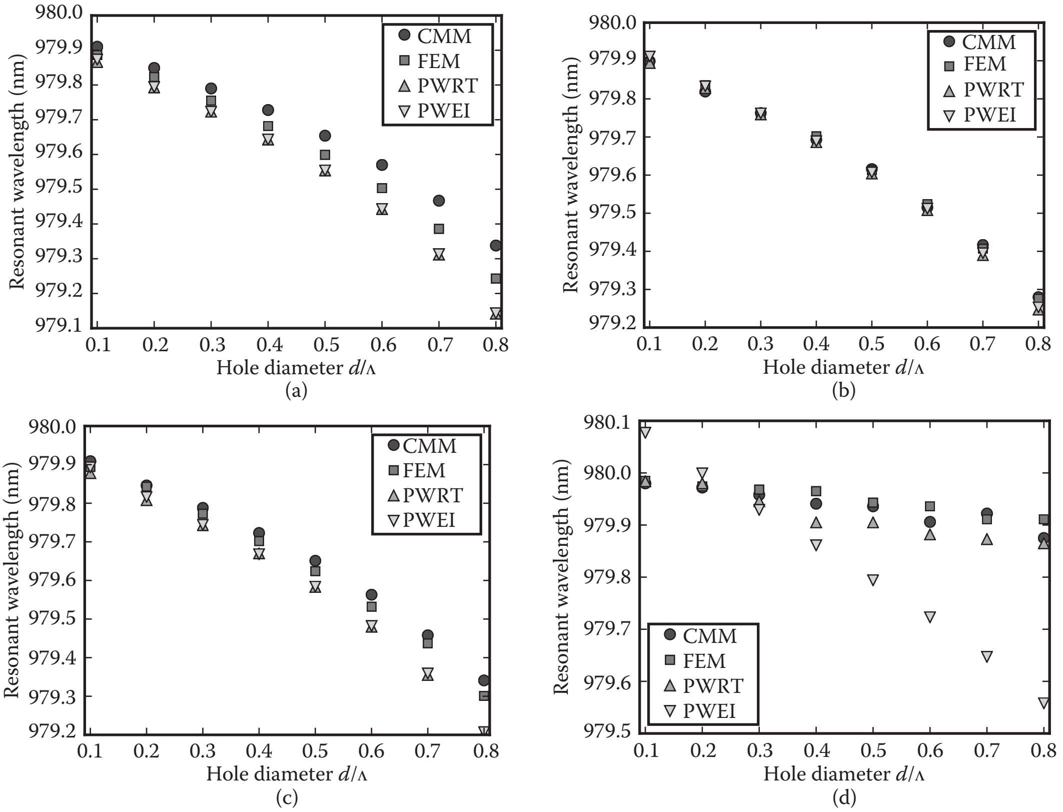


FIGURE 45.6 Resonant wavelengths of PC-VCSELs with photonic crystals etched in different parts of the structure ((a) whole structure, (b) cavity only, (c) 24 pairs of the top DBR, (d) 10 pairs of the top DBR) as a function of hole diameters *d*. The photonic crystal pitch Λ is 4 μm in all the plots.

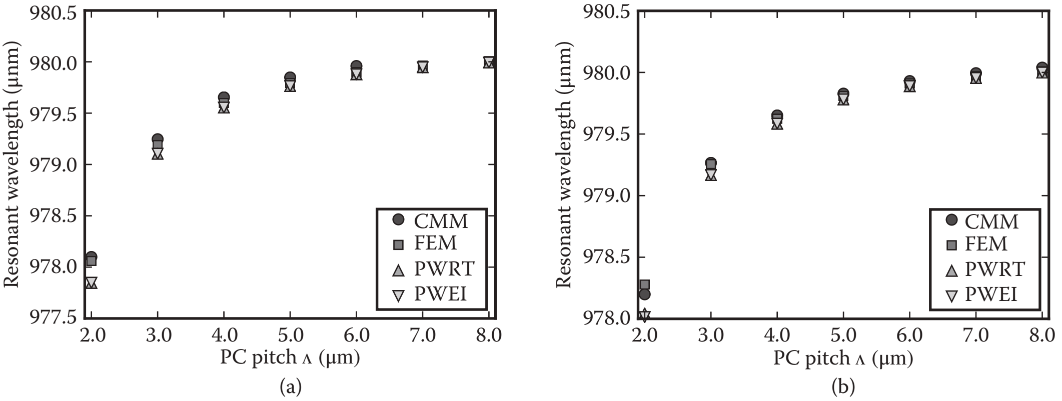
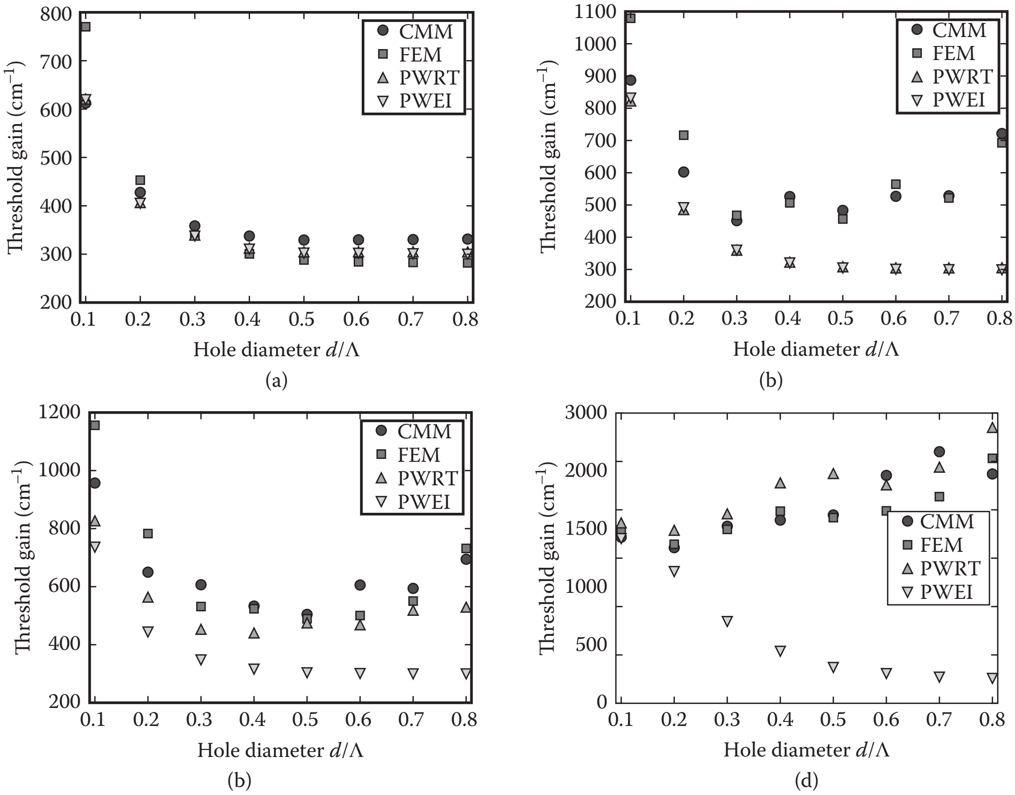


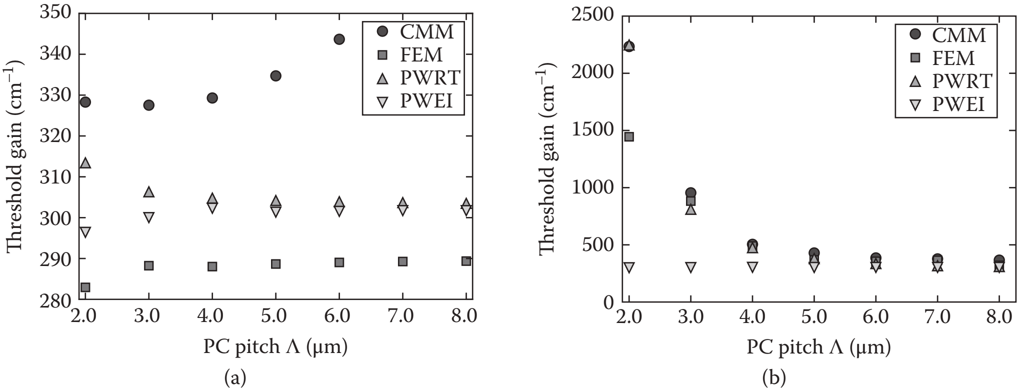
FIGURE 45.7 Resonant wavelengths of PC-VCSELs with photonic crystals etched in different parts of the structure ((a) whole structure, (b) cavity only, (c) 24 pairs of the top DBR, (d) 10 pairs of the top DBR) as a function of the photonic crystal pitch Λ. The diameter of the etched holes *d* is proportional to the pitch and equals 0.5 Λ.



(a) (b)

(b) (d)

FIGURE 45.8 Threshold gains of PC-VCSELs with photonic crystals etched in different parts of the structure ((a) whole structure, (b) cavity only, (c) 24 pairs of the top DBR, (d) 10 pairs of the top DBR) as a function of the hole diameters *d*. The photonic crystal pitch Λ is 4 μm in all the plots.



(a) (b)

FIGURE 45.9 Threshold gains of PC-VCSELs with photonic crystals etched in different parts of the structure ((a) whole structure, (b) cavity only, (c) 24 pairs of the top DBR, (d) 10 pairs of the top DBR) as a function of the photonic crystal pitch Λ. The diameter of the etched holes *d* is proportional to the pitch and equals 0.5 Λ.

The other factor is the increase of the effective refractive index contrast between the cavity and the photonic crystal surrounding it. Larger holes mean a larger average air-to-semiconductor ratio, which results in a lower effective index outside the cavity.

The opposite tendency is shown in Figure 45.7. Increasing the photonic crystal pitch results in a rapid increase in the resonant wavelength, as the cavity area increases in proportion to Λ. Due to the constant d/Λ ratio, the effective index contrast remains unchanged. The change is fastest for smaller pitches, where the light confinement is high. With increasing Λ, it increases asymptotically to 980 nm, which is the resonant wavelength of the presented VCSEL with no lateral confinement at all.

Much more interesting is the behavior of the threshold gain. In each case, it is large for small holes (Figure 45.8) because of the weak confinement and large overlap of the mode with the absorbing areas of the active region outside the gain aperture. As the size of the holes increases, the mode is increasingly better confined and therefore the threshold decreases. However, with shallow etching (Figure 45.8b through d), scattering starts to play an important role, and with larger holes the threshold starts increasing again. On the other hand, the photonic crystal pitch does not influence the threshold gain significantly (Figure 45.9), except when it drops too strongly (below 4 μm) in structures with holes etched only in the DBR (Figure 45.9b), in which case light scattering starts to play an important role.

In each case, the same optical properties were computed using four different numerical methods: PWRT presented above, the CMM [20], the FEM [25], and the plane-wave effective index (PWEI) [26]. These methods have already been rigorously compared [27]. Here, we show them to emphasize that the three-dimensional structure of a PC-VCSEL is complex to analyze numerically, even nowadays. Practically all of the methods have shown behavior qualitatively different to the others in at least one case. This shows how important it is to calibrate numerical models and to choose the right one, which will yield correct results for a particular case.

In most circumstances, however, numerical analysis allows some conclusions to be made regarding the optical properties of PC-VCSELs. The results depicted in Figures 45.6 and 45.8 show computed resonant wavelengths and threshold gains as functions of the hole diameters when the photonic crystal pitch Λ equals 4 μm. A pitch of 4 μm is a typical value and has been kept fixed for all etchings. Figures 45.7 and 45.9 show the results obtained from different sizes of the photonic crystal lattice constant, with holes etched through the whole cavity and all layers of the top DBR.

In most cases, the PC-VCSEL resonant wavelength is predicted consistently by all numerical methods. The only exception is with large, shallow etched holes, as can be seen in Figure 45.6d. The differences between the vector methods (CMM, FEM, and PWRT) and the scalar method (PWEI), which gives a too small wavelength, are clearly noticeable. The origin of these discrepancies is the significant light scattering at the bottoms of the holes, which is correctly considered by the vector methods, but neglected by the scalar method.

Unsurprisingly, greater differences are observable in the threshold gain data. The methods show good agreement only in case of the $\Lambda = 4$ μm wholly etched structure. All other situations require more detailed analysis. Figure 45.8c shows the threshold gains for holes etched only in the top DBR. Due to their complicated arrangement, significant scattering occurs at their edges, which is reflected in an increase in the threshold gain for $d/\Lambda > 0.4$. This effect is not predicted by PWEI, which—as a variation of the effective index method—is incapable of properly considering scattering losses. PWEI even more significantly underestimates the threshold gain for shallow holes (Figure 45.6d), where scattering is much stronger due to the weaker mode confinement.

45.4 Subwavelength Gratings in VCSELs

In the first part of this chapter, we described the two-dimensional photonic crystals used to provide light confinement of the optical mode. However, periodic subwavelength structures can offer much more. In what follows, we will look at specific one-dimensional photonic crystals: subwavelength gratings with high refractive index contrast (referred to as high-contrast gratings or HCGs). Such structures may effectively replace DBR mirrors in VCSELs, as they can provide very high reflectivity, while being significantly thinner than a multilayer DBR stack.

Because the light emitted by most VCSELs is linearly polarized, one-dimensional linear gratings can be efficiently applied. Furthermore, they ensure stable polarization of the emitted light, which otherwise might suddenly change at higher output powers [28]. For the same reason, circular gratings are much less suitable as mirror replacements for VCSELs, as they cannot provide uniform reflectivity for linearly polarized light.

In order to investigate the properties of VCSELs with subwavelength gratings, we first need to understand the behavior of grating mirrors. For this reason, in the next subsection, we will outline the basic physical principles behind the high reflectivity of HCGs. Next, we will study their reflectivity in various configurations. After that, we will move on to the determination of resonant modes in VCSEL cavities with subwavelength grating.

45.4.1 High-Contrast Subwavelength Grating Theory

Despite their very slight thicknesses—usually only a few hundreds of nanometers—HCGs can be effective mirrors with reflectivity comparable or exceeding that of thick multilayered DBRs. This behavior can be explained in two ways: either as an effect of the Fano resonance between the incident wave and the excited mode propagating inside the grating [29] or as an effect of destructive interference between transmitted and leaky modes [30]. Below, we briefly summarize the latter explanation.

Consider a grating suspended in air. It can be seen as a periodic array of waveguides, propagating light in a direction perpendicular to the grating plane (Figure 45.10). Assume an incident plane wave propagating in the z-direction and polarized parallel to the grating bars (with the nonzero E_y component). At the incidence side of the grating (region 1 in Figure 45.10), the electric field can be described as the single zero-order incident plane wave and multiple orders of the reflected wave:

$$E_1(x,z) = E_I \exp\left(-ik_0 z\right) + \sum_n E_R^n \cos\left(2n\pi x/L\right) \exp\left(+i\sqrt{k_0^2 - (2n\pi/L)^2}\,z\right), \tag{45.53}$$

where E_I is the amplitude of the incident field, E_R^n is the horizontal distribution of the reflected modes of different orders, $k_0 = \omega/c$ is the normalized frequency, and L is the grating period. Inside the grating, the field is the sum of forward and backward propagating waveguide array modes:

$$E_2(x,z) = \sum_m \left[E_F^m(x) \exp\left(-i\beta_m z\right) + E_B^m(x) \exp\left(+i\beta_m z\right)\right], \tag{45.54}$$

where β_m is the propagation constant of the mth mode and E_F^m and E_B^m are amplitudes of the forward and backward propagating modes, respectively. Finally, in the last section, the field consists only of the

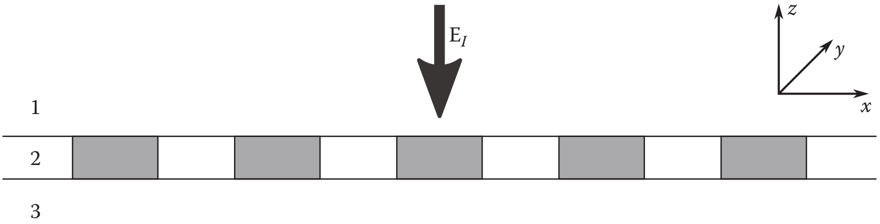


FIGURE 45.10 Grating as a periodic waveguide array.

transmitted modes with amplitudes E_T^n:

$$E_3(x, z) = \sum_n E_T^n \cos(2n\pi x/L) \exp\left(-i\sqrt{k_0^2 - (2n\pi/L)^2} z\right).$$

For subwavelength gratings $L < \lambda = 2\pi/k_0$, only the zero orders of both the reflected and transmitted modes propagate and the other modes are evanescent. Although they do not transport any energy, they are very important for ensuring proper boundary conditions for Maxwell equations. By applying these boundary conditions to both grating boundaries, it is possible to determine the amplitudes of both the reflected and transmitted modes (E_R^n and E_T^n) and the modes inside the grating (E_F^m and E_B^m). As only one transmitted mode (E_T^0) propagates energy, it is possible to tune the grating parameters in such a way that $E_T^0 = 0$. This corresponds to the zero energy transmission equivalent, as there is no gain inside the grating, up to 100% reflectivity.

For gratings operating in a near-subwavelength regime, the periodic waveguide array supports only two even modes with propagation constants β_0 and β_2 [31]. In such a case, E_T^0 can be derived analytically as

$$E_T^0 = \frac{1}{L}\int_0^L \left\{E_F^0(x) + E_B^0(x)\right\} dx + \frac{1}{L}\int_0^L \left\{E_F^2(x) + E_B^2(x)\right\} dx.$$

In other words, E_T^0 is a sum of the spatial averages of both waveguide array modes. If these averages are antimatched, i.e.,

$$\int_0^L \left\{E_F^0(x) + E_B^0(x)\right\} dx = -\int_0^L \left\{E_F^2(x) + E_B^2(x)\right\} dx,$$

there is no energy transmission through the grating. This corresponds to the destructive interference of these modes at the far edge of the grating.

For a fuller derivation and discussion of the modal properties of subwavelength gratings, we refer the reader to [31], where the interaction between gratings and light waves is described in detail.

45.4.2 Modeling of Subwavelength Grating Reflectivity

The reflectivity of subwavelength HCGs can be studied using the PWRT method presented at the beginning of this chapter. This method is mathematically equivalent to the popular rigorous coupled-wave analysis (RCWA) [32]. Both methods rely on expansion of the electromagnetic field into a complex Fourier series and both require the analyzed structure to be split into a set of layers, each of which is uniform in the surface normal direction (see Figure 45.11). PWRT, which was used to obtain all the results presented in this chapter, is applied in the same way as for cavity analysis, as described in Section 45.2.2 with the last step omitted. Instead, after determining the reflection matrix $\bar{\mathbf{R}}$, this is multiplied by a predefined incidence vector $\mathbf{F}$ to calculate the vector of reflected diffraction orders $\mathbf{B}$:

$$\mathbf{B} = \bar{\mathbf{R}}\,\mathbf{F}. \tag{45.55}$$

In most cases, the incident vector has only one nonzero component, which corresponds to the incident plane wave of the zero-order. It is possible to construct more complex incidence vectors, e.g., one forming a Gaussian beam. However, for subwavelength gratings the reflection of the zero-order plane wave is dominant and the results for a Gaussian beam do not differ significantly from those for a plane wave.

Figure 45.11 shows a typical single-period cell of a subwavelength grating. It is comprised of parallel bars made of a high-refractive-index material (usually Si or GaAs), surrounded on both sides by low-index

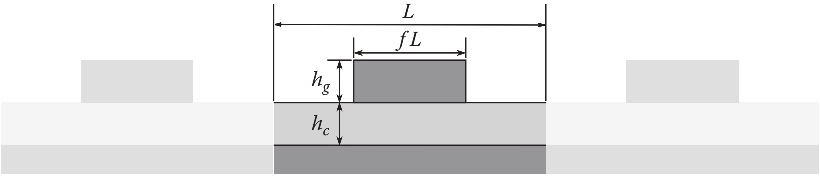


FIGURE 45.11 Schematic diagram of the computational domain for grating reflection computation. The structure is separated into a number of layers uniform in the perpendicular direction and only one grating period is analyzed.

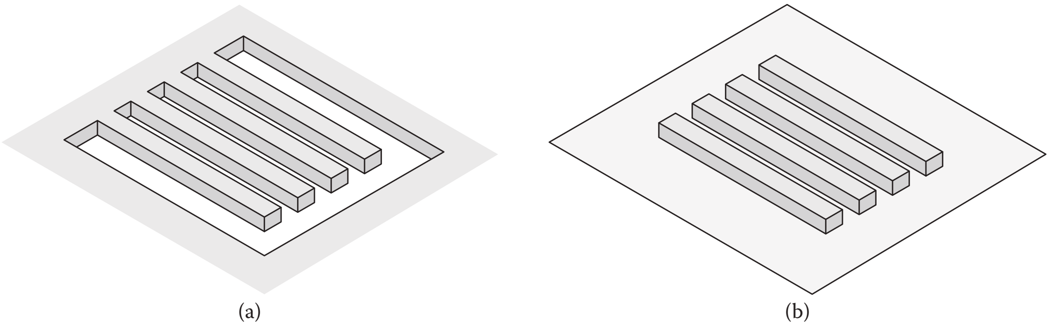


FIGURE 45.12 (a) High-contrast grating membrane suspended in air and (b) similar grating located on low-index cladding (e.g., SiO_2). The former structure is more difficult to fabricate; however, it allows the membrane to be shifted using piezoelectric forces for VCSEL cavity tuning.

layers. These layers can be either the air on both sides (in which case we speak of grating membranes; see Figure 45.12a) or one of the layers may be some kind of low-index solid (usually an oxide, such as amorphous SiO_2). The former are significantly more difficult to fabricate, as the process involves etching a sacrificial layer first and then dissolving it. The resulting grating is suspended in the air. A common problem is then buckling of the stripes, so some additional stress-relaxing etching may be necessary. The advantage of this solution is the fact that the membrane can be placed on a piezoelectric element and then moved freely, allowing for cavity tuning. On the other hand, the latter (Figure 45.12b) is usually much easier to fabricate and the bars are permanently attached to the cladding. The lack of tunability is compensated by the durable mount and no problems with buckling. If the refractive index of the cladding is low enough, there are no significant differences in the optical properties of either type of grating.

Although the gratings are very thin compared to multilayered DBRs, they can provide very high reflectivity, which in an ideal situation can even reach 100%. Due to their strongly anisotropic geometry, the gratings' reflective properties are qualitatively different, depending on the incident light polarization. For an ideal uniform grating, we can distinguish two distinct polarizations: one with an electrical field vector parallel to the grating bars (transverse electric, TE) and the other perpendicular to it (transverse magnetic, TM). Depending on the grating design, any of these polarizations can be strongly reflected, while the other is reflected only moderately. This effect can be used to design VCSELs with strong anisotropy and fixed polarization of the emitted light.

Sample reflectivities computed for a Si membrane suspended in air are shown in Figure 45.13a and b. These figures show reflectivity for both light polarizations with two different gratings. Their exact dimensions are summarized in Table 45.2. Both version were designed to provide high reflectivity at a wavelength of 1500 nm. As can be seen in Figure 45.13, both TE and TM polarization can be strongly reflected, with almost zero transmission. If two such peaks coincide, we observe a wide reflectivity bandwidth. This is clearly visible in Figure 45.13a, where the wide reflectivity area has a small dip in the middle, indicating that it consists of two independent peaks.

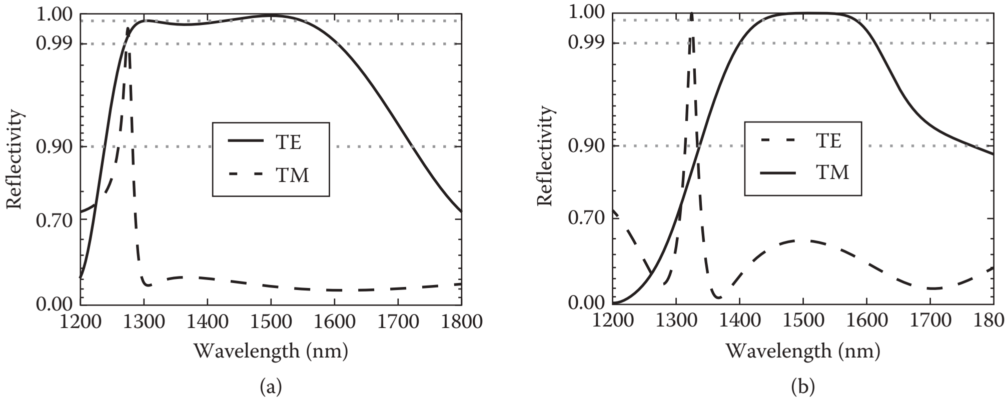


FIGURE 45.13 Silicon-on-silica grating reflectivity for TE and TM polarization. The dimensions of the gratings are optimized to provide high reflectivity around 1500 nm for (a) TE and (b) TM polarization. The exact grating dimensions are summarized in Table 45.2. The dotted lines show reflectivity levels important for laser operation: 90%, 99%, and 99.8%. Notice the non-linear scale in the plots.

TABLE 45.2 Dimensions for Gratings with Reflectivities Shown in Figure 45.13a and b.

	(a)	(b)
Grating period L [μm]	1.00	0.64
Grating fill-factor f	0.30	0.64
Grating thickness h_g [μm]	0.22	0.42
Air-gap/cladding thickness h_c [μm]	0.83	0.83

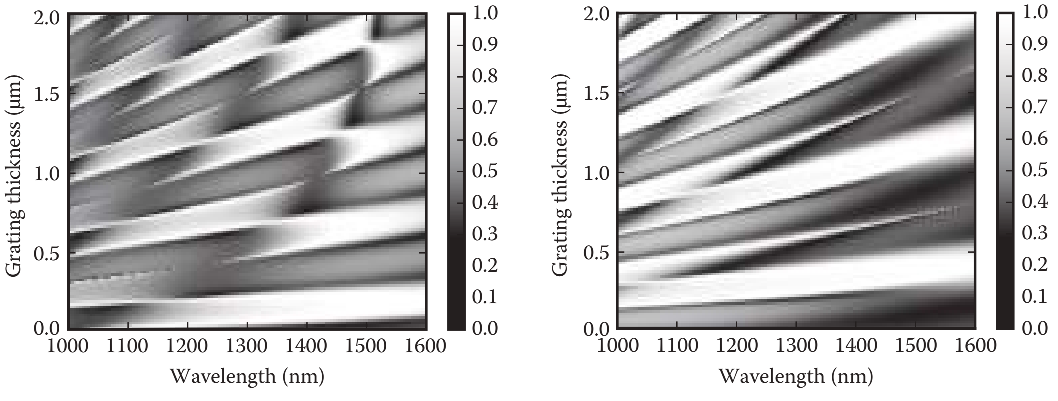
Note: The meaning of the parameters is explained in Figure 45.11. The refractive index of grating bars and substrate is 3.48 (silicon) and of the cladding is 1.47 (silica).

Due to the finite thickness of the grating, it can rather be considered as an array of cavities for the two waveguide array modes. The negative interference described in Section 45.4.1 corresponds to the resonance thickness of these cavities. If we plot grating reflectivity as a function of wavelength and grating thickness (Figure 45.14a and b), we can notice two overlapping periodic patterns, corresponding to the resonance conditions of both waveguide array modes (with propagation constants β_0 and β_2), giving a characteristic chessboard pattern.

Similar calculations can be performed for an HCG located on cladding (Figure 45.12b). With carefully chosen dimensions, it can reflect both TE and TM polarizations equally well. A sample reflectivity spectrum and reflectivity map for TM light polarization is shown in Figure 45.15. The grating parameters are as follows: $L = 0.70\,\mu\text{m}$, $h_g = 0.46\,\mu\text{m}$, $f = 0.75$, and $h_c = 0.83\,\mu\text{m}$. The cladding material is silica with a refractive index of 1.47. Such gratings are usually easier to manufacture and show the same patterns as air membranes.

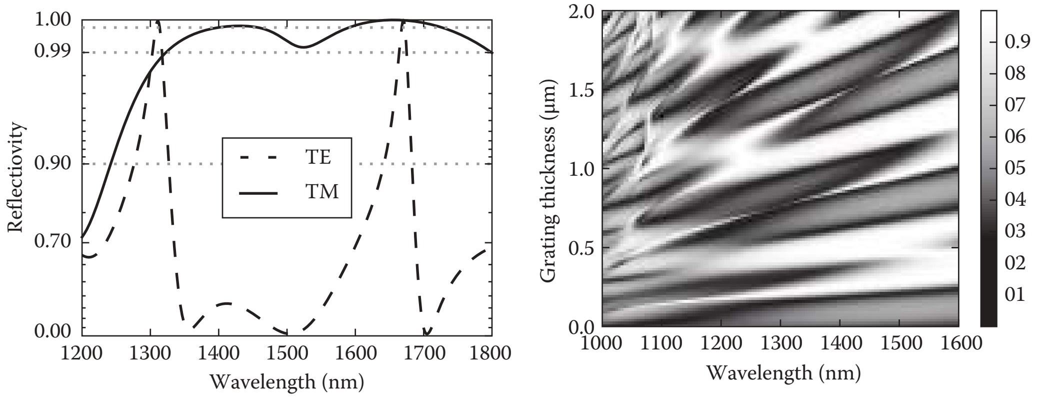
45.4.3 Cavities with Subwavelength Grating Mirrors

The grating mirrors presented in Section 45.4.2 can be used to replace multi-layered Bragg mirrors as reflectors in VCSELs. Typical structures for such lasers—both with an HCG membrane suspended in air and an HCG grating located on a low-index cladding—are presented in Figure 45.16. For manufacturing feasibility, the bottom mirrors are classical DBRs and the top ones are replaced with HCGs. In the case of the



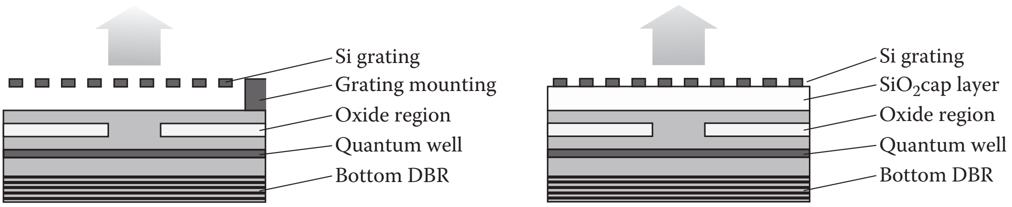
(a) (b)

FIGURE 45.14 Reflectivity maps for gratings optimized for (a) TE and (b) TM polarizations as a function of the wavelength and grating thickness.



(a) (b)

FIGURE 45.15 (a) Reflectivity spectrum and (b) TM reflectivity map for various thicknesses of a grating located on dielectric cladding (Figure 45.12b).


Si grating
Grating mounting
Oxide region
Quantum well
Bottom DBR
Si grating
SiO$_2$cap layer
Oxide region
Quantum well
Bottom DBR

(a) (b)

FIGURE 45.16 Typical structure of an HCG-VCSEL with (a) grating membrane suspended in air and (b) grating layer located on a low-index cladding.

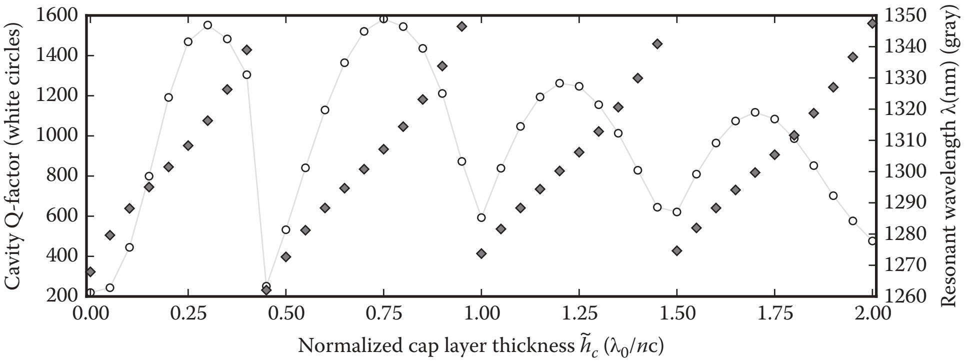


FIGURE 45.17 Cavity Q-factor (light circles) and resonant wavelength (gray diamonds) as functions of the normalized cap layer thickness.

design shown in Figure 45.16a, the membrane can be shifted by piezoelectric forces, effectively changing the cavity length and enabling the emitted wavelength to be tuned. Such a tunable HCG VCSEL has been demonstrated by [33].

In this section, we will analyze an HCG-VCSEL with the grating located on a low-index cladding layer (Figure 45.16b). It consists of a 2-λ GaAs/AlGaAs cavity[†] with a quantum well in one anti-node and an oxide aperture in the second. The bottom mirror is constructed from 29 pairs of GaAs/AlGaAs quarter-wavelength DBRs. At the top, there is an Si HCG separated by a SiO_2 cap layer that provides low index on both sides of the HCG. The cavity and the bottom DBR are designed for a wavelength of 1.3 μm and the HCG parameters are as follows (see Figure 45.11 for the meaning of the symbols): $L = 0.567\,\mu\text{m}$, $h_g = 0.373\,\mu\text{m}$, $f = 0.65$, and h_c varies. We are considering the cold-cavity properties of the laser, thus there is no gain in the quantum well. The oxide aperture is 4 μm and there are a total of seven grating periods within the aperture window. As the computations have to be fully three-dimensional, 891 plane waves have been used to represent the lateral electromagnetic field distribution in each layer: 11 in the direction parallel to the grating and 81 in the direction perpendicular to the grating.

Figure 45.17 shows the computed resonant wavelength of the emitted light and the cavity Q-factor[‡] as functions of cap layer thickness normalized to the target cavity wavelength λ_0 of 1.3 μm. This normalized thickness is computed as $\tilde{h}_c = h_c n_c/\lambda_0$. As can be seen, both the resonant wavelength and the cavity Q-factor depend periodically on $\tilde{h}_c$, with a period of $\lambda/2$ and the highest Q-factor corresponding to the quarter-wavelength cap thickness (i.e., $\tilde{h}_c = 1/4 + k/2$ with k being an arbitrary integer positive number). However, as the cap thickness increases over $\tilde{h}_c = 1.0$, the consequent maxima correspond to a lower and lower Q-factor. This can be attributed to the fact that the finite size of the grating causes an increase in scattering losses, as the divergent beam is not fully back-reflected to the active region.

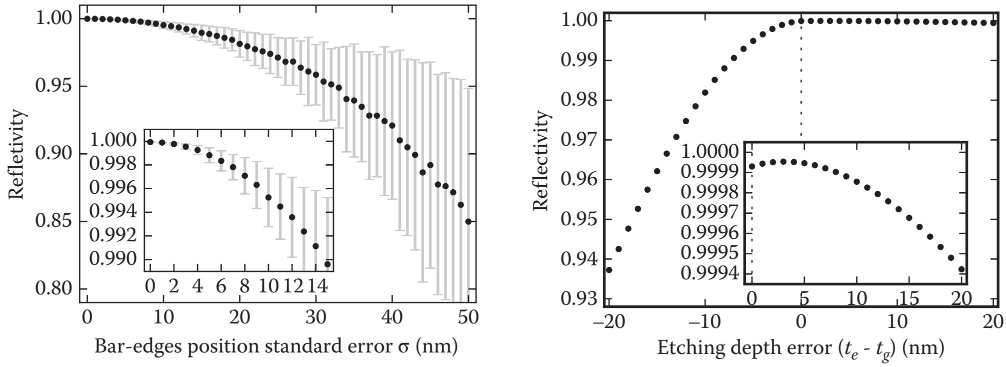
45.4.4 Manufacturing Sensitivity of High-Contrast Grating Mirrors

The results in Figure 45.17 show that HCGs can be effectively used to replace DBR mirrors in VCSELs. However, the question arises whether they are similarly robust. The need to etch a precise lithographic pattern can make them prone to manufacturing errors. We therefore investigated the reflectivity of a grating mirror with an imperfect periodic grating. We performed simulations of multiple grating periods in which grating bar thicknesses and positions were randomly varied with a specified standard deviation σ. Due

---

[†] 2-λ means that the optical cavity length is exactly two wavelengths.

[‡] Q-factor is a dimensionless rate of energy loss relative to the stored energy of the resonator. It can be determined from the equation $Q = \frac{1}{2}\text{Re}(\tilde{\omega})/\text{Im}(\tilde{\omega})$, where $\tilde{\omega}$ is the complex eigen-frequency of the resonant mode in the cavity.



(a)

(b)

FIGURE 45.18 (a) Computed reflectivity for an HCG mirror with the position of the bar edges varied randomly using different standard deviations. The vertical bars show the uncertainties of the results. The inset shows the magnified part of the full figure. (b) Dependence of reflectivity on the difference between etching depth (h_e) and the thickness of the Si layer (h_g, see Figure 45.11). Negative values mean too shallow holes while positive values designate holes that are too deep. The latter are shown magnified in the inset.

to the randomness introduced, we repeated our calculations 1000 times for each value of σ. The results are presented in Figure 45.18a. Their uncertainties (shown with gray vertical lines) are computed using Student's t-distribution for a confidence interval of 90%.

As can be seen, even a small variation causes a significant drop in reflectivity. As the precision of typical photo-lithography is around 10 nm, mirror reflectivity can drop by 0.5% for $\sigma = 10$ nm. Larger manufacturing errors can make the mirrors unusable in VCSELs, as reflectivity can drop well below 90%. The largest allowable error, for which the mirrors reflect more than 99% of the incident light in at least half of the cases, is $\sigma = 15$ nm. The stronger requirement of reflectivity exceeding 99.8% is fulfilled if the error is not greater than $\sigma = 8$ nm. This clearly shows that the main physical phenomenon behind the properties of HCG mirrors is Bragg reflection, in which perfect periodicity plays a crucial role. This contrasts with most modern designs for photonic crystal VCSELs, in which the most important factor is the reduced effective index of the photonic crystal area [8].

Another test we have performed is an analysis of the precision of etching depth. In the ideal case, the Si bars should rest fully separated on a SiO_2 layer. In other words, the spaces between the bars should be etched to a depth exactly equal to the thickness of the Si layer. In reality, this may not always be the case, as the etching can be slightly deeper or slightly shallower. However, locally one can assume uniform etching conditions, whereby the depths of adjacent etched areas should be the same.

In order to investigate the impact of etching depth precision on the performance of HCG mirrors, we computed their reflectivity for various etching depths h_e (h_e is the depth of etching spaces between the bars; in an ideal situation it should be equal to the thickness of the high-index grating layer h_g, see Figure 45.11). Incident light is the same as in the previous case, i.e., the polarization is perpendicular to the bars and the wavelength is 1.3 μm. The results are shown in Figure 45.18b. As one can see, too shallow etching causes a strong decrease in reflectivity. On the other hand, too deep etching does not influence the results significantly. Furthermore, etching 3 nm of the SiO_2 layer can increase reflectivity slightly, although this effect is so weak that it is negligible in practice. The main conclusion is that it is better to etch the space around the bars a little more than desired in order to provide optimum mirror reflectivity even in the case of manufacturing imprecisions.

Similar analysis can be performed for HCG VCSELs, as described in Section 45.4.3. As the reflectivity of HCGs drops with manufacturing imprecisions, so should the VCSEL cavity Q-factor. A numerical analysis of this effect is shown in Figure 45.19. The computation procedure is identical to that used for generating Figure 45.18. However, each point was averaged over 100 simulations instead of 1000 (due to the larger

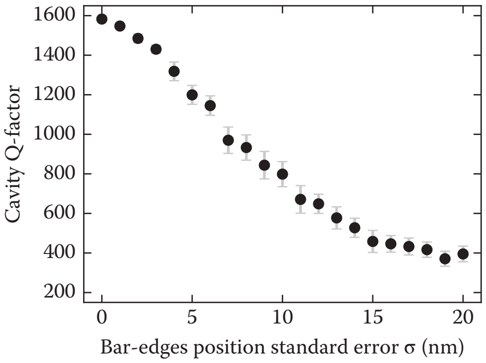


FIGURE 45.19 HCG-VCSEL cavity Q-factor with an imperfect mirror. The edges of the grating bars are varied randomly using different standard deviations. The vertical bars show the uncertainties of the results.

numerical effort necessary to determine the proper VCSEL modes). The resonant wavelength is obviously unaffected, as it is determined by the cavity itself and not the quality of the mirror. We can notice that even a 10 nm error can lead to a twofold reduction in the Q-factor. This must be taken into consideration in designs for devices that are to be produced on an industrial scale. Furthermore, the analysis presented does not consider the effects of the surface roughness of the bars, which to some extent is inevitable in the lithography process.

45.5 Summary

Periodic structures have been used in laser diodes for a long time. Their most basic form are one-dimensional distributed Bragg reflectors, commonly used in VCSELs. However, two- or three-dimensional photonic crystals allow further tuning of laser properties, e.g., to achieve a single-mode light emission. One-dimensional subwavelength high-contrast gratings can make efficient broadband laser mirrors with reduced size and wavelength tuning ability.

The downside of such structures is the increased laser complexity, which requires more advanced numerical methods for the device optical modeling. In particular, only vectorial methods like the finite-difference time-domain method, the plane-wave expansion, the finite element method, the coupled mode model, the plane-wave admittance method, or the plane-wave reflection transfer can yield proper optical characteristics of PC-VCSELs.

References

1. Strutt, J. W. (1887). On the maintenance of vibrations by forces of double frequency, and on the propagation of waves through a medium endowed with a periodic structure. *Phil. Mag.* 27, 145–159.
2. Yablonovitch, E. (1987). Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* 58, 2059–2062.
3. John, S. (1987). Strong localization of photons in certain disordered dielectric superlattices. *Phys. Rev. Lett.* 58, 2486–2489.
4. Kittel, C. (2004). *Introduction to Solid State Physics* (8th ed.). Hoboken, NJ: Wiley.
5. Joannopoulos, J. D., R. D. Meade, and J. N. Winn (1995). *Photonic Crystals: Molding the Flow of Light*. Princeton, NY: Princeton University Press.
6. Checoury, X., P. Boucaud, J.-M. Lourtioz, F. Pommereau, C. Cuisin, E. Derouin, O. Drisse, L. Legouezigou, O. L. Legouezigou, F. Lelarge, F. Poingt, G.-H. Duan, S. Bonnefont, D. Mulin, J.

Valentin, O. Gauthier-Lafaye, F. Lozes-Dupuy, and A. Talneau (2005). Distributed feedback-like laser emission in photonic crystal waveguides on InP substrate. *IEEE J. Sel. Top. Quant. Electron*. 11, 1180–1186.

7. Painter, O., J. Vuckovic, and A. Scherer (1999). Defect modes of a two-dimensional photonic crystal in an optically thin dielectric slab. *J. Opt. Soc. Am. B*. 16, 275–285.
8. Danner, A. J., J. J. Raftery Jr., L. P. O., and K. D. Choquette (2006). Single mode photonic crystal vertical cavity lasers. *Appl. Phys. Lett*. 88, 091114.
9. Danner, A. J., J. J. Raftery Jr., N. Yokouchi, and K. D. Choquette (2004). Transverse modes of photonic crystal vertical-cavity lasers. *Appl. Phys. Lett*. 84, 1031–1033.
10. Leisher, P. O., A. J. Danner, and K. D. Choquette (2006). Single-mode 1.3 μm photonic crystal vertical-cavity surface-emitting laser. *IEEE Photon. Technol. Lett*. 18, 2156–2158.
11. Song, D. S., S. H. Kim, H. G. Park, C. K. Kim, and Y. H. Lee (2002). Single-fundamental mode photonic-crystal vertical-cavity surface-emitting lasers. *Appl. Phys. Lett*. 80, 3901–3903.
12. Yokouchi, N., A. J. Danner, and K. D. Choquette (2003). Two-dimensional photonic crystal confined vertical-cavity surface-emitting lasers. *IEEE J. Sel. Top. Quant. Electron*. 9, 1439–1445.
13. Liu, H., M. Yan, P. Shum, H. Ghafouri-Shiraz, and D. Liu (2004). Design and analysis of anti-resonant reflecting photonic crystal VCSEL lasers. *Opt. Express*. 12, 4269–4274.
14. Czyszanowski, T., M. Dems, and K. Panajotov (2007a). Impact of the hole depth on the modal behaviour of long wavelength photonic crystal VCSELs. *J. Phys. D: Appl. Phys*. 40, 2732–2735.
15. Czyszanowski, T., M. Dems, R. P. Sarzała, W. Nakwaski, and K. Panajotov (2011). Precise lateral mode control in photonic crystal vertical-cavity surface-emitting lasers. *IEEE J. Quant. Electron*. 47, 1291–1296.
16. Birks, T. A., J. C. Knight, and P. S. J. Russell (1997). Endlessly single-mode photonic crystal fiber. *Opt. Lett*. 22, 961–963.
17. Taflove, A. and S. C. Hagness (2000). *Computational Electrodynamics: The Finite-Difference Time-Domain Method* (2nd ed.). Boston, MA: Artec House.
18. Johnson, S. G. and J. D. Joannopoulos (2001). Block iterative frequency-domain methods for Maxwell's equations in a planewave basis. *Opt. Express*. 8, 173–190.
19. Nyakas, P. (2007). Full-vectorial three-dimensional finite element optical simulation of vertical-cavity surface-emitting lasers. *IEEE J. Lightwave Techn*. 25, 2427–2434.
20. Bava, G. P., P. Debernardi, and L. Fratta (2001). Three-dimensional model for vectorial fields in vertical-cavity surface-emitting lasers. *Phys Rev. A*. 63, 23816.
21. Dems, M., R. Kotynski, and K. Panajotov (2005). Plane-wave admittance method—A novel approach for determining the electromagnetic modes in photonic structures. *Opt. Express*. 13, 3196–3207.
22. Dems, M. (2011). Modelling of high-contrast grating mirrors. The impact of imperfections on their performance in VCSELs. *Opto-Electr. Rev*. 19, 340–345.
23. Conradi, O., S. F. Helfert, and R. Pregla (2001). Comprehensive modeling of vertical-cavity laser-diodes by the method of lines. *IEEE J. Quant. Electron*. 37, 928–935.
24. Czyszanowski, T., M. Dems, and K. Panajotov (2007b). Single mode condition and modes discrimination in photonic-crystal 1.3 μm AlInGaAs/InP VCSEL. *Opt. Express*. 15, 5604–5609.
25. Gedney, S. D. (1996). An anisotropic perfectly matched layer-absorbing medium for the truncation of FDTD lattices. *IEEE T. Antenn. Propag*. 44, 1630–1639.
26. Dems, M. (2009). Semi-vectorial method based on effective index for VCSEL analysis. *J. Opt. Soc. Am. B*. 26, 792–796.
27. Dems, M., I.-S. Chung, P. Nyakas, S. Bischoff, and K. Panajotov (2010). Numerical methods for modeling photonic-crystal VCSELs. *Opt. Express*. 18, 16042–16054.
28. Panajotov, K., B. Ryvkin, J. Danckaert, M. Peeters, H. Thienpont, and I. Veretennicoff (1998). Polarization switching in VCSEL's due to thermal lensing. *IEEE Photon. Technol. Lett*. 10, 6–8.
29. Fan, S., W. Suh, and J. Joannopoulos (2003). Temporal coupled-mode theory for the Fano resonance in optical resonators. *J. Opt. Soc. Am. A*. 20 (3), 569–572.

30. Karagodsky, V., F. G. Sedgwick, and C. J. Chang-Hasnain (2010). Theoretical analysis of subwavelength high contrast grating reflectors. *Opt. Express*. 18, 16973–16988.
31. Chang-Hasnain, C. J. and W. Yang (2012). High-contrast gratings for integrated optoelectronics. *Adv. Opt. Photon*. 4, 379–440.
32. Rosenblatt, D., A. Sharon, and A. A. Friesem (1997). Resonant grating waveguide structures. *IEEE J. Quant. Electron*. 33, 2038–2059.
33. Huang, M. C. Y., Y. Zhou, and C. J. Chang-Hasnain (2008). A nanoelectromechanical tunable laser. *Nat. Photon*. 2, 180–184.

46

Single-Photon Sources

46.1 Introduction 585
Characteristics • Single-Photon Generation Schemes • Photon Extraction Strategies
46.2 Modeling of Spontaneous Emission 590
Classical Dipole • Spontaneous Emission • Simulation Techniques
46.3 Cavity QED Approach 594
Micropillar
46.4 Waveguide QED Approach 596
Photonic Nanowire • Photonic Crystal Waveguide
46.5 Planar Dielectric Antenna Approach 600
46.6 Decoherence Effects 601
Quantum Emitter Master Equation • Emitter-Cavity Master Equation • Spectral Wandering • Coupling to Longitudinal Acoustic Phonons
46.7 Summary 604

Niels Gregersen

Dara P. S. McCutcheon

and

Jesper Mørk

46.1 Introduction

The single-photon source (SPS) is a solid-state device capable of emitting single photons on demand. It is a key component within the emerging field of quantum information processing, where the information is encoded on single photons. In these applications, the quantum mechanical properties of the single photon are exploited, e.g., to enable secure communication where undetected eavesdropping is impossible or to perform quantum calculations taking advantage of massive parallelization.

This tutorial chapter reviews the characteristic figure of merits of the deterministic SPS based on a quantum emitter in a semiconductor solid-state system. The theory and the modeling methods used to describe single-photon emission are discussed. Photon extraction strategies based on cavity and waveguide quantum electrodynamics (QED) are presented, and finally we discuss decoherence mechanisms due to the solid-state environment.

46.1.1 Characteristics

For use in quantum information technology, the SPS must meet three basic performance requirements: pure single-photon emission, indistinguishability, and high efficiency, as discussed separately in the three following subsections.

46.1.1.1 Pure Single-Photon Emission

The most fundamental characteristic of the SPS is that it should emit only one photon at a time. This ability is tested in the so-called Hanbury, Brown, and Twiss (HBT) experiment illustrated in Figure 46.1.

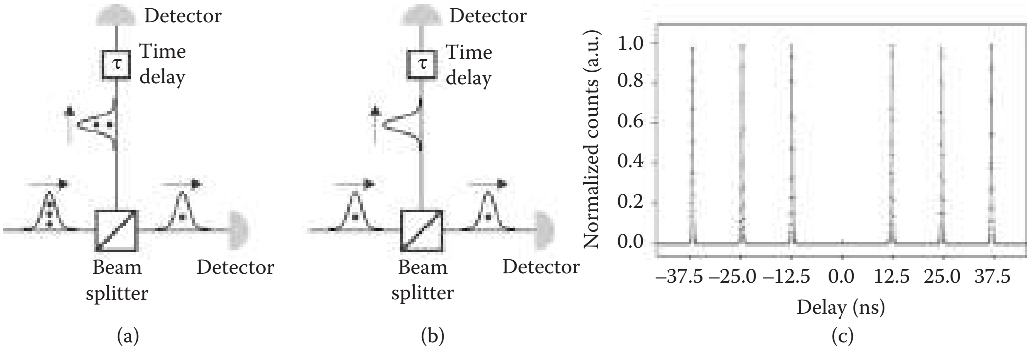


FIGURE 46.1 The Hanbury, Brown, and Twiss experiment illustrated for input pulses containing (a) multiple photons and (b) single photons. (c) The $g^{(2)}(\tau)$ autocorrelation function for a micropillar SPS. (Reprinted with permission from Ding, X. et al., *Physical Review Letters*, 116, 020401, 2016. Copyright 2016 by the American Physical Society.)

Here an optical pulse ideally containing only one photon is directed toward a beam splitter and is either reflected or transmitted toward two detectors. A variable time delay τ is implemented in one of the two output paths, and it is adjusted such that the photons from either of the two ports of the beam splitter arrive simultaneously at the two detectors for zero time delay. The single-photon emission ability is now characterized by measuring the autocorrelation function $g^{(2)}(\tau)$. The experiment is carried out by sending in light pulses from the SPS and by measuring the number of simultaneous detection events in the two detectors as function of time delay.

If the input pulse contains multiple photons as depicted in Figure 46.1a, some photons may be reflected while others are transmitted leading to coincidence clicks in the two detectors at zero time delay such that $g^{(2)}(\tau = 0) \neq 0$. However, if the pulse contains only a single photon as illustrated in Figure 46.1b, this photon will be either transmitted or reflected and no coincidence events are measured. An example of such a HBT measurement is given in Figure 46.1c (Ding et al., 2016), where we observe coincidence counts as function of delay for a SPS generating pulses every 12 ns. The peaks in the measurement correspond to integer values of the pulse time separation, where photons from different pulses may lead to coincidence clicks. However, at zero time delay, the correlation counts are strongly suppressed corresponding to an ideal single-photon emission with $g^{(2)}(\tau = 0) = 0$.

46.1.1.2 Indistinguishability

A second requirement for the SPS is that the emitted photons should be quantum mechanically indistinguishable, meaning that the emitted photons should have the same wavelength, phase variation across the pulse, polarization, optical mode profile, etc. The single-photon indistinguishability is characterized in the Hong–Ou–Mandel (HOM) experiment sketched in Figure 46.2 (Hong et al., 1987).

In this experiment, two single-photon pulses are inbound on a 50–50 beam splitter, where the two pulses are either transmitted or reflected into the output paths. A variable time delay τ is implemented in one of the two output paths adjusted such that the photons reach the detectors simultaneously at zero time delay. The experiment then takes place by sending in single-photon pulses along the input paths and by measuring the correlation function $g^{(2)}(\tau)$ describing the simultaneous detection events for this setup as function of time delay.

The impinging photons may leave the beam splitter either in separate output paths or together as illustrated in Figure 46.2a and b, respectively. In the case where the input photons are distinguishable, the output photons may leave the beam splitter along separate paths leading to $g^{(2)}(\tau = 0) \neq 0$. However, when the input photons are completely indistinguishable and arrive simultaneously at the beam splitter, they always exit the beam splitter along the same output path as depicted in Figure 46.2b, resulting in a $g^{(2)}(\tau = 0) = 0$. This phenomenon is known as the HOM effect and can be understood from a quantum-optical consideration of the probabilities of the four possible output states sketched in Figure 46.2c. The

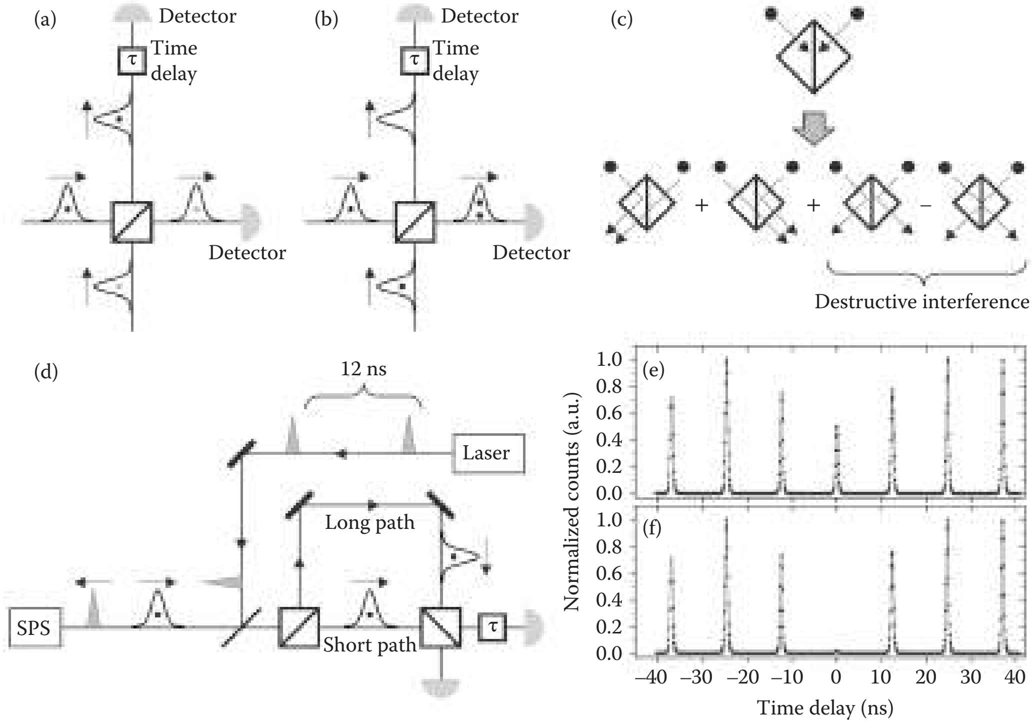


FIGURE 46.2 Outcome of the Hong–Ou–Mandel experiment for (a) distinguishable and (b) indistinguishable photons. (c) Various possible output states. Sketch of a typical experimental setup. Measured coincidence counts in (e) cross and (f) parallel polarization. (Reprinted figure with permission from Ding, X. et al., *Physical Review Letters*, 116, 020401, 2016. Copyright 2016 by the American Physical Society.)

probability amplitude of the photons leaving the beam splitter along separate output paths consists of two contributions corresponding to the photons being both either reflected or transmitted at the beam splitter, and for indistinguishable photons these two contributions add up destructively (Gerry and Knight, 2008).

In most HOM experiments, the impinging photons are generated by the same source using the setup sketched in Figure 46.2d. A laser generates classical pulses at a typical repetition rate of 80 MHz. The pulses are transmitted to the SPS, and each laser pulse triggers a single-photon emission event leading to a train of single photons with 12.5-ns time separation inbound on the first beam splitter. Two subsequent photons in the train can propagate along the short and the long paths in various configurations. The long and short path lengths are adjusted such that a first photon taking the long path and a second photon taking the short path will arrive simultaneously at the beam splitter. Typical HOM measurements for indistinguishable photons having orthogonal or parallel polarization are illustrated in Figure 46.2e and f, respectively. For orthogonal polarizations, the photons are distinguishable and $g^{(2)}(\tau = 0) \neq 0$, whereas for parallel polarizations the HOM effect with $g^{(2)}(\tau = 0) = 0$ is observed. Again, $g^{(2)}(\tau)$ has peaks for time delays equal to an integer times the pulse separation. These peaks correspond to coincidence events for photon pairs that do not arrive simultaneously at the beam splitter and are thus not subject to the HOM effect. The indistinguishability is then quantified as one minus the normalized area of the $\tau = 0$ peak and should ideally equal unity.

46.1.1.3 Efficiency

The third important characteristic of the SPS is its efficiency or brightness (Gregersen et al., 2013). The ideal SPS should be deterministic and provide single photons on demand. It should act as a “single-photon gun” such that every trigger releases exactly one photon into the collection optics. The efficiency ε is then defined as the number of photons detected by the collection optics per trigger. Within the QED design approaches discussed in Sections 46.3 and 46.4, the efficiency is characterized using two governing parameters as illustrated in Figure 46.3. The first is the spontaneous emission factor β describing the fraction of light coupled from the emitter into the optical mode of interest. The β factor can be written as

$$\beta = \frac{\Gamma_{\mathrm{M}}}{\Gamma_{\mathrm{M}} + \Gamma_{\mathrm{Rad}}}, \tag{46.1}$$

where Γ_{M} is the spontaneous emission rate into the optical mode and Γ_{Rad} is the emission rate into background radiation modes. One may also add a term Γ_{NonRad} to the denominator in the definition (46.1) to take into account nonradiative recombination channels of the quantum emitter. Whereas Γ_{M} and Γ_{Rad} can be modified by tailoring the photonic environment, the parameter Γ_{NonRad} represents inherent properties of the emitter that generally cannot easily be controlled.

The second governing parameter is the outcoupling efficiency γ describing the fraction of light transmitted from the optical mode to the first objective lens. In this simple picture, the efficiency is given by $\varepsilon = \beta\gamma$. In a multiphoton interference experiment involving j photons, the total success probability of the experiment will scale as ε^j. For scalable optical quantum information technologies involving a large number of photons, a single-photon generation efficiency close to unity is thus required.

46.1.2 Single-Photon Generation Schemes

A major workhorse for generating single photons has been the spontaneous parametric down-conversion (SPDC) process (Brendel et al., 1999). Let us recall that in a nonlinear isotropic material, the polarization **P** is related to the electric field **E** as

$$\mathbf{P} = \varepsilon_0 \left(\chi^{(1)}\mathbf{E} + \chi^{(2)}\mathbf{E}^2 + \chi^{(3)}\mathbf{E}^3 + \dots \right), \tag{46.2}$$

where $\chi^{(n)}$ is the nth-order susceptibility. In the SPDC scheme, a crystal featuring a $\chi^{(2)}$ nonlinearity such as lithium niobate is illuminated with a strong laser pump pulse with an energy $\hbar\omega_p$ as sketched in Figure 46.4. Provided that proper phase-matching conditions are met, the $\chi^{(2)}$ nonlinearity then enables conversion of a pump photon to two single photons, the signal and the idler photons of energies $\hbar\omega_s$ and $\hbar\omega_i$, respectively, such that $\hbar\omega_p = \hbar\omega_s + \hbar\omega_i$. The advantage of the SPDC source is its simplicity and the high

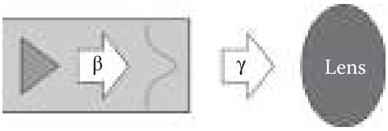


FIGURE 46.3 Efficiency model for the quantum electrodynamics approach. The single-photon emitter is represented by the triangle.

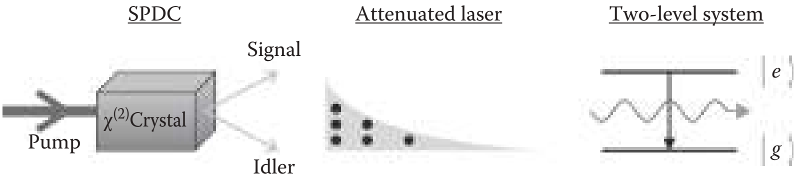


FIGURE 46.4 Major single-photon generation schemes.

degree of indistinguishability of the signal and idler photons. Its main drawback is that the downconversion process is inherently probabilistic, leading to efficiencies below 0.01. While the efficiency may be improved by cranking up the pump power, this occurs at the cost of significant increase of the probability of multiphoton emission. The probabilistic nature thus limits SPDC to experiments involving only a few photons.

Another commercially important source is the attenuated laser SPS (Buttler et al., 2000). A pulse generated using a classical laser propagates through an absorbing material and is attenuated to the single-photon level. The advantage of the attenuated laser is its simplicity, and similar to the SPDC process its main drawback is the probabilistic nature of the attenuation leading to either low single-photon generation efficiency or high probability of multiphoton events. Additionally, the generated photons are not indistinguishable, preventing their use in many quantum optics and quantum information experiments. However, the attenuated laser remains an important device in commercial products for quantum key distribution, where clean single-photon emission and indistinguishability are not required.

The quantum-dot emitter (Shields, 2007) has recently emerged as a solution, overcoming the probabilistic limitations of the SPDC and attenuated laser strategies. It consists of a two-level system as shown in Figure 46.4, where an electron can occupy either a ground state $|g\rangle$ or an excited state $|e\rangle$ characterized by energies E_g and E_e, respectively. The system is initially in the ground state and is then excited using an optical or electrical pulse, after which the electron returns to the ground state by the spontaneous emission of a single photon of energy $\hbar\omega_0 = E_e - E_g$. A main advantage of the two-level emitter is that the upper level can contain only one excitation at a time, thereby ensuring the single-photon nature of the emitted light, combined with the deterministic nature of the spontaneous emission event and the possibility of structuring the medium to realize highly directional emission, potentially allowing for unity efficiency.

Any solid-state system featuring discretized energy levels can in principle be employed as a SPS. The InAs quantum dot (QD) grown in a GaAs host material as shown in Figure 46.5a represents a mature platform for a deterministic SPS. The QD consists of a pyramidal-shaped island of InAs atoms surrounded by GaAs. Due to the lower bandgap of InAs, the electron wavefunction will be fully confined in all three dimensions leading to discretized energy levels in the conduction and valence bands depicted in Figure 46.5b. The emission of InAs QD ensembles features strong inhomogeneous broadening with typical emission wavelength between 900 and 950 nm. The lowest energy *s* state in the conduction band is typically used as the excited state. Due to the spin degree of freedom, the *s* state supports a biexciton leading to two-photon emission. However, the InAs QD features a fine-structure splitting arising from the exciton binding energy of a few nanometers allowing for the selection of the excitonic line using spectral filtering thus enabling single-photon emission at cryogenic temperatures. The indistinguishability of the emitted photons depends greatly on the charge environment and the excitation scheme employed, as discussed in detail in Section 46.6.

Other two-level systems including diamond nitrogen-vacancy centers (Babinec et al., 2010) and organic molecules (Siyushev et al., 2014) have been pursued for single-photon generation. These systems feature the advantages of room temperature operation and very long spin coherence times. However, the maturity

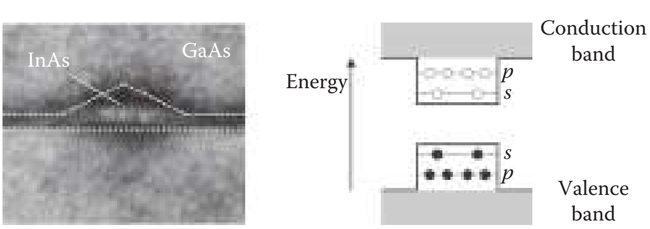


FIGURE 46.5 Scanning electron micrograph (a) and energy level diagram (b) of an InAs quantum dot.

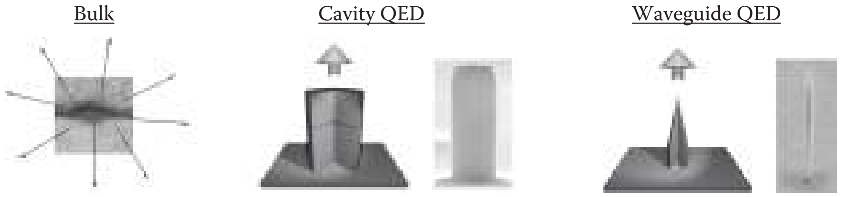


FIGURE 46.6 Major photon extraction strategies. (Reprinted from Gregersen, N et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2013 IEEE and Reprinted by permission from Macmillan Publishers Ltd: Claudon, J et al., *Nature Photonics* 4(3): 174–177, copyright 2010.)

of III–V semiconductor cleanroom fabrication has made the InAs QD the leading platform for highly efficient SPSs, and for this reason we focus in this tutorial on the InAs QD.

46.1.3 Photon Extraction Strategies

While the self-assembled InAs QD features an internal quantum efficiency η of almost unity, a major challenge in single-photon engineering lies in the efficient extraction of the emitted photon (Gregersen et al., 2013). Figure 46.6 summarizes the main SPS photon extraction strategies. The simplest approach consists of a planar sample with a QD in a bulk material. The main advantages of this approach are its simplicity as no additional structuring is needed after QD growth as well as the absence of fluctuating surface charges near the QD as discussed further in Section 46.6. The disadvantage is the poor extraction efficiency. The symmetry of the system leads to light emission in all directions limiting the efficiency to only a few percent.

The first strategy proposed to improve the efficiency and control the light emission was based on cavity QED, where the QD was placed in a microcavity (Gérard, 2003). Here the Purcell effect ensures a preferential single-photon emission from the QD into the cavity mode. This preferential coupling requires a careful spectral alignment between the QD emission line and the cavity line as discussed in detail in Section 46.3. This approach has led to efficiencies of 0.8 (Gazzano et al., 2013) as well as an indistinguishability of 0.99 from the micropillar cavity SPS (Ding et al., 2016; Somaschi et al., 2016), representing the current state of the art in highly efficient sources of indistinguishable photons.

A second design strategy is based on waveguide QED, where the QD is positioned inside a waveguide (Claudon et al., 2013). The preferential coupling to the waveguide mode is ensured by a dielectric screening effect (Bleuse et al., 2011). The main advantage of this nonresonant approach is that no spectral alignment is required. While an efficiency of 0.75 has been demonstrated for the photonic nanowire SPS, the demonstration of highly indistinguishable photons from a waveguide-based SPS is still lacking.

46.2 Modeling of Spontaneous Emission

A major challenge in engineering deterministic SPSs lies in optimizing the photon extraction efficiency and the indistinguishability to achieve values close to unity. Proper modeling of the light emission from a two-level system generally requires a QED description of the spontaneous emission process. However, SPSs generally operate in the weak coupling regime, and here it turns out that the spontaneous emission rate can be calculated in a purely classical manner (Novotny and Hecht, 2012).

46.2.1 Classical Dipole

We first consider the electromagnetic field generated by a classical current distribution $\mathbf{j}$ emitting light at the angular frequency ω_0. We thus work in the frequency domain with harmonic time dependence such that $\mathbf{E}(\mathbf{r}, t) = \mathbf{E}(\mathbf{r})e^{-i\omega_0 t}$ and $\mathbf{j}(\mathbf{r}, t) = \mathbf{j}(\mathbf{r})e^{-i\omega_0 t}$. Furthermore, we assume nonmagnetic materials and the absence of free charges for simplicity. The classical Maxwell's equations are then combined to yield the second-order wave equation for the electric field:

$$\nabla \times \nabla \times \mathbf{E}(\mathbf{r}) - \varepsilon_r(\mathbf{r})k_0^2\mathbf{E}(\mathbf{r}) = i\omega_0\mu_0\mathbf{j}(\mathbf{r}), \tag{46.3}$$

where $k_0 = \omega_0/c$ is the free-space wavenumber and ε_r is the dielectric constant.

Equation 46.3 can be solved in an elegant manner by introducing the dyadic Green's function $\overline{\overline{G}}$ (Novotny and Hecht, 2012). It is defined as the solution to

$$\nabla \times \nabla \times \overline{\overline{G}}(\mathbf{r}, \mathbf{r}') - \varepsilon_r(\mathbf{r})k_0^2\overline{\overline{G}}(\mathbf{r}, \mathbf{r}') = \overline{\overline{I}}\delta\left(\mathbf{r} - \mathbf{r}'\right), \tag{46.4}$$

where $\overline{\overline{I}}$ is the unit dyad. $\overline{\overline{G}}$ and $\overline{\overline{I}}$ are both 3 × 3 tensors and we stress that $\overline{\overline{G}}(\mathbf{r}, \mathbf{r}')$ depends on both $\mathbf{r}$ and $\mathbf{r}'$. Now, using this definition of the dyadic Green's function, we note that

$$\mathbf{E}(\mathbf{r}) = i\omega_0\mu_0 \int_V \overline{\overline{G}}(\mathbf{r}, \mathbf{r}')\mathbf{j}(\mathbf{r}')\,\mathrm{d}\mathbf{r}' \tag{46.5}$$

is a solution to Equation 46.3. The Green's function $\overline{\overline{G}}(\mathbf{r}, \mathbf{r}')$ can thus be interpreted physically as the field generated at the position $\mathbf{r}$ by a current at the position $\mathbf{r}'$. Once the Green's function has been calculated, the entire field profile can be determined from Equation 46.5.

The standard procedure for computing the power P emitted by the current distribution is to integrate the time average of the normal component of the Poynting vector over the surface S of volume V enclosing the current distribution as

$$P = \frac{1}{2}\int_S Re\left(\mathbf{E} \times \mathbf{H}^*\right) \cdot \mathbf{n}_S \mathrm{d}S, \tag{46.6}$$

where $\mathbf{n}_S$ is a unit vector normal to the surface S. Now, the average energy dissipation rate W inside the volume is given by

$$W = -\frac{1}{2}\int_V Re\left(\mathbf{j}^* \cdot \mathbf{E}\right)\mathrm{d}\mathbf{r}, \tag{46.7}$$

and according to Poynting's theorem, we have $P = W$.

We now consider the current distribution generated by a point dipole $\mathbf{d}$ positioned at $\mathbf{r}_0$, as illustrated in Figure 46.7, such that

$$\mathbf{j}(\mathbf{r}) = -i\omega_0\mathbf{d}\delta\left(\mathbf{r} - \mathbf{r}_0\right). \tag{46.8}$$

By inserting Equation 46.8 into Equation 46.5 and subsequently Equation 46.5 into Equation 46.7, we obtain an expression for the power P in terms of the dipole moment given by

$$P = \frac{\omega_0^3\mu_0\,|\mathbf{d}|^2}{2}\mathrm{Im}\left(\mathbf{n}_\mathrm{d} \cdot \overline{\overline{G}}(\mathbf{r}_0, \mathbf{r}_0) \cdot \mathbf{n}_\mathrm{d}\right), \tag{46.9}$$

where $\mathbf{n}_\mathrm{d} = \mathbf{d}/\,|\mathbf{d}|$ is the orientation of the dipole moment. The power P emitted by the point dipole can thus be obtained by evaluating the imaginary part of the Green's function at the position of the dipole.

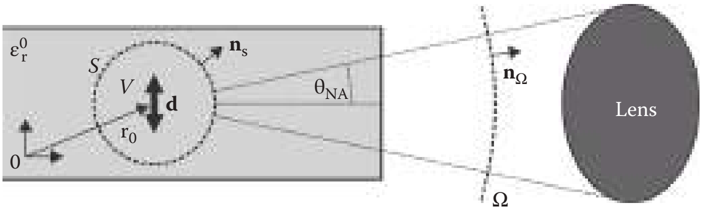


FIGURE 46.7 Light emission from a classical dipole **d** placed at the position $\mathbf{r}_0$ inside a volume V in a material of dielectric constant ε_r^0. The volume is limited by the surface S having normal unit vector $\mathbf{n}_S$. The entire structure is enclosed in a sphere Ω with normal unit vector $\mathbf{n}_\Omega$.

46.2.2 Spontaneous Emission

In the weak coupling regime, the total spontaneous emission rate Γ of a two-level system at the position $\mathbf{r}_0$, with transition energy $\hbar\omega_0$ and with dipole moment **d,** is derived from Fermi's golden rule as

$$\Gamma = \frac{\pi\omega_0}{\hbar\varepsilon_0} |\mathbf{d}|^2 \rho_L \left(\mathbf{n}_\mathrm{d}, \mathbf{r}_0, \omega_0\right). \tag{46.10}$$

Here, $\rho_L \left(\mathbf{n}_\mathrm{d}, \mathbf{r}_0, \omega_0\right)$ is the local density of states (LDOS) defined as

$$\rho_L \left(\mathbf{n}_\mathrm{d}, \mathbf{r}_0, \omega_0\right) = \int \left|\mathbf{n}_\mathrm{d} \cdot \widetilde{\mathbf{E}}_\alpha(\mathbf{r}_0)\right|^2 \delta(\omega_0 - \omega_\alpha) d\alpha, \tag{46.11}$$

where $\widetilde{\mathbf{E}}_\alpha$ are the so-called normal modes solutions to the wave equation in the absence of free currents and charges:

$$\nabla \times \nabla \times \widetilde{\mathbf{E}}_\alpha(\mathbf{r}) - \varepsilon(\mathbf{r}) \left(\frac{\omega_\alpha}{c}\right)^2 \widetilde{\mathbf{E}}_\alpha(\mathbf{r}) = 0, \tag{46.12}$$

and the index α is a short notation for all appropriate discrete and continuous parameters describing a mode. The normal modes are normalized such that

$$\int \varepsilon_r(\mathbf{r}) \widetilde{\mathbf{E}}_\alpha^*(\mathbf{r}) \cdot \widetilde{\mathbf{E}}_{\alpha'}(\mathbf{r}) d\mathbf{r} = \delta(\alpha - \alpha'). \tag{46.13}$$

We now write the dyadic Green's function as an expansion in normal functions. Referring the reader interested in the derivation to Novotny and Hecht (2012), we simply state the result:

$$Im \left(\overline{\overline{G}}(\mathbf{r}_0, \mathbf{r}_0, \omega_0)\right) = \frac{\pi c^2}{2\omega_0} \int \widetilde{\mathbf{E}}_\alpha^*(\mathbf{r}_0) \widetilde{\mathbf{E}}_\alpha(\mathbf{r}_0) \delta(\omega_0 - \omega_\alpha) d\alpha, \tag{46.14}$$

where the integrand contains the outer product of the normal modes. Comparing this expression to Equation 46.11, we observe that the imaginary part of the Green's function is closely related to the LDOS. Indeed, outer multiplications of the dyad with the dipole unit vector and using the definition (46.11) lead to

$$\rho_L \left(\mathbf{n}_\mathrm{d}, \mathbf{r}_0, \omega_0\right) = \frac{2\omega_0}{\pi c^2} \mathrm{Im} \left(\mathbf{n}_\mathrm{d} \cdot \overline{\overline{G}}(\mathbf{r}_0, \mathbf{r}_0, \omega_0) \cdot \mathbf{n}_\mathrm{d}\right), \tag{46.15}$$

and we observe that, except for physical constants, the LDOS is identical to the average power emitted by a classical dipole as given in Equation 46.9. This means that we obtain the very important relationship

$$\frac{\Gamma}{\Gamma_{\text{Bulk}}} = \frac{P}{P_{\text{Bulk}}}, \tag{46.16}$$

which states that the spontaneous emission rate, Γ, of a two-level system embedded in a medium with dielectric constant ε_r^0 normalized to its value, Γ_{Bulk}, in a bulk material of dielectric constant ε_r^0 is equal to the power dissipation, P, from a classical dipole normalized to the rate, P_{Bulk}, in the bulk material.

46.2.3 Simulation Techniques

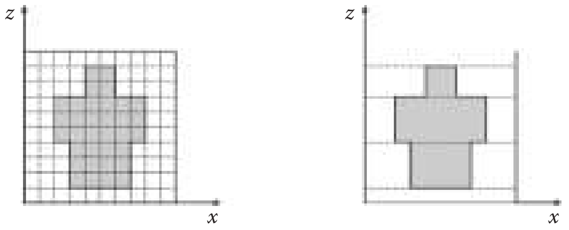
Analytic expressions for the Green's function are only available in a limited number of geometries and generally we rely on numerical simulation techniques to compute the electric field. When using spatial discretization methods such as the finite-difference time-domain technique and the finite-elements method (Lavrinenko et al., 2014), the field is expanded on a grid as sketched in Figure 46.8a and either Maxwell's equations or a vectorial wave equation are directly solved to give the optical field.

Alternatively, one may use a modal method (Lavrinenko et al., 2014). Here, the geometry is sliced into layers uniform along a propagation axis as illustrated in Figure 46.8b and the eigenmodes of each layer are then computed. The optical field is then expanded on the eigenmodes, and the fields at each side of a layer interface are connected using the scattering matrix formalism. This allows for direct access to scattering coefficients and the possibility of studying the various parts of the SPS individually.

As we have seen, the QD can be modeled as a classical point dipole placed inside the geometry under study. The electric field profile generated by the dipole is then computed, and the classical power P can then be obtained from Equation 46.7. We note that, for the current source defined by Equation 46.8, the real part of the electric field diverges at the position of the dipole, and an evaluation of the power P by integrating over a small sphere using Equation 46.6 is thus often preferred for numerical stability. The power collected by the lens of numerical aperture NA can be evaluated by computing the electromagnetic field on the surface of a large sphere Ω using a standard near-field to far-field transformation (Balanis, 1989) and integrating the Poynting vector over the unit solid angle subject to the condition $\theta < \theta_{\text{NA}}$, where $\theta_{\text{NA}} = \text{asin}(\text{NA})$. The efficiency ε is then given as

$$\varepsilon = \frac{\frac{1}{2}\int\limits_{\theta<\theta_{\text{NA}}} \text{Re}\,(\mathbf{E}\times\mathbf{H}^*)\cdot\mathbf{n}_\Omega d\Omega}{P}. \tag{46.17}$$

While Equation 46.17 is exact, this direct procedure does not give complete insight into the governing physics. For this reason, the efficiency in the QED SPS designs is instead treated using an elements-splitting



(a) (b)

FIGURE 46.8 Discretization schemes for (a) the finite-difference time-domain technique and for (b) the modal method.

approach where β, Γ_{M}, Γ_{Rad}, and γ are analyzed and optimized separately and the efficiency is computed using $\varepsilon = \beta\gamma$. The rates Γ_{M} and Γ_{Rad} are available from Equation 46.11 as

$$\Gamma_{\mathrm{M}} = \frac{\pi\omega_0}{\hbar\varepsilon_0} |\mathbf{d}|^2 \left|\mathbf{n}_{\mathrm{d}} \cdot \widetilde{\mathbf{E}}_{\mathrm{M}}(\mathbf{r}_0)\right|^2 \tag{46.18}$$

$$\Gamma_{\mathrm{Rad}} = \frac{\pi\omega_0}{\hbar\varepsilon_0} |\mathbf{d}|^2 \int\limits_{\alpha \neq \mathrm{M}} \left|\mathbf{n}_{\mathrm{d}} \cdot \widetilde{\mathbf{E}}_{\alpha}(\mathbf{r}_0)\right|^2 \delta(\omega_0 - \omega_\alpha) d\alpha, \tag{46.19}$$

and we observe that their evaluation requires access to the electric field profiles of the optical mode of interest as well as the radiation modes.

When performing numerical simulations of SPS designs, the advantages of the spatial discretization methods are their maturity, their availability in commercial user-friendly software packages (Lumerical, n.d.; JCMWave, n.d.), and their ability to exploit parallelization as required in the treatment of large geometries. Their major weakness is that they do not give direct access to the mode profiles $\widetilde{\mathbf{E}}_\alpha(\mathbf{r})$. While the total power P can be obtained from Equation 46.6 or Equation 46.7, the individual rates Γ_{M} and Γ_{Rad} are not directly available and β thus cannot be directly evaluated. On the other hand, in the modal expansion methods, the mode profiles $\widetilde{\mathbf{E}}_\alpha(\mathbf{r})$ are directly available, and this represents a major advantage. However, the drawbacks of the modal method are the lack of user-friendly software packages as well as its difficulty in handling large 3-D geometries due to the lack of eigenvalue solvers capable of exploiting parallelization.

46.3 Cavity QED Approach

Inspection of Equation 46.1 reveals two options for optimizing β. The first is to increase Γ_{M} using cavity QED effects in the weak coupling regime. Here, the emitter is placed inside a cavity with quality factor Q, mode volume V, and resonance wavelength $\lambda_{\mathrm{r}} = 2\pi c/\omega_{\mathrm{r}}$ as illustrated in Figure 46.9a. If the emitter is placed at a maximum of the optical field profile and the emitter and cavity lines are spectrally aligned such that $\omega_{\mathrm{r}} = \omega_0$, the spontaneous emission rate Γ_{M} into the cavity mode is enhanced by the factor F_{P} known as the Purcell factor (Purcell, 1946). It describes the rate Γ_{M} relative to the rate Γ_{Bulk} for an emitter in a bulk medium and is given by (Gérard, 2003)

$$F_{\mathrm{P}} = \frac{\Gamma_{\mathrm{M}}}{\Gamma_{\mathrm{Bulk}}} = \frac{3}{4\pi^2} \frac{Q}{V} \left(\frac{\lambda_r}{n}\right)^3, \tag{46.20}$$

where n is the refractive index of the material enclosing the emitter. The mode volume is computed as[†]

$$V = \frac{\int \varepsilon_r(\mathbf{r}) \left|\widetilde{\mathbf{E}}_{\mathrm{M}}(\mathbf{r})\right|^2 d\mathbf{r}}{\varepsilon_r(\mathbf{r}_{\max}) \left|\widetilde{\mathbf{E}}_{\mathrm{M}}(\mathbf{r}_{\max})\right|^2}, \tag{46.21}$$

where the denominator is evaluated at the position $\mathbf{r}_{\max}$ of a maximum of the electric field intensity profile.

The expression (Equation 46.20) holds provided the emitter line is narrow compared to the cavity linewidth. If this is not the case, an integration over the electronic density-of-states should be performed,

---

[†] When integrated over all space, the integral in Equation 46.21 diverges. To remedy this, it is possible to define a generalized mode volume (Kristensen et al., 2012) that converges in an unambiguous way. However, in high-Q cavities, one can usually employ Equation 46.21 and limit the integration to the device geometry with negligible error.

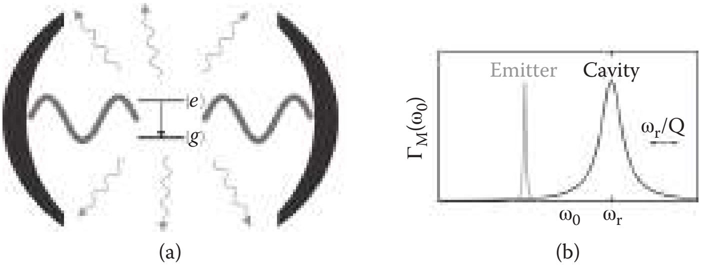


FIGURE 46.9 (a) Cavity-enhanced spontaneous emission. (b) Cavity emission rate Γ_M as a function of frequency.

which generally leads to a reduction in the enhancement. The β factor can be written in terms of the Purcell factor as

$$\beta = \frac{F_P}{F_P + (\Gamma_{Rad}/\Gamma_{Bulk})} \approx \frac{F_P}{F_P + 1}. \tag{46.22}$$

For micropillars with diameters in the few micrometer range, the spontaneous emission into radiation modes is approximately equal to the rate in a bulk material, and we can make the approximation $\Gamma_{Rad} = \Gamma_{Bulk}$ leading to the second equality in Equation 46.22.

We observe from Equations 46.20 through 46.22 that a high-efficiency cavity-based SPS requires a large Q factor combined with a low mode volume, and the optical design challenge for the cavity-based SPS is to maximize the ratio Q/V while maintaining a high outcoupling coefficient γ.

The expression (46.20) assumes perfect spectral alignment between the emitter and the cavity line. If they are detuned, a correction factor given by a Lorentzian lineshape function is introduced, such that (Lorke et al., 2013)

$$\frac{\Gamma_M(\omega_0)}{\Gamma_{Bulk}} = \frac{3}{4\pi^2}\frac{Q}{V}\left(\frac{\lambda_r}{n}\right)^3 \frac{(\delta\omega_r)^2}{(\delta\omega_r)^2 + (\omega_0 - \omega_r)^2}, \tag{46.23}$$

where $\delta\omega_r = \omega_r/Q$ is the cavity linewidth. The cavity emission rate Γ_M is plotted in Figure 46.9b as function of frequency and we observe that for an emitter-cavity misalignment of only a few linewidths $\delta\omega_r$, the enhancement is lost. For the cavity design approach to work, it is thus absolutely necessary to control the spectral alignment.

46.3.1 Micropillar

The AlGaAs microcavity pillar geometry (Reitzenstein and Forchel, 2010) illustrated in Figure 46.10a is one of the most popular platforms for constructing a high-efficiency SPS. It is a vertical structure consisting of a GaAs cavity with an embedded InAs QD surrounded by GaAs/AlAs distributed Bragg reflectors (DBRs). The optical cavity mode profile is confined vertically by the DBRs and laterally by total internal reflection. The emitted light is collected by an objective lens from above.

The thicknesses of the cavity and the DBR layers should be chosen as λ_d/n_{eff} and $\lambda_d/(4n_{eff})$, respectively, where λ_d is the design wavelength and n_{eff} is the effective index of the fundamental mode. These thicknesses ensure that the resonance is centered on the DBR stopband for maximum reflection (Coldren et al., 2012).

In the ideal structure, light escapes only through the bottom and top DBRs. The light emitted into the substrate is lost and the reflectivity of the bottom DBR should thus be significantly higher than that of the top DBR to obtain a high efficiency. The transmission γ to the lens for a micropillar of infinite diameter is presented in Figure 46.10b, and we observe that $\gamma > 0.99$ requires approximately 18 extra layer pairs in the bottom DBR.

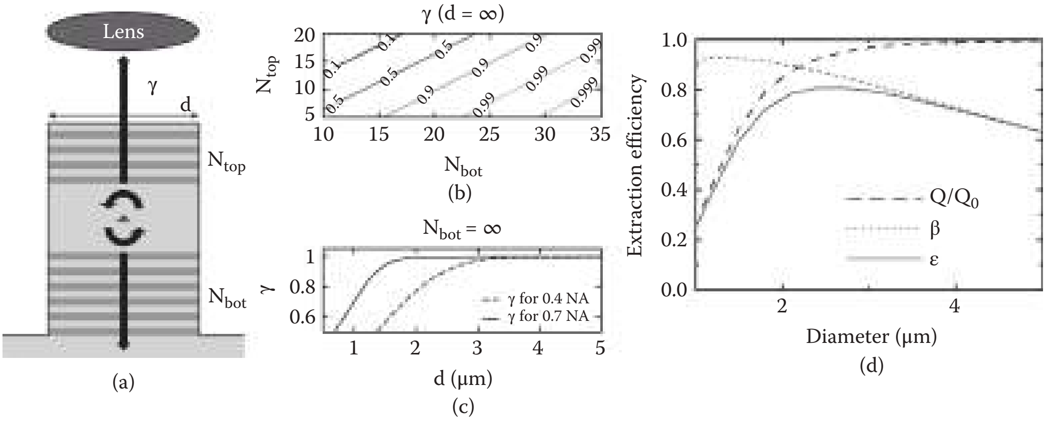


FIGURE 46.10 (a) The micropillar geometry. Transmission γ as function of DBR layer pairs (b) and as function of the pillar diameter (c). (d) Extraction efficiency as function of the pillar diameter. (Reprinted from Gregersen, N. et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2013 IEEE and reprinted by permission from Macmillan Publishers Ltd: Gazzano, O. et al., *Nature Communications* 4: 1425, copyright 2013.)

After the numbers of layer pairs are fixed, the pillar diameter should be chosen. The transmission γ as function of the pillar diameter is shown in Figure 46.10c for a perfectly reflecting bottom DBR. The far-field emission pattern is related to the Fourier transform of the near-field profile. For small diameters, the far-field emission pattern is broad and the transmission is low. For larger diameters, the beam divergence is reduced and the transmission improves. We observe that a pillar diameter above ~2 μm leads to near-unity transmission for a 0.7 NA lens.

In the optimization of the extraction efficiency, two additional mechanisms should be considered. While the ideal structure emits light only along the vertical axis, fabricated pillar structures suffer from sidewall imperfections leading to in-plane scattering of light. The amount of light lost can be estimated by measuring the Q factor and comparing with the quality factor Q_0 of a planar sample without microstructuring. The light lost due to imperfections is given by $1 - Q/Q_0$ and the predicted efficiency should then be corrected by the fraction Q/Q_0 such that $\varepsilon = \beta\gamma Q/Q_0$. The dashed curve in Figure 46.10d shows a fit to experimentally measured Q/Q_0 ratios for the pillar geometry described in Gazzano et al. (2013). As the diameter increases, the relative overlap of the mode profile with the imperfections decreases, and Q thus increases with diameter toward the planar value.

While a large pillar diameter optimizes γ and Q/Q_0 toward unity, it also increases the mode volume V leading to a reduced F_P. When choosing the diameter, it is thus necessary to consider the magnitude of β as function of diameter as given by Equation 46.22. The calculated ratio $\beta = F_P/(F_P + 1)$ is displayed in Figure 46.10d as the dotted curve and it decreases with the diameter due to the increase in V. The product $\varepsilon = \beta Q/Q_0$ is shown in the figure as the full curve, and we observe that a maximum of ~ 0.8 is obtained for d ~ 2.5 μm. This estimation of the efficiency assumes a transmission γ of unity, which is a good approximation when using a 0.7 NA lens.

This micropillar SPS reported in Gazzano et al. (2013) indeed features an efficiency of 0.79. The simultaneous spectral and spatial alignment was achieved using an *in situ* fabrication method pioneered by the group.

46.4 Waveguide QED Approach

The second option for optimizing β as defined by Equation 46.1 is to decrease Γ_{Rad} and this represents the waveguide QED approach. Two main geometries taking advantage of this approach are the photonic nanowire presented in Section 46.4.1 and the photonic crystal waveguide discussed in Section 46.4.2.

Assuming that the QD is positioned at a maximum of the guided mode, the spontaneous emission rate to the guided mode in an infinite uniform waveguide is given by (Claudon et al., 2013)

$$\frac{\Gamma_{\mathrm{M}}}{\Gamma_{\mathrm{Bulk}}} = \frac{3}{4\pi A}\left(\frac{\lambda}{n}\right)^2 \frac{n_{\mathrm{g}}}{n}. \tag{46.24}$$

Here, n_{g} is the group index and

$$A = \frac{\int \varepsilon_r(\mathbf{r}_\perp)\,|\mathbf{E}(\mathbf{r}_\perp)|^2\,\mathrm{d}\mathbf{r}_\perp}{\varepsilon_r(\mathbf{r}_{\max})\,|\mathbf{E}(\mathbf{r}_{\max})|^2}, \tag{46.25}$$

where the integration is over the lateral plane normal to the waveguide axis. In the case of the periodic waveguide of periodicity a, the rate is given by

$$\frac{\Gamma_{\mathrm{M}}}{\Gamma_{\mathrm{Bulk}}} = \frac{3a}{4\pi V}\left(\frac{\lambda}{n}\right)^2 \frac{n_{\mathrm{g}}}{n}, \tag{46.26}$$

where the integration (46.21) of the mode volume V is limited to a single periodic cell.

The waveguide QED approach relies on significant suppression of the spontaneous emission into radiation modes, and this rate Γ_{Rad} should be quantified. The suppression occurs for waveguide diameters on the order of λ/n, and Γ_{Rad} can no longer be approximated to its value in a bulk material. Instead, Γ_{Rad} should be computed using full 3-D numerical simulations. Whereas the calculation of Γ_{M} is usually straightforward, a precise computation of Γ_{Rad} is generally quite demanding because the radiation modes propagate away from the device and perfect absorbing boundary conditions are required to avoid artificial reflections from the boundary walls. It is often determined indirectly by first computing the total spontaneous emission rate Γ_{T}, which for most methods is more accessible, and then subtracting Γ_{M} such that $\Gamma_{\mathrm{Rad}} = \Gamma_{\mathrm{T}} - \Gamma_{\mathrm{M}}$.

46.4.1 Photonic Nanowire

The photonic nanowire (Claudon et al., 2013) is a vertical GaAs cylinder with an embedded QD sitting on a substrate. The geometry is illustrated in Figure 46.11a. Light is guided in the nanowire by total internal reflection at the nanowire-air interface. Single photons emitted by the QD are predominantly coupled to the fundamental HE_{11} waveguide mode as will be discussed in the following. The emission is again collected from an objective lens above the nanowire.

Similar to the micropillar SPS, a bottom mirror is required to avoid loss of light into the substrate. Furthermore, the small nanowire diameter leads to a highly divergent far-field emission pattern, and a top taper strategy is required to reduce this divergence. The reader is referred to (Claudon et al., 2013; Gregersen et al., 2013) for more details on the mirror and the taper strategies. In this tutorial, we will simply study spontaneous emission in the infinite nanowire of Figure 46.11b.

The physical origin of the strong suppression of the spontaneous emission into radiation modes can be understood from Figure 46.11c and d. We first consider a radiation mode defined from the illumination of an infinite photonic nanowire by a plane wave $E_x(\mathbf{r}) = E_0 \exp(ik_0 y)$ propagating along the y axis. The corresponding field intensity profile is shown in Figure 46.11c for a nanowire with diameter $d << \lambda$. Outside the nanowire we observe some field enhancement; however, inside the nanowire the field is strongly suppressed. This field suppression is due to a dielectric screening effect (Bleuse et al., 2011) illustrated in Figure 46.11d, which displays the refractive index profile $n(x)$ and the normalized electric field component $E_x(x)/E_0$ in along the $y = 0$ axis. Now, at the nanowire-air interface, continuity is dictated by Maxwell's equations for the component $D_x(x) = n(x)^2 E_x(x)$ of the displacement field normal to the boundary.

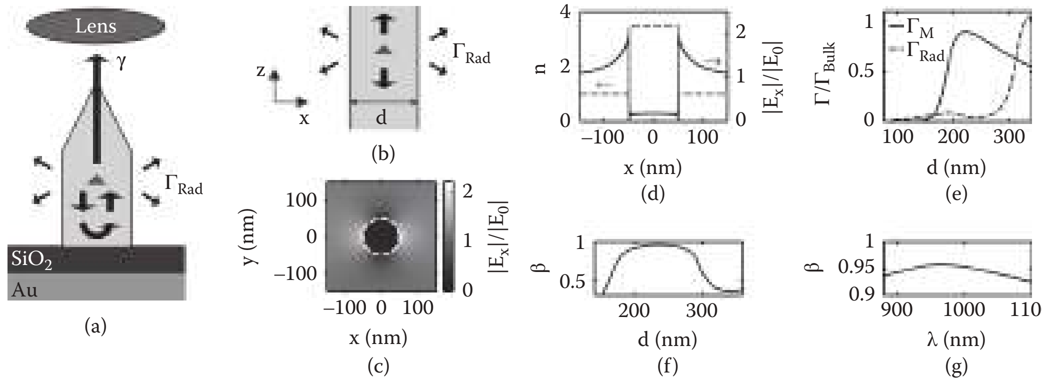


FIGURE 46.11 (a) The full photonic nanowire SPS geometry. (b) The infinite nanowire geometry. (c) The normalized E_x field amplitude of a plane wave propagating along the *y* axis and scattered by a nanowire of diameter $d = 100$ nm. The nanowire boundary is represented by the dashed white line. (d) Normalized E_x field amplitude (solid) and refractive index profile (dashed) at the position $y = 0$. (e) Spontaneous emission rates into the fundamental waveguide mode (solid) and into radiation modes (dashed) as function of nanowire diameter *d*. Spontaneous emission β factor (f) as function of *d* and (g) as function of λ for $d = 240$ nm. Other parameters: $\lambda = 950$ nm and $n_{\text{GaAs}} = 3.46$ unless otherwise noted. (Reprinted from Gregersen, N. et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2013 IEEE.)

Since the refractive index of GaAs is large, around 3.5, the E_x component inside the nanowire must be correspondingly suppressed. Indeed, an analytic expression for the spontaneous emission inside the nanowire is available in the electrostatic limit, where $d << \lambda$, and is given by

$$\frac{\Gamma}{\Gamma_{\text{Bulk}}} = \frac{\Gamma_{\text{Rad}}}{\Gamma_{\text{Bulk}}} = \frac{4}{n\left(n^2+1\right)^2}. \tag{46.27}$$

We observe that for GaAs the suppression is on the order of 150. The expression (46.27) is valid for a dipole oriented in the plane normal to the axis, which is the case for InAs QDs. However, for a dipole oriented along the *z* axis, the relevant boundary condition for the radiation modes at the nanowire-air interface is simply continuity of the E_z component. The mechanism ensuring suppression of the electric field is no longer present, and Γ_{Rad} thus remains substantial even when $d << \lambda$.

The normalized spontaneous emission rates Γ_M and Γ_{Rad} are shown in Figure 46.11e for finite nanowire diameters for an on-axis radially oriented dipole. For $d < 150$ nm, both rates are low. As *d* increases toward 200 nm, the nanowire becomes large enough to support a well-confined fundamental mode and Γ_M increases toward unity. However, for $d < 300$ nm, Γ_{Rad} remains below 0.1 Γ_{Bulk} due to the dielectric screening effect discussed above. The corresponding β factor is presented in Figure 46.11f, and we observe a maximum β of ~0.95 for a diameter of ~240 nm.

Importantly, this screening effect is not a resonant effect. The β factor as a function of wavelength is presented in Figure 46.11g, and we observe that $\beta > 0.93$ is obtained in a 100-nm range. The waveguide QED is thus broadband, and no spectral alignment between an emitter line and a narrow cavity is required. This represents a significant advantage in the fabrication and a major asset of the waveguide QED approach.

46.4.2 Photonic Crystal Waveguide

For integrated quantum photonics, in-plane emission of the emitted light into a planar waveguide is desired. A platform allowing for the efficient in-plane emission of single photons is the photonic crystal (PhC) membrane with an embedded QD (Yao, Manga Rao, and Hughes, 2010) as illustrated in Figure 46.12a. By

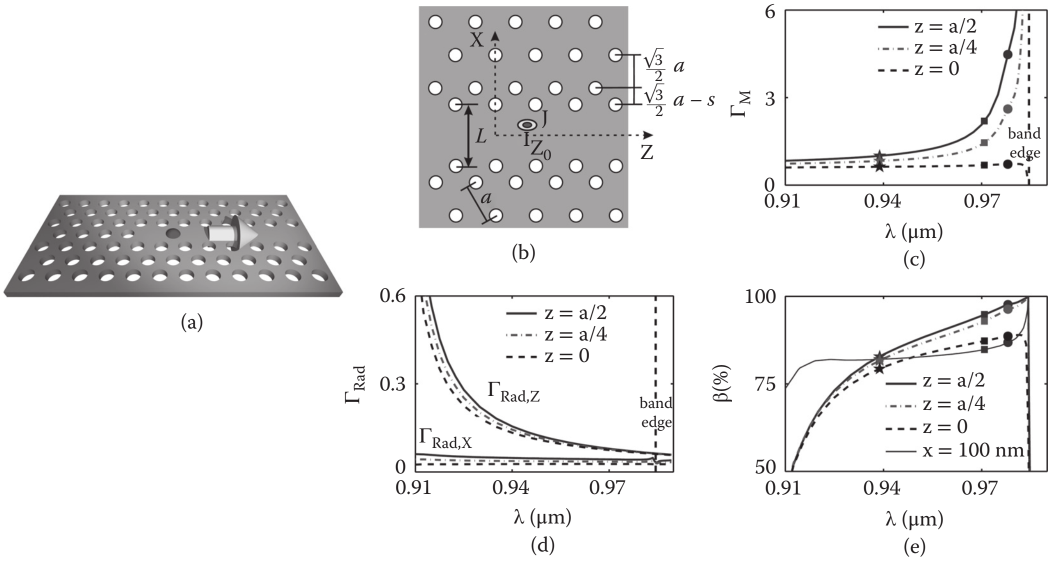


FIGURE 46.12 (a) The PhC waveguide SPS. (b) Geometry of the infinite periodic PhC waveguide. (c) Normalized SE decay rates Γ_M as function of λ. Normalized SE decay rates (d) into the radiation modes for the two in-plane dipole moment orientations X and Z. (e) Associated β factors. The thin solid curve is obtained for an off-axis dipole location ($x = 100$ nm and $z = a/2$). Circles, squares, and stars are reference marks locating the fundamental Bloch-mode group velocities equal to $c/20$, $c/10$, and $c/5$, respectively, where c is the speed of light. (Reprinted with permission from Lecamp, G. et al., *Physical Review Letters*, 99, 023902, 2007. Copyright 2007 by the American Physical Society.)

removing a row of holes, a waveguide can be formed. The waveguide confines light in the plane using the PhC bandgap and vertically using total internal reflection.

In the following, we study the spontaneous emission of the infinite periodic PhC waveguide illustrated in Figure 46.12b. In this tutorial, let us simply consider a dipole positioned in the center of the waveguide $x = 0$ at the position $z = a/2$ with hole displacement $s = 0$. The corresponding results are displayed in Figure 46.12c through 46.12e as full curves. The spontaneous emission rate Γ_M into the fundamental PhC waveguide mode for a dipole oriented along the x axis is shown in Figure 46.12c. For short λ, the rate is similar to that in a bulk material; however, near the band edge Γ_M diverges. This strong enhancement is due to the slow-light effect occurring near the band edge, where the group refractive index diverges (Baba, 2008).

The emission into radiation modes is illustrated in Figure 46.12d for dipoles aligned along the x and the z axes. $\Gamma_{Rad,X}$ is strongly suppressed in the entire wavelength regime due to the same dielectric screening effect discussed for the photonic nanowire. As for the photonic nanowire, this suppression does not take place for the dipole oriented along the z axis and $\Gamma_{Rad,Z}$ is small only near the band edge.

The average spontaneous emission β factor for dipoles oriented along the x and z axes can be written as

$$\beta = \frac{\Gamma_M}{\Gamma_M + \Gamma_{Rad,X} + \Gamma_{Rad,Z}} \tag{46.28}$$

and is depicted in Figure 46.12e. We observe that β approaches unity near the band edge thanks to the suppression of Γ_{Rad} combined with the enhanced Γ_M. The β factor remains larger than 0.75 over a 50 nm wavelength range, an effect predominantly due to the broadband suppression of the spontaneous emission into radiation modes. By displacing the QD x position, the range can be increased as illustrated by the thin solid curve for $x = 100$ nm in Figure 46.12e.

46.5 Planar Dielectric Antenna Approach

A final SPS design discussed in this tutorial is the planar dielectric antenna (PDA) illustrated in Figure 46.13a. This design approach (Chen, Götzinger, and Sandoghdar, 2011) differs conceptually from the previous QED strategies. Whereas the quantum emitter in the QED approaches couples to a specific optical cavity or waveguide mode with the coupling coefficient β, the quantum emitter in the PDA design instead couples to a continuous range of radiation modes and the collection efficiency ε is computed by integrating the contributions from this entire range.

We consider the geometry illustrated in the inset of Figure 46.13b, where a quantum emitter with horizontal electric dipole (HED) or vertical electric dipole (VED) moment orientation is placed in a material of refractive index n_2. In this geometry, light will be collected by an objective lens placed below the emitter.

Without any precautions, light is emitted in all directions due to symmetry, and it is necessary to implement a top mirror to reflect light propagating upward back toward the bottom. While a gold mirror can provide a high broadband reflectivity, coupling to surface plasmons reduces the reflectivity for certain angles of incidence (Chen, Götzinger, and Sandoghdar, 2011). However, the high reflectivity can be restored by introducing an intermediate low-index layer of index $n_3 < n_2$.

With the introduction of the metal mirror, the light is now propagating downward with a highly divergent far-field pattern. To reduce this divergence, a solid immersion lens of index $n_1 > n_2$ is introduced and the downward propagating field is now limited to a cone of polar angle $\theta = \sin^{-1}(n_2/n_1)$. The power density and the collection efficiency as function of the polar angle θ is shown in Figure 46.13b for the two dipole orientations. We observe that a collection efficiency above 0.99 is obtained for an angle $\theta = 55°$. The wavelength dependence on the efficiency is presented in Figure 46.13c, and we observe that $\varepsilon > 0.99$ is maintained over a wavelength range of several hundred nanometers for both emitter orientations. The inset of Figure 46.13c shows the influence of the spacer thickness s on the collection efficiency. As s decreases, the increased coupling of the vertical components of the electric field generated by a VED with surface plasmons reduces the efficiency. However, the field generated by the HED has much weaker vertical components; the coupling to surface plasmons is reduced and the efficiency for this dipole orientation thus remains high.

Whereas the PDA design features record-high broadband efficiency, the condition $n_1 > n_2$ imposes constraints on the choice of material for the medium 1 if the emitter is, e.g., a QD in GaAs with index $n_2 = 3.5$. This can be remedied by introducing an intermediate low-index buffer layer between media 1 and 2. However, an important currently unresolved issue for the PDA design remains, namely the highly non-Gaussian far-field emission pattern illustrated in Figure 46.13b, which makes efficient coupling of the emitted photons to the Gaussian mode profile of a single-mode fiber a challenge.

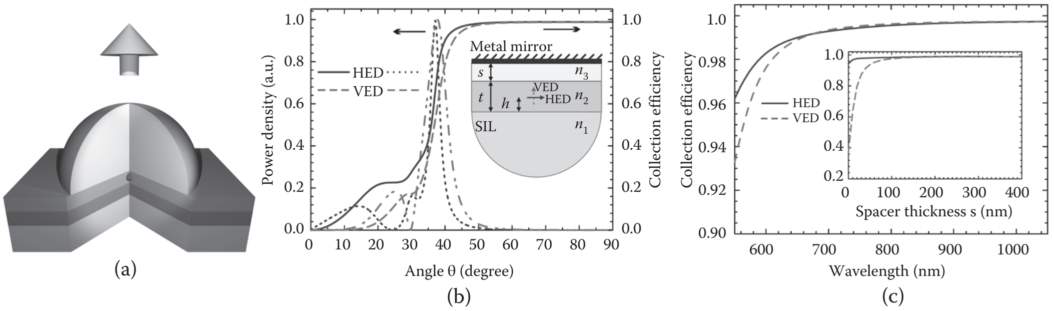


FIGURE 46.13 (a) Planar dielectric antenna. (b) Power density (left axis) and collection efficiency (right axis) versus emission/collection angle θ at a wavelength of 637 nm. The inset shows the schematics of the device with vertically and horizontally oriented dipole emitters. (c) Collection efficiency as a function of wavelength. The inset plots the collection efficiency versus spacer thickness s at 637 nm. (Reprinted from Chen, X,-W. et al., *Optics Letters* 36, 3545–3547, 2011. With permission of Optical Society of America.)

46.6 Decoherence Effects

As discussed above, as well as producing single photons and having a high extraction efficiency, an ideal SPS would produce photons that are quantum mechanically indistinguishable. With reference to Figure 46.2, when two indistinguishable photons are incident on a 50–50 beam splitter they emerge in the same output arm, which provides the fundamental quantum interference necessary for quantum information applications. For photons that are not completely indistinguishable, there remains a nonzero probability that a single photon emerges in both output arms. Based on these arguments, the indistinguishability of a SPS can be quantified by Kaer et al. (2013):

$$I = \frac{\int\limits_0^\infty \int\limits_0^\infty \left| \left\langle \hat{A}^\dagger(t+\tau)\hat{A}(t) \right\rangle \right|^2 dt d\tau}{\int\limits_0^\infty \int\limits_0^\infty \left\langle \hat{A}^\dagger(t)\hat{A}(t) \right\rangle \left\langle \hat{A}^\dagger(t+\tau)\hat{A}(t+\tau) \right\rangle dt d\tau}, \tag{46.29}$$

where $I = 1$ for perfectly indistinguishable photons and $I = 0$ for distinguishable photons. The operators $\hat{A}^\dagger$ and $\hat{A}$ are the creation and annihilation operators for the electric field incident on the beam splitter. Further details of the derivation leading to Equation 46.29 can be found in Unsleber et al. (2015). The challenge in modeling the indistinguishability of photons emitted from a semiconductor-based SPS lies in (a) relating the correlation function and expectation values in Equation 46.29 to operators pertaining to the quantum emitter, and (b) ensuring that the model of the emitter itself captures all important processes that affect the photon indistinguishability.

46.6.1 Quantum Emitter Master Equation

To gain some insight, we first consider a QD emitting in a bulk material or a very low-Q cavity. In this case, provided the emitter is sufficiently far from the beam splitter optics, the field operators in Equation 46.29 can be replaced with operators describing the emitter. We write

$$\hat{A}(t) \sim \hat{\sigma}(t), \tag{46.30}$$

where $\hat{\sigma} = |g\rangle \langle e|$ annihilates an excitation of the emitter (Kiraz et al., 2004). The proportionality constants that we have omitted are independent of time, and with reference to Equation 46.29 we see that they will cancel when the indistinguishability is calculated.

Our task now is to calculate the correlation function in the numerator of Equation 46.29, which can be achieved with help from the quantum regression theorem. To begin, we must first write down a master equation describing the dynamics of the quantum emitter, which takes the form of an equation of motion for the density operator $\hat{\rho}$ (Carmichael, 1999). This can be done to various levels of rigor, but in order to capture the most important qualitative features it suffices to write

$$\frac{d}{dt}\hat{\rho} = \Gamma_{\text{Bulk}} L_{\hat{\sigma}}(\hat{\rho}) + 2\gamma^* L_{\hat{\sigma}^\dagger\hat{\sigma}}(\hat{\rho}), \tag{46.31}$$

where the Lindblad operators are defined to satisfy $L_{\hat{O}}(\hat{\rho}) = \hat{O}\hat{\rho}\hat{O}^\dagger - \left(\hat{O}^\dagger\hat{O}\hat{\rho} + \hat{\rho}\hat{O}^\dagger\hat{O}\right)/2$ and Γ_{Bulk} is the bulk spontaneous emission rate. The first term on the right side of Equation 46.31 describes the decay of the excited state while the second term describes the decoherence of the two-level system. The quantity γ^* is the dephasing or decoherence rate, which we will discuss further in the following sections. Equation 46.31

allows us to establish equations of the motion for expectation values of the emitter. The dipole expectation value, e.g., obeys

$$\frac{d}{dt}\left\langle\hat{\sigma}\right\rangle = -\frac{1}{2}\left(\Gamma_{\text{Bulk}} + 2\gamma^*\right)\left\langle\hat{\sigma}\right\rangle. \tag{46.32}$$

Armed with Equation 46.31, we can now invoke the quantum regression theorem, which states that the correlation function in Equation 46.29 obeys the same equation of motion as the dipole operator in Equation 46.32 but with respect to τ, namely $\frac{d}{d\tau}\left\langle\hat{\sigma}^\dagger(t+\tau)\hat{\sigma}(t)\right\rangle = -\frac{1}{2}\left(\Gamma_{\text{Bulk}} + 2\gamma^*\right)\left\langle\hat{\sigma}^\dagger(t+\tau)\hat{\sigma}(t)\right\rangle$. The initial condition is found by setting $\tau = 0$, from which we find $\left\langle\hat{\sigma}^\dagger(t)\hat{\sigma}(t)\right\rangle = e^{-\Gamma_{\text{Bulk}}t}$ if we assume the emitter is initially in its excited state at time $t = 0$. This allows us to find the correlation function, and inserting this solution into Equation 46.29 we find that the indistinguishability is given by Kaer et al. (2013):

$$I = \frac{\Gamma_{\text{Bulk}}}{\Gamma_{\text{Bulk}} + 2\gamma^*}. \tag{46.33}$$

This expression demonstrates that for our simplistic model, the indistinguishability depends solely on the ratio, $\Gamma_{\text{Bulk}}/\gamma^*$ that we should seek to maximize.

46.6.2 Emitter-Cavity Master Equation

So far we have considered a QD emitting into a bulk medium. However, using similar techniques to those above it is also possible to calculate the indistinguishability of photons emitted from a coupled emitter-cavity system. The calculation is performed by replacing the master equation in Equation 46.31 with the more complicated expression (Kaer et al., 2013)

$$\frac{d}{dt}\hat{\rho} = -i\left[\hat{H},\hat{\rho}\right] + \Gamma_{\text{Bulk}}L_{\hat{\sigma}}(\hat{\rho}) + 2\gamma^* L_{\hat{\sigma}^\dagger\hat{\sigma}}(\hat{\rho}) + \kappa L_{\hat{a}}(\hat{\rho}). \tag{46.34}$$

Here, $\hat{H} = g\left(\hat{\sigma}\hat{a}^\dagger + \hat{a}\hat{\sigma}^\dagger\right)$ describes the interaction between the emitter and the cavity with g the emitter-cavity coupling strength and $\hat{a}$ the annihilation operator of the cavity mode. The last term in Equation 46.34 describes the outcoupling of light from the cavity with κ the cavity mode decay rate. Broadly speaking, provided the losses in the system are weaker than the emitter-cavity coupling strength, the cavity mode can be adiabatically eliminated from this equation of motion. For details of this procedure, we refer the reader to Kaer et al. (2013). This results in a master equation describing the emitter only, which is equal to Equation 46.31 but with Γ_{Bulk} replaced with $\Gamma = \Gamma_{\text{Bulk}} + \Gamma_{\text{M}}$ with $\Gamma_{\text{M}} = \Gamma_{\text{Bulk}}F_{\text{P}}$ the emission rate into the cavity, and the Purcell factor now given by

$$F_{\text{P}} = \frac{4g^2}{\kappa\Gamma_{\text{Bulk}}}. \tag{46.35}$$

We note that the Purcell factor given here is in fact equivalent to that given in Equation 46.20. The apparent difference is a matter of notation, with Equation 46.35 being preferred in the present context since it makes explicit the dependence on the emitter-cavity coupling strength g. With this modification, the indistinguishability measure for a weakly coupled emitter-cavity system becomes

$$I = \frac{\left(F_{\text{P}} + 1\right)\Gamma_{\text{Bulk}}}{\left(F_{\text{P}} + 1\right)\Gamma_{\text{Bulk}} + 2\gamma^*}. \tag{46.36}$$

This expression encompasses the two main strategies used to increase photon indistinguishability. First, as one might expect, eliminating sources of decoherence, and thus reducing γ^*, serves to increase I. Methods to achieve this are briefly discussed in the following section. Second, placing the emitter in a photonic structure such as a micropillar can give rise to a Purcell enhancement of the spontaneous emission rate, which will also increase the indistinguishability. In either case, the strategy amounts to extracting the single photon on a timescale faster than the decoherence rate. We note, however, that these simple arguments only apply as long as the dephasing rate γ^* and the Purcell factor are independent parameters. As we will see below, dephasing caused by coupling to phonons can increase as the Purcell factor is increased, meaning that increasing photon indistinguishability by increasing the spontaneous emission rate may not always be advisable.

46.6.3 Spectral Wandering

Having introduced the basic theory behind photon indistinguishability, we shall now discuss two main mechanisms of dephasing in semiconductor emitters. We begin with spectral wandering, which refers to the process by which the spectral emission line of a QD moves (wanders) over time (Berthelot et al., 2006). In such a case, successively emitted photons may have different frequencies and therefore cannot be considered indistinguishable. Many processes in semiconductor systems can give rise to spectral wandering, with the amplitude and frequency distribution being dependent on the nature and timescale of the process. The most common process in charge-neutral QDs is spectral diffusion caused by the temporary trapping of charges in nearby sites, which induces a change of the electrostatic environment of the QD.

In order to model spectral diffusion, it is typically sufficient to include a phenomenological dephasing term in the master equation describing the QD degrees of freedom as in Equation 46.31 above. As we have seen, dephasing of this form can be combated by increasing the Purcell effect. Recalling also that the efficiency of a source increases with the Purcell effect, we conclude that in the case of simple phenomenological dephasing, such as that caused by spectral wandering, an increased Purcell factor is always advantageous.

46.6.4 Coupling to Longitudinal Acoustic Phonons

Coupling to phonons, present in any semiconductor sample, represents an additional source of decoherence. It has been demonstrated that the dominant phonon-mediated dephasing mechanism is due to coupling of the QD exciton to longitudinal acoustic phonons via the deformation potential (Ramsay et al., 2010). Phonons deform the QD structure and give rise to fluctuations in the exciton energy, which in turn cause a dephasing of superpositions of the ground and excited state. As such phonons are sensitive to the QD dressed state energy difference, which itself depends on the strength of any QD-cavity coupling g. Therefore, in contrast to spectral wandering, phonons cannot be accurately included by the addition of a simple phenomenological dephasing rate that is independent of other parameters.

For these reasons, correctly incorporating phonon coupling into the description of a QD-based SPS is particularly challenging. One approach is to include the exciton-phonon coupling at a Hamiltonian level, where one essentially takes a large but finite number of harmonic oscillators to model the phonon environment. The dynamics describing the complete system can then be solved numerically, giving results that are exact to within a set numerical tolerance (Kaer et al., 2012, 2013). This method, however, requires a large amount of computing power and can obscure the underlying physical processes. Another less numerically demanding method involves deriving a time-local master equation describing the emitter-cavity system using projection operators. This results in more tractable expression from which the indistinguishability can be calculated, though is valid only to second order in the exciton–phonon coupling strength. A comparison of these methods and a simple phenomenological approach can be found in Kaer et al. (2013).

Here we will adopt a simple model that manages to capture the salient qualitative features while being simple to implement. We replace the constant dephasing rate γ^* in Equation 46.34 with the expression (Ramsay et al., 2010)

$$\gamma_{\text{ph}} = \frac{\pi}{4} J(2g) \coth\left(\frac{g}{k_{\text{B}}T}\right), \tag{46.37}$$

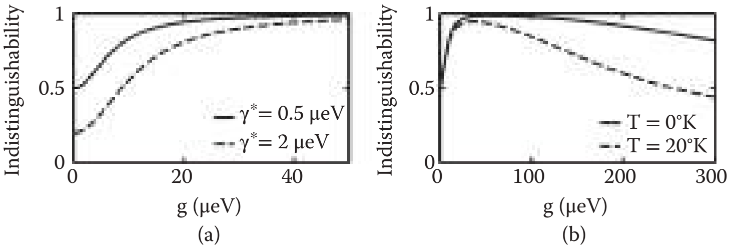


FIGURE 46.14 (a) Photon indistinguishability as a function of emitter-cavity coupling strength for a simple constant dephasing rate of $\gamma^* = 0.5$ μeV (solid) and $\gamma^* = 2$ μeV (dashed). (b) A similar plot where the dephasing rate is composed of a constant value of $\gamma^* = 0.5$ μeV plus a phonon-induced contribution given by Equation 46.37 for temperatures of $T = 0$ K (solid) and $T = 20$ K (dashed). Other parameters are $\Gamma_{\text{Bulk}} = 1$ μeV, $\kappa = 100$ μeV, $\eta = 0.032$ ps^2 and $\omega_c = 0.95$ μeV.

where T is the sample temperature and the function $J(\omega)$ is the exciton-phonon spectral density. For typical GaAs QDs, the spectral density takes the form

$$J(\omega) = \eta\omega^3 e^{-\left(\frac{\omega}{\omega_c}\right)^2}, \tag{46.38}$$

where η is a quantity describing the overall strength of the interaction and can be determined from material parameters, while ω_c is a characteristic cutoff frequency that is inversely proportional to the linear dimension of the QD size (Ramsay et al., 2010). Using Equation 46.37 in Equation 46.34 and again invoking the quantum regression theorem, Equation 46.29 can be evaluated to explore the indistinguishability of photons emitted from a coupled QD-cavity system subject to dephasing caused by phonons.

One final consideration is the correct operator to insert into Equation 46.29. Since in the majority of this chapter we are interested in the emission via the cavity mode, if this mode cannot be adiabatically eliminated it is more appropriate to use the cavity operator $\hat{a}$ rather than the emitter operator in Equation 46.29. In doing so, we are explicitly considering the indistinguishability of photons emitted into the cavity mode of interest.

In Figure 46.14a, we show the photon indistinguishability as a function of the emitter-cavity coupling strength g for a constant dephasing rate of $\gamma^* = 0.5$ μeV (solid) and $\gamma^* = 2$ μeV (dashed). For $g = 0$, the indistinguishability is given by its value in bulk (Equation 46.33). As g is increased, the coupling of the emitter to the cavity gives rise to an enhancement of the Purcell effect that accelerates the spontaneous emission, thus increasing the indistinguishability. Figure 46.14b shows a similar plot where we now include a constant dephasing rate $\gamma^* = 0.5$ μeV and also a phonon contribution given by Equation 46.37. For small values of g, we again observe an increase in I due to the Purcell enhancement. However, as g is increased further, the phonon contribution to the dephasing increases, which serves to reduce the indistinguishability. These plots demonstrate the subtle interplay between the beneficial effects of the Purcell enhancement and the detrimental effects of increased coupling to phonons. The reader is referred to (Kaer et al., 2012, 2013; Kaer and Mørk, 2014) for a more extensive discussion of these effects.

46.7 Summary

In this chapter, we have presented the key figures of merit for the single-photon source. For applications in quantum information processing, clean single-photon emission, efficient photon extraction, and indistinguishable photon emission are required. Sources based on spontaneous parametric down-conversion or attenuation are simple to implement but suffer inherently from low efficiency. We have presented the two-level system with focus on the semiconductor quantum dot as an attractive platform for realizing a deterministic highly efficient single-photon source. Design strategies for controlling the light emission and ensuring a near-unity coupling efficiency based on cavity and waveguide quantum electrodynamics approaches as well as the planar dielectric antenna have been presented. We have discussed the theory and major numerical simulation techniques for modeling the light emission. Finally, we have discussed the influence of several decoherence processes arising from the solid-state environment including simple pure dephasing mechanisms such as charge fluctuations and the more advanced phonon-induced decoherence process due to interaction with longitudinal acoustic phonons.

While, in recent years, tremendous progress has been witnessed towards achieving efficient and pure emission of indistinguishable photons, the simultaneous realization of near-unity efficiency and indistinguishability remains elusive. Fundamental challenges persist both on the design level and on the fabrication level making single-photon sources an interesting research topic most likely for many years to come.

References

Baba, T. 2008. Slow light in photonic crystals. *Nature Photonics* 2 (8): 465–473. doi:10.1038/nphoton.2008.146.

Babinec, T. M., B. J. M. Hausmann, M. Khan, Y. Zhang, J. R. Maze, P. R. Hemmer, and M. Lončar. 2010. A diamond nanowire single-photon source. *Nature Nanotechnology* 5 (3): 195–199. doi:10.1038/nnano.2010.6.

Balanis, C. A. 1989. *Advanced Engineering Electromagnetics*. 1st ed. Hoboken, NJ: Wiley.

Berthelot, A., I. Favero, G. Cassabois, C. Voisin, C. Delalande, P. Roussignol, R. Ferreira, and J-M. Gérard. 2006. Unconventional motional narrowing in the optical spectrum of a semiconductor quantum dot. *Nature Physics* 2 (11): 759–764. doi:10.1038/nphys433. http://arxiv.org/abs/cond-mat/0610346/nhttp://www.nature.com/doifinder/10.1038/nphys433.

Bleuse, J., J. Claudon, M. Creasey, N. S. Malik, J-M. Gérard, I. Maksymov, J-P. Hugonin, and P. Lalanne. 2011. Inhibition, enhancement, and control of spontaneous emission in photonic nanowires. *Physical Review Letters* 106 (10): 103601. doi:10.1103/PhysRevLett.106.103601.

Brendel, J., N. Gisin, W. Tittel, and H. Zbinden. 1999. Pulsed energy-time entangled twin-photon source for quantum communication. *Physical Review Letters* 82 (12): 2594–2597. doi:10.1103/PhysRevLett.82.2594. http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.82.2594.

Buttler, W. T., R. J. Hughes, S. K. Lamoreaux, G. L. Morgan, J. E. Nordholt, and C. G. Peterson. 2000. Daylight quantum key distribution over 1.6 Km. *Physical Review Letters* 84 (24): 5652–5655.

Carmichael, H. J. 1999. *Statistical Methods in Quantum Optics*. 1st ed. Berlin-Heidelberg: Springer.

Chen, X-W., S. Götzinger, and V. Sandoghdar. 2011. 99% efficiency in collecting photons from a single emitter. *Optics Letters* 36 (18): 3545–3547. doi:10.1364/OL.36.003545. http://ol.osa.org/abstract.cfm?URI=ol-36-18-3545.

Claudon, J., J. Bleuse, N. S. Malik, M. Bazin, P. Jaffrennou, N. Gregersen, C. Sauvan, P. Lalanne, and J-M. Gérard. 2010. A highly efficient single-photon source based on a quantum dot in a photonic nanowire. *Nature Photonics* 4 (3): 174–177. doi:10.1038/nphoton.2009.287. http://dx.doi.org/10.1038/nphoton.2009.287.

Claudon, J., N. Gregersen, P. Lalanne, and J-M. Gérard. 2013. Harnessing light with photonic nanowires: Fundamentals and applications to quantum optics. *ChemPhysChem* 14 (11): 2393–2402. doi:10.1002/cphc.201300033.

Coldren, L. A., S. W. Corzine, and M. L. Mashanovitch. 2012. *Diode Lasers and Photonic Integrated Circuits*. 2nd ed. Hoboken, NJ: Wiley.

Ding, X., Y. He, Z.-C. Duan, N. Gregersen, M.-C. Chen, S. Unsleber, S. Maier, et al. 2016. On-demand single photons with high extraction efficiency and near-unity indistinguishability from a resonantly driven quantum dot in a micropillar. *Physical Review Letters* 116 (2): 020401. doi:10.1103/PhysRevLett.116.020401. http://arxiv.org/abs/1601.00284.

Gazzano, O., S. Michaelis de Vasconcellos, C. Arnold, A. Nowak, E. Galopin, I. Sagnes, L. Lanco, A. Lemaître, and P. Senellart. 2013. Bright solid-state sources of indistinguishable single photons. *Nature Communications* 4: 1425. doi:10.1038/ncomms2434. http://www.ncbi.nlm.nih.gov/pubmed/23385570.

Gérard, J.-M. 2003. Solid-state cavity-quantum electrodynamics with self-assembled quantum dots. In *Single Quantum Dots, Topics in Applied Physics*, vol. 90, edited by P. Michler, 269–315. Berlin, Heidelberg: Springer-Verlag. doi:10.1007/978-3-540-39180-77. http://www.springerlink.com/index/4nn94hfbm1gpmyxq.pdf/nhttp://www.springerlink.com/content/4NN94HFBM1GPMYXQ.

Gerry, C. C. and P. L. Knight. 2008. *Introductory Quantum Optics*. New York, NY: Cambridge University Press.

Gregersen, N., P. Kaer, and J. Mørk. 2013. Modeling and design of high-efficiency single-photon sources. *IEEE Journal of Selected Topics in Quantum Electronics* 19 (5): 9000516. doi:10.1109/JSTQE.2013.2255265. http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6493378.

Hong, C. K., Z. Y. Ou, and L. Mandel. 1987. Measurement of subpicosecond time intervals between two photons by interference. *Physical Review Letters* 59 (18): 2044–2046. doi:10.1103/PhysRevLett.59.2044. http://journals.aps.org.globalproxy.cvt.dk/prl/abstract/10.1103/PhysRevLett.59.2044.

JCMWave. n.d. JCMSuite. http://www.jcmwave.com.

Kaer, P., N. Gregersen, and J. Mørk. 2013. The role of phonon scattering in the indistinguishability of photons emitted from semiconductor cavity QED systems. *New Journal of Physics* 15: 035027. doi:10.1088/1367-2630/15/3/035027.

Kaer, P. and J. Mørk. 2014. Decoherence in semiconductor cavity QED systems due to phonon couplings. *Physical Review B* 90 (3): 035312. doi:10.1103/PhysRevB.90.035312. http://journals.aps.org/prb/abstract/10.1103/PhysRevB.90.035312.

Kaer, P., T. R. Nielsen, P. Lodahl, A.-P. Jauho, and J. Mørk. 2012. Microscopic theory of phonon-induced effects on semiconductor quantum dot decay dynamics in cavity QED. *Physical Review B* 86 (8): 085302. doi:10.1103/PhysRevB.86.085302.

Kiraz, A., M. Atatüre, and A. Imamoğlu. 2004. Quantum-dot single-photon sources: Prospects for applications in linear optics quantum-information processing. *Physical Review A* 69 (3): 032305. doi:10.1103/PhysRevA.69.032305. http://journals.aps.org/pra/abstract/10.1103/PhysRevA.69.032305.

Kristensen, P. T., C. Van Vlack, and S. Hughes. 2012. Generalized effective mode volume for leaky optical cavities. *Optics Letters* 37 (10): 1649–1651. doi:10.1364/OL.37.001649.

Lavrinenko, A. V., J. Lægsgaard, N. Gregersen, F. Schmidt, and T. Søndergaard. 2014. *Numerical Methods in Photonics*. Boca Raton, FL: CRC Press. http://www.crcpress.com/product/isbn/9781466563889.

Lecamp, G., P. Lalanne, and J. P. Hugonin. 2007. Very large spontaneous-emission β factors in photonic-crystal waveguides. *Physical Review Letters* 99 (2): 023902. doi:10.1103/PhysRevLett.99.023902.

Lorke, M., T. Suhr, N. Gregersen, and J. Mørk. 2013. Theory of nanolaser devices: Rate equation analysis versus microscopic theory. *Physical Review B* 87 (20): 205310. doi:10.1103/PhysRevB.87.205310.

Lumerical. n.d. FDTD Solutions. http://www.lumerical.com/.

Novotny, L. and B. Hecht. 2012. *Principles of Nano-Optics*. 2nd ed. New York, NY: Cambridge University Press.

Purcell, E. M. 1946. Spontaneous emission probabilities at radio frequencies. *Physical Review* 69: 681.

Ramsay, A. J., A. V. Gopal, E. M. Gauger, A. Nazir, B. W. Lovett, A. M. Fox, and M. S. Skolnick. 2010. Damping of exciton rabi rotations by acoustic phonons in optically excited InGaAs/-GaAs quantum dots. *Physical Review Letters* 104 (1): 017402. doi:10.1103/PhysRevLett.104.017402. http://journals.aps.org/prl/abstract/10.1103/PhysRevLett.104.017402.

Reitzenstein, S. and A. Forchel. 2010. Quantum dot micropillars. *Journal of Physics D: Applied Physics* 43 (3): 033001. doi:10.1088/0022-3727/43/3/033001.

Shields, A. J. 2007. Semiconductor quantum light sources. *Nature Photonics* 1 (4): 215–223. doi:10.1038/nphoton.2007.46. http://www.nature.com/nphoton/journal/v1/n4/full/nphoton.2007.46.html.

Siyushev, P., G. Stein, J. Wrachtrup, and I. Gerhardt. 2014. Molecular photons interfaced with alkali atoms. *Nature* 509 (7498): 66–70. doi:10.1038/nature13191. http://www.ncbi.nlm.nih.gov/pubmed/24784217.

Somaschi, N., V. Giesz, L. De Santis, J. C. Loredo, M. P. Almeida, G. Hornecker, S. L. Portalupi, et al. 2016. Near-optimal single-photon sources in the solid state. *Nature Photonics* 10 (5): 340–345. doi:10.1038/nphoton.2016.23. http://dx.doi.org/10.1038/nphoton.2016.23.

Unsleber, S., D. P. S. McCutcheon, M. Dambach, M. Lermer, N. Gregersen, S. Höfling, J. Mørk, C. Schneider, and M. Kamp. 2015. Two-photon interference from a quantum dot microcavity: Persistent pure dephasing and suppression of time jitter. *Physical Review B* 91 (7): 075413. doi:10.1103/PhysRevB.91.075413.

Yao, P., V. S. C. Manga Rao, and S. Hughes. 2010. On-chip single photon sources using planar photonic crystals and single quantum dots. *Laser & Photonics Reviews* 4 (4): 499–516. doi:10.1002/lpor.200810081. http://doi.wiley.com/10.1002/lpor.200810081.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

47

Nanoplasmonic Lasers and Spasers

47.1 Introduction 609
47.2 Models for Metals and Gain 610
Drude–Lorentz Model • Multilevel Maxwell–Bloch Theory
47.3 Transfer Matrix Method 613
47.4 Evolutionary Algorithm 616
47.5 Finite-Difference Time-Domain Method 618
47.6 Loss Compensation 620
47.7 Lasing Dynamics 623
47.8 Stopped-Light Lasing 623
47.9 Summary 624

A Freddie Page
and
Ortwin Hess

47.1 Introduction

Semiconductor-based electronics has, in recent years, enjoyed a miniaturization trend down to structure sizes on nanometer scales. However, corresponding photonic devices, such as, in particular, lasers, have not been able to follow suit with a similarly dramatic pace [1]. This is because the generation of photons in small volumes appeared challenging due to the diffraction limit, which apparently seemed to set the minimum wavelength of light to be approximately inversely proportional to its frequency. Considering typical semiconductor band gaps with associated frequencies, this would imply that we would have a lower limit of optical devices and lasers of typically around 1 μm. Yet the availability of nanoscale lasers would be a true starting point for on-chip optical circuits, with applications in telecommunications, data storage, optical computing and sensing. So, can we shrink lasers to be much smaller than the wavelength of the light they emit, particularly down to the nanoscale?

Let us recall that lasers operate on the basis of two features: feedback and gain. In general, a cavity provides feedback by supporting resonant lasing modes (closed loops where light constructively interferes with itself on the round trip). On each loop, the gain medium amplifies the light, and when the amplification is greater than the threshold of energy lost (to dissipation in the material or by escaping the cavity), the system can enter a lasing regime. The physical size of a laser is thus dependent on the cavity supporting lasing modes and this, in turn, is limited by diffraction. Metallic cavities, on the other hand, allow for the circumvention of this diffraction limit by confining light in the subwavelength surface plasmon modes [2,3]. On a metal-insulator interface, light couples to the oscillations of surface charges as *surface plasmon polariton* (SPP) modes. But how does this help? There are two primary features of SPPs that make them useful in nano-optics applications: First, SPPs break the diffraction limit, i.e., they have wavevectors larger than what is generally allowed for a photon of a given frequency. Second, SPPs have or are associated with large field enhancements at the material interface. This combination of shorter wavelengths and higher field densities then significantly enhances the interaction of light with the electronic systems of active media

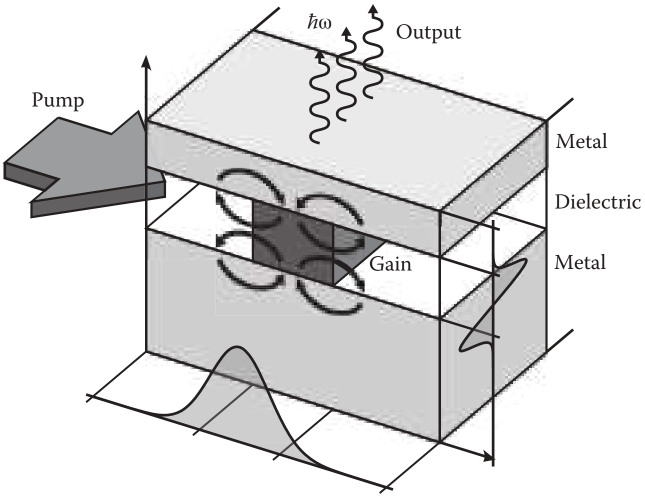


FIGURE 47.1 Nanoplasmonic laser illustration. Light is trapped in a metallic structure such that it forms a closed path (black arrows) that forms in overlap with a gain medium (dark box). In this setup, the gain is pumped from the side and emits from the top. A field strength profile is also shown superimposed on the structure.

[4,5], particularly those interactions that are usually weak. Having subwavelength confinement is, however, frequently traded off to a certain degree with not insignificant dissipative material losses in the metallic components, which requires a corresponding round-trip gain in order to be compensated [6].

Lasers that operate below the diffraction limit on nanometer scales by using metallic cavities are hence called nanoplasmonic lasers. A schematic illustration is given in Figure 47.1. Here there are two options for the light emission: Generated light can be channeled into photonic modes, as in a conventional laser, but can also generate surface plasmons, in which case the laser is more aptly described as a *spaser* for "surface plasmon amplification by stimulated emission of radiation" [7]. Indeed, the possibility to make nanoplasmonic lasers has been explored in various configurations such as, nanorod [8], double-fishnet [9], whispering gallery [10], and stopped light lasers [11].

This chapter will outline a range of techniques that enable the design, optimization, and simulation of nanoplasmonic lasers while tracking the example of a stopped light laser to outline concepts.

47.2 Models for Metals and Gain

47.2.1 Drude–Lorentz Model

In the following, we will chart our theoretical basis for describing the optical response of both metals (which provide subwavelength confinement) and semiconductor gain media (which provides the amplification in a nanoplasmonic laser), involving a combination of a Drude–Lorentz model and a multilevel Maxwell–Bloch theory, respectively, in the harmonic oscillator form. The Drude–Lorentz model describes the oscillations of charges in response to an applied electric field, while a Maxwell–Bloch approach is used to describe the behavior of dipole transitions such as in a four-level system model of a semiconductor.

In its most general terms, the Drude–Lorentz model relates the time evolution of the polarization of a material, **P**, to a driving electric field, **E**, by a second-order differential equation:

$$\frac{\partial^2 \mathbf{P}}{\partial t^2} + \gamma_\mathrm{L} \frac{\partial \mathbf{P}}{\partial t} + \omega_\mathrm{L}^2 \mathbf{P} = \omega_\mathrm{p}^2 \varepsilon_0 \mathbf{E} \,, \tag{47.1}$$

where γ_L is a damping term, ω_L is the restoring force frequency, and ω_p is the plasma frequency—the coupling strength to the driving **E** field. It is a causal response that captures the behavior of most resonance phenomena. In the frequency domain, it is expressed as

$$\tilde{\mathbf{P}} = \varepsilon_0 \frac{\omega_p^2}{\omega_L^2 - \omega(\omega + i\gamma_L)} \tilde{\mathbf{E}} \, , \tag{47.2}$$

where the Drude–Lorentz susceptibility can be extracted:

$$\chi_L(\omega) = \frac{\omega_p^2}{\omega_L^2 - \omega(\omega + i\gamma_L)} \, . \tag{47.3}$$

For $\gamma_L > 0$, this obeys the Kramers–Kronig relations relating real and imaginary parts as required by causality.

The Drude–Lorentz model describes noble metals well, and one may explain the origins of each of the terms with a simple model for a free electron gas. For each electron, let the displacement from its nucleus be given as $\mathbf{x}(t)$. A Newtonian force equation can be built assuming the electron acts as a free particle under the Lorentz force law of an external field **E**, i.e.,

$$m^* \frac{\partial^2 \mathbf{x}}{\partial t^2} = -e\mathbf{E} \, , \tag{47.4}$$

where m^* is the effective mass of an electron and e is its charge. Additional forces can be included, such as a collision term proportional to the velocity and inversely to an average collision time $F_{\mathrm{col}} = -\frac{\partial \mathbf{x}}{\partial t}/\tau$. In this free electron picture, there is no restoring force as for metals, electrons are not bound to their lattice sites. The macroscopic polarization can be related to these electron displacements, as each displacement induces an electronic dipole moment, and these can be averaged to become the polarization by multiplication by the electron number density, $\mathbf{P} = -en\mathbf{x}$. A comparison of terms allows for the associating:

$$\omega_p^2 \to \frac{ne^2}{m^*} \, , \; \gamma \to \tau^{-1} \, , \; \omega_L \to 0 \, . \tag{47.5}$$

This is the Drude model, and with an additional static background permittivity ε_{bg}, it describes metals well.

Real materials typically have multiple resonances, which can be incorporated by adding multiple Drude–Lorentz resonances:

$$\chi = \sum_i \frac{\omega_{pi}^2}{\omega_{Li}^2 - \omega(\omega + i\gamma_{Li})} \, . \tag{47.6}$$

These can have a variety of origins, i.e., in addition to the plasma mode discussed, rotational modes, electronic transitions, etc., and indeed the magnetic response, χ_m, can also be described as a sum of Drude–Lorentz resonances.

47.2.2 Multilevel Maxwell–Bloch Theory

A Drude–Lorentz (harmonic oscillator) type system can also model the electronic transitions between two levels of a semiconductor system, with an energy difference $\hbar\omega_e$. In this picture, there is, however, a field-dependent occupation density of upper and lower levels. Light interacts with this *two-level system* by stimulated emission and absorption processes. When the lower level is more occupied than the upper level, photons are absorbed and electromagnetic energy is dissipated within the system; however, when the reverse is true, photons can be emitted, stimulated by other photons, adding gain to the system. This dissipative/amplifying response is accompanied by a corresponding refractive response. The strength of the resonance is proportional to the *inversion density* of the two levels, i.e., how much more the higher level is occupied than the lower one, $\Delta N = N_2 - N_1$, such that a layer with emitters embedded may be represented by

$$\varepsilon = \varepsilon_{bg} + \frac{-\Delta N \omega_{pe}^2}{\omega_e^2 - \omega(\omega + i\gamma_e)}, \tag{47.7}$$

$$\omega_{pe}^2 = \sqrt{\varepsilon_{bg}}\gamma_e \sigma_e c, \tag{47.8}$$

where ε_{bg} is the permittivity of the layer hosting the resonance, σ_e is the emission cross-section, and γ_e is the width of the resonance [6]. Note that for negative inversion, the emitter becomes an absorber as there is a higher density of emitters in the lower state. In time, the inversion varies as light is emitted or absorbed. The auxiliary equations to take this into account are

$$\frac{\partial N_1}{\partial t} = -\frac{1}{\hbar\omega_e}\left(\frac{\partial \mathbf{P}_e}{\partial t} + \frac{\gamma_e}{2}\mathbf{P}_e\right) \cdot \mathbf{E} + \frac{N_2}{\tau_{21}}, \tag{47.9}$$

$$\frac{\partial N_2}{\partial t} = +\frac{1}{\hbar\omega_e}\left(\frac{\partial \mathbf{P}_e}{\partial t} + \frac{\gamma_e}{2}\mathbf{P}_e\right) \cdot \mathbf{E} - \frac{N_2}{\tau_{21}}, \tag{47.10}$$

with the first terms representing stimulated emission, and the final ones to account for non-radiative recombination at a rate τ_{21}^{-1}.

The two-level system is appropriate for modeling a single electronic transition; however, there is no way in which a two-level system can reach a state of inversion by relying solely on the processes of spontaneous emission and absorption—the best that can be achieved is equal occupation of the levels. A *four-level system*, on the other hand, can be constructed such that there can be a dynamically maintained population inversion between two of its levels, from where stimulated emission can occur. The construction contains two two-level systems, labeled e and a for emission and absorption; e with energy levels 1 and 2 and energy gap $\hbar\omega_e$, in-between a with energy levels 0 and 3 and greater energy gap $\hbar\omega_a$, as depicted in Figure 47.2. The two upper levels, 2 and 3, are coupled by a fast non-radiative relaxation channel, with rate τ_{32}^{-1}, as are the two lower levels, 0 and 1 (τ_{10}^{-1}). This has the effect of rapidly depleting the first and third levels shortly

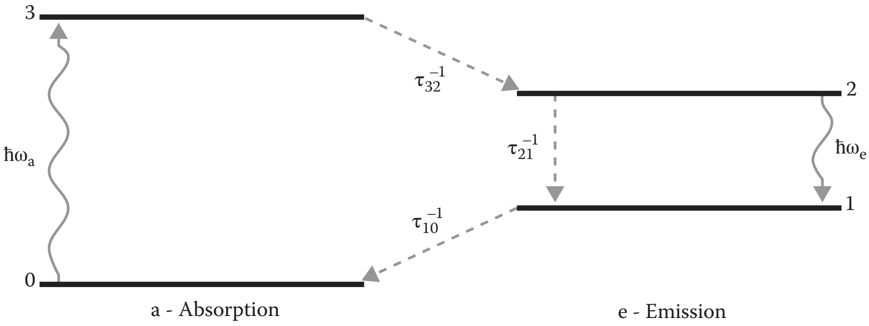


FIGURE 47.2 Schematic of the four-level system. A four-level system comprises two two-level systems, a and e, with active transitions (0 ↔ 3) and (1 ↔ 2), respectively, which are coupled by nonradioactive transition rates τ_{32}^{-1} and τ_{10}^{-1}. In this scheme, the emission subsystem permits radiative transitions at an energy of $\hbar\omega$ and has a slow nonradiative recombination rate τ_{21}^{-1}.

after they become occupied. Between the levels of the emission two-level subsystem, there is additionally a (slow) nonradiative channel (τ_{21}^{-1}).

The key point of a four-level system is that electrons that are pumped from levels 0 and 3, will quickly decay to level 2, leading to an inversion density of level 2 over level 1, which is available for stimulated emission. This allows the optical pumping between levels 0 and 3 to generate inversion between levels 1 and 2.

Four-level systems are incorporated into time-domain simulations using the differential equation for the polarization given in Equation 47.1 for each transition:

$$\frac{\partial^2 \mathbf{P}_\mathrm{a}}{\partial t^2} + \gamma_\mathrm{a} \frac{\partial \mathbf{P}_\mathrm{a}}{\partial t} + \omega_\mathrm{a}^2 \mathbf{P}_\mathrm{a} = -\Delta N \omega_\mathrm{pa}^2 \varepsilon_0 \mathbf{E}, \tag{47.11}$$

$$\frac{\partial^2 \mathbf{P}_\mathrm{e}}{\partial t^2} + \gamma_\mathrm{e} \frac{\partial \mathbf{P}_\mathrm{e}}{\partial t} + \omega_\mathrm{e}^2 \mathbf{P}_\mathrm{e} = -\Delta N \omega_\mathrm{pe}^2 \varepsilon_0 \mathbf{E}\,. \tag{47.12}$$

Here the polarization has been split off into a two parts $\mathbf{P}_\mathrm{a}$ and $\mathbf{P}_\mathrm{e}$ that are connected with the radiative resonances, which are added to the total polarization. The corresponding level occupation densities update with the auxiliary equations [6]

$$\frac{\partial N_3}{\partial t} = \left(\frac{\partial \mathbf{P}_\mathrm{a}}{\partial t} + \frac{\gamma_\mathrm{a}}{2}\mathbf{P}_\mathrm{a}\right) \cdot \mathbf{E} - \frac{N_3}{\tau_{32}}, \tag{47.13}$$

$$\frac{\partial N_2}{\partial t} = \frac{N_3}{\tau_{32}} + \frac{1}{\hbar\omega_\mathrm{e}}\left(\frac{\partial \mathbf{P}_\mathrm{e}}{\partial t} + \frac{\gamma_\mathrm{e}}{2}\mathbf{P}_\mathrm{e}\right) \cdot \mathbf{E} - \frac{N_2}{\tau_{21}}, \tag{47.14}$$

$$\frac{\partial N_1}{\partial t} = \frac{N_2}{\tau_{21}} - \frac{1}{\hbar\omega_\mathrm{e}}\left(\frac{\partial \mathbf{P}_\mathrm{e}}{\partial t} + \frac{\gamma_\mathrm{e}}{2}\mathbf{P}_\mathrm{e}\right) \cdot \mathbf{E} - \frac{N_1}{\tau_{10}}, \tag{47.15}$$

$$\frac{\partial N_0}{\partial t} = \frac{N_1}{\tau_{10}} - \left(\frac{\partial \mathbf{P}_\mathrm{a}}{\partial t} + \frac{\gamma_\mathrm{a}}{2}\mathbf{P}_\mathrm{a}\right) \cdot \mathbf{E}\,, \tag{47.16}$$

which encode both the stimulated terms and nonradiative recombinations.

47.3 Transfer Matrix Method

The geometry of the perhaps simplest types of nanoplasmonic lasers is planar waveguide structures. Here, planar layers of heterogeneous materials are stacked to form a waveguide in the center. The modes of such structures depend strongly on the width and material of each layer and can be calculated using the transfer matrix method (TMM), a linear, frequency-domain analysis that shall be used to design and optimize structures before simulation in time domain.

Isotropic homogeneous materials support transverse plane wave solutions in bulk. These waves have the dispersion relation

$$\varepsilon\mu\frac{\omega^2}{c^2} - k^2 = 0\,, \tag{47.17}$$

and the relative amplitudes of the electric and magnetic fields set by

$$E = \sqrt{\frac{\mu}{\varepsilon}} Z_0 H\,, \tag{47.18}$$

where Z_0 is the impedance of free space. When two materials meet at a planar interface, the solutions in each space must be matched together, and the prescription here is that the E and H field components parallel to the interface are continuous. Each interface couples a forward traveling and a backward traveling wave in each layer together.

Let us consider a stack with an axis parallel to the z-direction and waves traveling in a plane perpendicular to the y-direction. There are two independent polarization solutions, *transverse magnetic* (TM) where H is perpendicular to the plane of incidence (parallel to y) with E parallel to the plane, and *transverse electric* (TE) where E is perpendicular to the plane of incidence. TM equations shall be considered herein as this is the character of the SPP solutions. The corresponding TE equations can be recovered by making the substitutions: $\varepsilon \to \mu$, $\mu \to \varepsilon$, $\mathbf{E} \to -Z_0\mathbf{H}$, $Z_0\mathbf{H} \to \mathbf{E}$.

The system can be solved by performing a change of basis between forward and backward propagating waves and electric and magnetic field components, i.e.,

$$\begin{pmatrix} E_{xi} \\ Z_0 H_{yi} \end{pmatrix} = \underbrace{\begin{pmatrix} u_i & -u_i \\ 1 & 1 \end{pmatrix}}_{\mathbf{U}} \begin{pmatrix} A_i^+ \\ A_i^- \end{pmatrix}, \tag{47.19}$$

and translating the forward and backward field components within a layer:

$$\left.\begin{pmatrix} A_i^+ \\ A_i^- \end{pmatrix}\right|_z = \underbrace{\begin{pmatrix} e^{ik_{zi}(z-z')} & 0 \\ 0 & e^{-ik_{zi}(z-z')} \end{pmatrix}}_{\mathbf{T}} \left.\begin{pmatrix} A_i^+ \\ A_i^- \end{pmatrix}\right|_{z'}, \tag{47.20}$$

with $u_i = ck_{zi}/(\varepsilon_{xi}\omega)$. These are the matrices of the TMM. The matrices, $\mathbf{U}$ and $\mathbf{T}$, are assigned such that $\mathbf{U}$ converts from the basis of forward and backward traveling waves to electric and magnetic field components, and $\mathbf{T}$ propagates forward and backward traveling waves from a point z' to a point z.

For a stack of N planar slabs on a substrate as in Figure 47.3, each with its own material properties, we demonstrate how the TMM can determine the bound modes. Let the top interface of the ith slab be at z_i, vector components with that label, e.g., $A_i^\pm$, be measured from there, i.e., just below the top interface in that

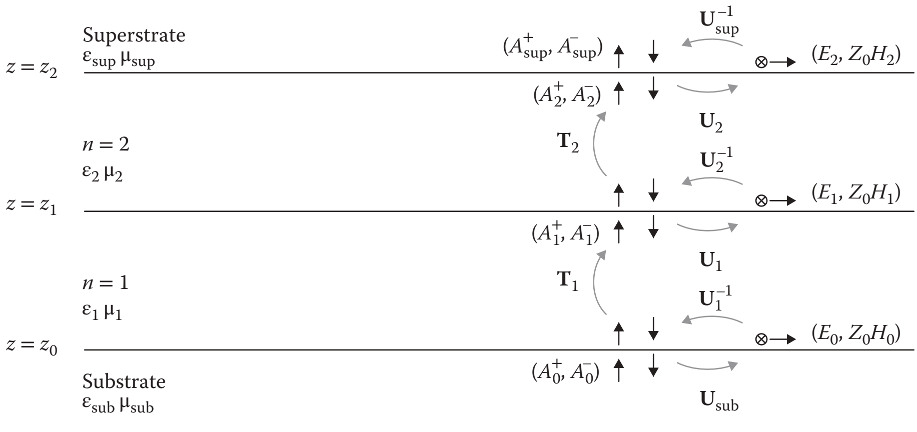


FIGURE 47.3 Schematic of the transfer matrix method. Fields (black arrows) are operated on by matrix transformations (gray arrows). The forward and backward components at the top of the substrate layer (A_0^+, A_0^-) are transferred through the stack to forward and backward components at the bottom of the superstrate layer (A^+_{sup}, A^-_{sup}). Fields move between bases of forward and backward waves (↑ ↓) and electric and magnetic components (⊗ →).

layer. Starting in the forward and backward basis, one may use the transfer matrices to propagate through the stack from the one layer to the next:

$$\begin{pmatrix} A_{i+1}^{+} \\ A_{i+1}^{-} \end{pmatrix} = \mathbf{T}_{i+1}\mathbf{U}_{i+1}^{-1}\mathbf{U}_{i}\begin{pmatrix} A_{i}^{+} \\ A_{i}^{-} \end{pmatrix} . \tag{47.21}$$

The exception to this being the superstrate, the $(N+1)$th layer, which by definition does not have a top interface. As such, $E_{\text{sup}}^{\pm}$ is defined specially as the forward and backward components just above the top interface of the Nth layer, i.e.,

$$\begin{pmatrix} A_{\text{sup}}^{+} \\ A_{\text{sup}}^{-} \end{pmatrix} = \mathbf{U}_{\text{sup}}^{-1}\mathbf{U}_{N}\begin{pmatrix} A_{N}^{+} \\ A_{N}^{-} \end{pmatrix} . \tag{47.22}$$

The full transfer matrix $\mathcal{M}$ will map forward and backward traveling waves in the substrate to the superstrate:

$$\begin{pmatrix} A_{\text{sup}}^{+} \\ A_{\text{sup}}^{-} \end{pmatrix} = \mathcal{M}\begin{pmatrix} A_{\text{sub}}^{+} \\ A_{\text{sub}}^{-} \end{pmatrix} , \tag{47.23}$$

and is given as the product of each layer hop:

$$\mathcal{M} = \mathbf{U}_{\text{sup}}^{-1}\mathbf{U}_{N}\cdots\mathbf{T}_{1}\mathbf{U}_{1}^{-1}\mathbf{U}_{\text{sub}} . \tag{47.24}$$

Terms with similar indices can be grouped to give a matrix for each layer:

$$\mathbf{M}_{i} = \mathbf{U}_{i}\mathbf{T}_{i}\mathbf{U}_{i}^{-1} , \tag{47.25}$$

which maps electric and magnetic fields at the bottom of the layer through to the top. As such, the transfer matrix for the whole stack is

$$\mathcal{M} = \mathbf{U}_{\text{sup}}^{-1}\left(\prod_{i=N}^{1}\mathbf{M}_{i}\right)\mathbf{U}_{\text{sub}} , \tag{47.26}$$

noting the reverse order on the product.

The eigenmodes of the system that have non-zero fields for a zero driving field, i.e.,

$$\begin{pmatrix} A_{\text{sup}}^{+} \\ 0 \end{pmatrix} = \mathcal{M}\begin{pmatrix} 0 \\ A_{\text{sub}}^{-} \end{pmatrix} , \tag{47.27}$$

with the prescription that $\operatorname{Im} k_z > 0$ in both the superstrate and substrate such that fields on both sides decay away to infinity. This yields the condition that $\mathcal{M}_{22} = 0$ for bound modes. For dispersive media, this is satisfied for either one of the frequency ω or wavevector q being a complex number, with the other remaining real. In general, analytical dispersion relation solutions do not exist, but rather solutions for $\mathcal{M}_{22}(\mathbf{q}, \omega) = 0$ are found implicitly, using a complex root-finding algorithm.

To illustrate the method, let's consider the SPP solution. Consider a metal substrate with a planar interface to a dielectric layer. In this case, the transfer matrix is given as

$$\mathcal{M} = \mathbf{U}_{\text{sup}}^{-1}\mathbf{U}_{\text{sub}} \tag{47.28}$$

$$= \frac{1}{2}\begin{pmatrix} \frac{\varepsilon_{\text{bg}}\omega}{ck_{z,\text{air}}} & 1 \\ -\frac{\varepsilon_{\text{bg}}\omega}{ck_{z,\text{air}}} & 1 \end{pmatrix}\begin{pmatrix} \frac{ck_{z,\text{metal}}}{\varepsilon(\omega)\omega} & -\frac{ck_{z,\text{air}}}{\varepsilon(\omega)\omega} \\ 1 & 1 \end{pmatrix}, \tag{47.29}$$

which gives the SPP condition:

$$\mathcal{M}_{22} = \frac{1}{2}\left(\frac{\varepsilon_{\text{bg}}\omega}{ck_{z,\text{air}}}\frac{ck_{z,\text{metal}}}{\varepsilon(\omega)\omega} + 1\right) = 0, \tag{47.30}$$

$$\frac{\varepsilon_{\text{bg}}}{\varepsilon(\omega)} + \frac{k_{z,\text{air}}}{k_{z,\text{metal}}} = 0\,. \tag{47.31}$$

One can see that for this equation to have solutions, given $\varepsilon_{\text{bg}} > 0$ and $\operatorname{Im} k_z > 0$, it must hold that $\operatorname{Re}\varepsilon(\omega) < 0$, which is characteristic of metals for frequencies below their plasma frequency.

For a lossless Drude metal, i.e., $\varepsilon(\omega) = \varepsilon_\infty - (\omega_{\text{p}}/\omega)^2$, an analytical solution exists:

$$q^2 = \varepsilon_{\text{bg}}\left(\frac{\omega}{c}\right)^2 \frac{\varepsilon_\infty\omega^2 - (\varepsilon_{\text{bg}} + \varepsilon_\infty)\omega_{\text{sp}}^2}{(\varepsilon_{\text{bg}} + \varepsilon_\infty)(\omega^2 - \omega_{\text{sp}}^2)}, \tag{47.32}$$

where $\omega_{\text{sp}} = \omega_{\text{p}}/\sqrt{\varepsilon_{\text{bg}} + \varepsilon_\infty}$ is the surface plasmon frequency.

47.4 Evolutionary Algorithm

For a planar stratified structure whose properties can be determined by a TMM, one may wish to optimize some aspect of its response to electromagnetic radiation. This may be in the dispersion relation of bound modes or in the transmission and reflection spectra. For example, to prescribe the frequency of a bound mode, or a reflection peak; or set whether there is only a single mode or many modes within a frequency range; or control the amplitude and phase response to an incoming wavepacket. These properties are entirely determined by the composition of the structure; however, it is often not obvious how changes in the structure lead to changes in the response.

The structures to be considered can be described by a finite number of numeric or discrete parameters, i.e., for each layer: the thickness of the layer, the material model to use, and the parameters of that model. This is well suited to optimization using an *evolutionary algorithm* (EA) [12].

In brief, an EA operates by storing multiple copies of these parameter sets, perturbing the values of the parameters randomly, ordering the sets by how the system they describe performs against a fitness metric, and keeping only the best variants to the next round.

More fully, in the language of an EA, this would be to say that a *population* of *chromosomes* is kept, and on each iteration of the algorithm, new chromosomes are generated from *mutations* of the current ones; the next-generation population is filled by *selection* of these chromosomes by a *fitness function*. The goal of the EA is to find a chromosome that extremizes the fitness function globally, and as such an EA is most suitable when the evaluation of the fitness function can be performed relatively quickly as a large number of members of the state space must be evaluated and compared.

A chromosome is simply the set of parameters that describes a system, i.e., a list of numbers or discrete option values. They are atomic entities, independent of other chromosomes. Each chromosome may have its fitness function evaluated to determine how suitable its corresponding physical system is to a problem.

Each iteration of the EA holds a set of chromosomes, ordered by their fitness function, known as a *population*. The first-generation population is either seeded with randomly generated structures, or structures that have been tuned by hand, perhaps by approximative theoretical methods. Populations of subsequent generations are constructed by first taking a small quantile of the fittest chromosomes from the previous iteration, with the remaining places filled with *mutations* of chromosomes selected from the previous set. The progress of the algorithm can be tracked by examining the fittest chromosome at the end of each iteration. At the end of each iteration, a *terminating condition* is evaluated, which determines if the algorithm should stop, returning the fittest chromosome, e.g., if the best structure in the population is within a defined tolerance of the target output, or if a set number of iterations have elapsed. If the terminating condition is not met, then the algorithm continues for another iteration. The operation of the EA is described as a flowchart in Figure 47.4.

The mutation process generates a new chromosome by randomly perturbing the values of parameters of a parent chromosome. It is by mutation that the state space of chromosomes is mapped.

For the layered structures considered here, the mutation process will perturb both the properties of the layers themselves and the relationship between layers. Numeric parameters, such as layer thicknesses and those used in material models, are perturbed by a Gaussian random variable, and the discrete parameters, such as the layer model, can be randomly selected from a list. Typically, only one or few parameters will be changed in each mutation. The parameters of the layers themselves do not sit in isolation. The number of layers in these structures is variable and mutation allows for adding or removing of layers or indeed swapping the order of layers and shifting the boundary between adjacent layers.

Whereas the EA is widely applicable to problems in nano-optics, the fitness function is not general and has to be constructed for each problem based on the type and kind of characteristics that are to be selected for. As an example, the EA can be applied to the case of stopped-light lasing. Here structures are desired that support a bound mode with two points of zero group velocity (ZGV) in their dispersion. An SL structure is deemed better if the ratio of the separations of the wavevectors and frequencies of the ZGV points, the band index $|c/v_b| = (k_2 - k_1)/(\omega_2 - \omega_1)$, is higher. This ensures that a wide range of wavevectors can be held at a single frequency, which reduces the group velocity for the whole band and the wave dispersion that will occur for a wavepacket made from Fourier components in the band. Figure 47.5 shows an example of

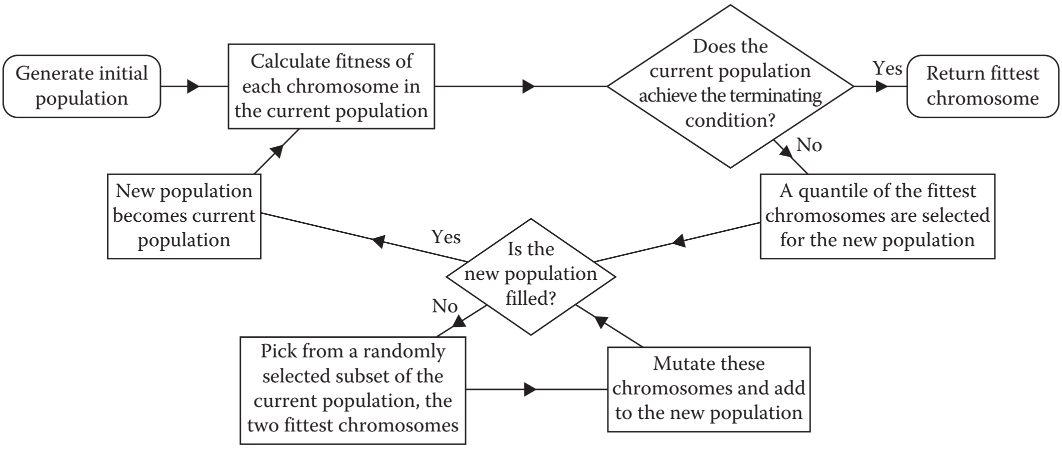
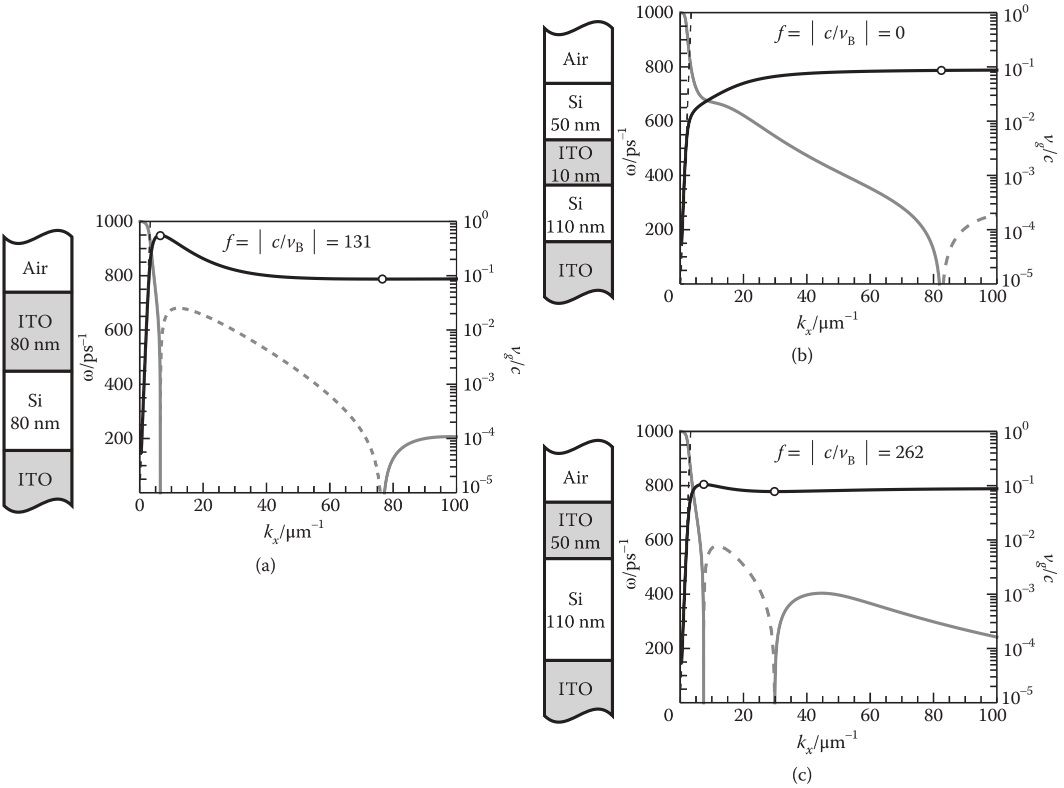

Generate initial population
Calculate fitness of each chromosome in the current population
Does the current population achieve the terminating condition?
Yes
Return fittest chromosome
No
A quantile of the fittest chromosomes are selected for the new population
New population becomes current population
Yes
Is the new population filled?
No
Pick from a randomly selected subset of the current population, the two fittest chromosomes
Mutate these chromosomes and add to the new population

FIGURE 47.4 Evolutionary algorithm flowchart.



(a)

(b)

(c)

FIGURE 47.5 Diagrams of planar plasmonic nanostructures and corresponding dispersion relations (black solid curve). The group velocity is also shown (gray solid, positive; gray dashed, negative) as well as zero group velocity points (circles). An evolutionary algorithm, using *band index* as a metric, given a. as a starting chromosome may produce b. and c. as child chromosomes. b. has a metric of zero and is thrown out, whereas c. may stay to the next generation as it has a higher metric than its parent.

potential SL structures, their dispersion relations, and group velocity. Each structure is labeled by its band index, and for a parent structure, two possible child structures are presented.

47.5 Finite-Difference Time-Domain Method

When describing processes such as gain, a frequency-domain analysis can only bring insight to a certain level as it has limitations to describe nonlinearities. When plasmons or phonons are emitted, electrons are demoted from higher energetic states to lower ones. This depletes the population inversion, reducing the available gain over time, leading to a nonlinear dependence in the field strength. In addition, spatially resolved effects, such as spatial hole burning, cannot be considered in TMM that assumes uniformity in the direction of propagation, whereas the level of inversion can vary both in this direction and perpendicular to the stacking. Depending on the mode formed, some regions may host higher field densities that can deplete the local gain. Thus a dynamic, spatially resolved, time-domain simulation is required to capture all the aspects of a nanoplasmonic lasing device. Here we will discuss and use the *finite-difference time-domain* (FDTD) method [13,14].

In FDTD, the goal is to solve Maxwell's equations for successive instants in time on a discretized lattice. This is approached by focusing on the curl equations and making the time derivatives of the fields the subject. Let us assume that there are no external sources:

$$\frac{\partial \mathbf{E}}{\partial t} = \frac{1}{\varepsilon_0}\left(\nabla \times \mathbf{H} - \frac{\partial \mathbf{P}}{\partial t}\right) \tag{47.33}$$

$$\frac{\partial \mathbf{H}}{\partial t} = -\frac{1}{\mu_0}\left(\nabla \times \mathbf{E} - \frac{\partial \mathbf{M}}{\partial t}\right) , \tag{47.34}$$

where **P** and **M** are the polarization and magnetization, respectively. Now, space and time are discretized as $(t, x, y, z) \to (n\Delta_t, i\Delta_x, j\Delta_x, k\Delta_x)$ and derivatives are turned into finite differences, i.e.,

$$\frac{\partial \mathbf{E}}{\partial t}(\mathbf{x}, t) \to \frac{\mathbf{E}^{n+1}_{i,j,k} - \mathbf{E}^{n}_{i,j,k}}{\Delta_t} . \tag{47.35}$$

Strictly, the above is a central difference and approximates the time derivative at a time $(n + 1/2)\Delta_t$. Therefore, the corresponding right-hand side terms must be evaluated at this time. The curl operator is handled similarly:

$$\nabla \times \mathbf{E}(\mathbf{x}, t)|_x \to \frac{E_z|^n_{i,j+1,k} - E_z|^n_{i,j,k}}{\Delta_x} - \frac{E_y|^n_{i,j,k+1} - E_y|^n_{i,j,k}}{\Delta_x} . \tag{47.36}$$

In order to make best use of this central difference, the lattice is arranged as a Yee grid [13], where fields are stored separately by component and **E** and **P** fields are specified on integer time coordinates, whereas **H** and **M** are specified on half-integer time coordinates. The lattice coordinates are more complicated; for **E**, components are stored at half-integer coordinates in their own direction and at integer coordinates otherwise. This is the opposite for the **H** field, as illustrated in Figure 47.6. From here, one can write the *update equations*:

$$\begin{aligned} E_x|^{n+1}_{i+\frac{1}{2},j,k} &= E_x|^{n}_{i+\frac{1}{2},j,k} - \varepsilon_0^{-1} P_x|^{n+1}_{i+\frac{1}{2},j,k} + \varepsilon_0^{-1} P_x|^{n}_{i+\frac{1}{2},j,k} \\ &+ \frac{c\Delta_t}{\Delta_x}\left(Z_0 H_z|^{n+\frac{1}{2}}_{i+\frac{1}{2},j+\frac{1}{2},k} - Z_0 H_z|^{n+\frac{1}{2}}_{i+\frac{1}{2},j-\frac{1}{2},k}\right) \\ &- \frac{c\Delta_t}{\Delta_x}\left(Z_0 H_y|^{n+\frac{1}{2}}_{i+\frac{1}{2},j,k+\frac{1}{2}} - Z_0 H_y|^{n+\frac{1}{2}}_{i+\frac{1}{2},j,k-\frac{1}{2}}\right) \end{aligned} \tag{47.37}$$

$$\begin{aligned} Z_0 H_x|^{n+\frac{1}{2}}_{i,j+\frac{1}{2},k+\frac{1}{2}} &= Z_0 H_x|^{n-\frac{1}{2}}_{i,j+\frac{1}{2},k+\frac{1}{2}} + cM_x|^{n+\frac{1}{2}}_{i,j+\frac{1}{2},k+\frac{1}{2}} - cM_x|^{n-\frac{1}{2}}_{i,j+\frac{1}{2},k+\frac{1}{2}} \\ &+ \frac{c\Delta_t}{\Delta_x}\left(E_z|^{n}_{i,j+1,k+\frac{1}{2}} - E_z|^{n}_{i,j,k+\frac{1}{2}}\right) \\ &- \frac{c\Delta_t}{\Delta_x}\left(E_y|^{n}_{i,j+\frac{1}{2},k+1} - E_y|^{n}_{i,j+\frac{1}{2},k}\right) , \end{aligned} \tag{47.38}$$

with corresponding equations for the omitted components.

With the updated equations of the primary fields determined, it now remains to be described how the auxiliary fields are updated. In general, the constitutive equations are functionals of all past values of all fields, though this would prove tricky to implement directly. In practice, material responses can be cast in terms of a finite difference if they can be expressed in terms of a temporal differential equation. This lends the Drude–Lorentz model well to inclusion within FDTD by converting its auxiliary equations into finite differences.

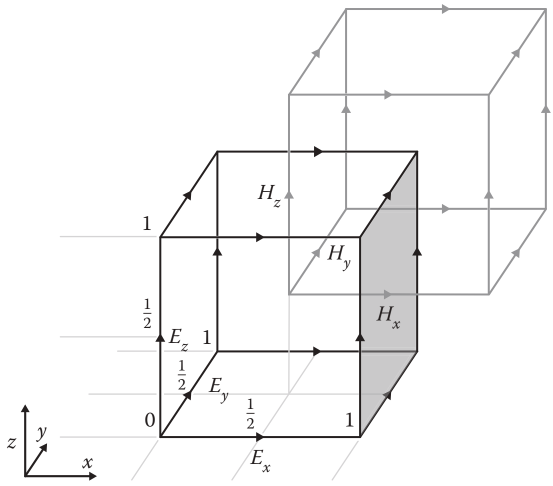


FIGURE 47.6 Schematic of FDTD Yee cell. Electric (black) and magnetic (gray) field grids. They are shifted half a step from each other, forming an interlocking link. Each field line passes through the center of a square formed from lines of the other field, e.g., H_x here passes through the center of the highlighted E-field face, where the finite difference $\nabla \times \mathbf{E}$ is defined.

The FDTD algorithm proceeds by iterating over, and updating in order, the **P** field, **E** field, **M** field, and **H** field over all space for each timestep.

Of course, there are limitations to FDTD. Its staggered grid structure means that material interfaces are not sharply defined, and any curvature on an interface will also experience a staircasing effect that can lead to spurious hotspots especially in plasmonic structures where geometric effects play a key role. The grid is also regular, which is to say it cannot be adaptively refined in areas of interest, rather all points must be simulated to the same resolution. This adds to the numerical load, as the smallest wavelength component of the fields, as well as the smallest geometrical features must be provided with sufficient resolution. This is again especially relevant in plasmonics, where field confinement is a feature.

Because of its nature as a full-wave time-domain solver, FDTD can often be the final say on a theoretical investigation. Short of doing a real-world experiment, FDTD numerical simulations find their utility as a heavy-duty tool. That is, the Maxwell's equations are solved at every discretized point in space, so it is often slower and more cumbersome, but can as a result of the inherent thoroughness produce results for questions that other techniques are unable to.

47.6 Loss Compensation

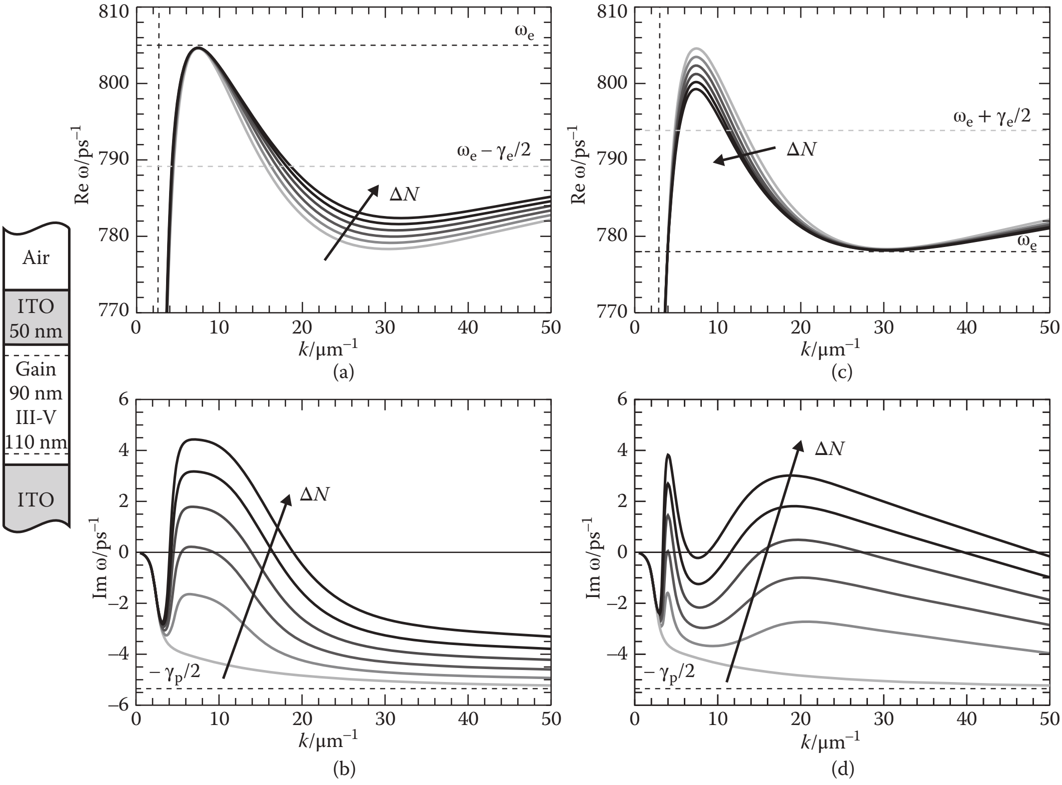
The high field enhancements afforded by plasmon modes supported by negative permittivity in metals must be traded off against Ohmic losses in the optical frequencies. Loss compensation is the incorporation of gain elements into a system in order to offset the loss in a particular frequency range while retaining desired properties from the lossy structure.

The properties one might wish to retain, e.g., negative effective index [6], stopped light points [11], or hyperbolic dispersion [15], etc., are frequency dependent. Causality will influence the position and width of the frequency range where losses can be compensated and desirable properties are to be found simultaneously [16,17], and adding gain to a system will necessarily change the refractive properties in accordance with the Kramers–Kronig relations.

The induced change in the mode structure by the addition of gain can be analyzed in a frequency-domain analysis, modeling gain as a four-level system in frequency domain with constant electron occupation densities. Using the TMM, this would mean replacing a part or whole of a dielectric layer with an equivalent Drude–Lorentz model with properties that map to the dielectric layer for equilibrium carrier occupation densities.

To illustrate this, the SL structure as optimized previously is modified by replacing the dielectric layer with a quantum emitter. The emission frequency is set to match either one of the ZGV points, and the other parameters are set representing the inclusion of realistic laser dye molecules [18]. The resulting dispersion and loss relations are shown in Figure 47.7 for inversion densities ΔN varying between 0 and N, with a fixed emitter density, and excitation about ZGV1 and ZGV2. The first point to note is that, even on full inversion, the addition of gain does not significantly change the dispersion, with the maximum shift in frequency being around 0.6%. The presence of ZGV points is preserved, though they may drift slightly, i.e., ZGV2 moves slightly right with emission about ZGV1. There is no change at the frequency that is being excited because the permittivity change of a Lorentzian is zero at the resonant frequency. Adding gain has the side effect of making the structure a slightly better SL structure as the band velocity is marginally reduced.

The key change though is in the loss. As the inversion density increases, the loss decreases for wavevectors where the dispersion is within the gain width. Initially the plasmonic loss is reduced as the inversion increases, and then for an inversion of around $\Delta N/N \approx 0.4$, some wavevectors become undamped, and even eventually experience gain. Plasmons sitting in these modes will grow exponentially in amplitude while small enough to remain in the small-signal gain regime. As ZGV2 is flatter than ZGV1, when the



(a) (c)

(b) (d)

FIGURE 47.7 Perturbation of dispersion relation and loss with Lorentzian gain. The change of mode shape plotted as Lorentzian gain is introduced to the system. For subfigures on the left, emission is about ZGV1, and on the right about ZGV2. a and b show a zoom in of how the dispersion relation changes with increasing inversion density, while c and d show the corresponding loss/gain of the mode.

emission is about this point, a wider range of frequencies fall within the gain width, leading to a larger range of k values that can become undamped.

An analysis of loss compensation can also be performed in the time domain. This has been done in Reference [6] where the structure was a negative refractive index double-fishnet metamaterial. Pump-probe spectroscopy was performed in FDTD in order to retrieve effective parameters using realistic parameters of rhodamine 800 dye molecules as a four-level system. It was shown that the imaginary part of the effective refractive index, which encodes the loss, transitioned from positive to zero as pumping increased while retaining a negative real part of the effective index between wavelengths of 706 and 716 nm. The results are shown in Figure 47.8.

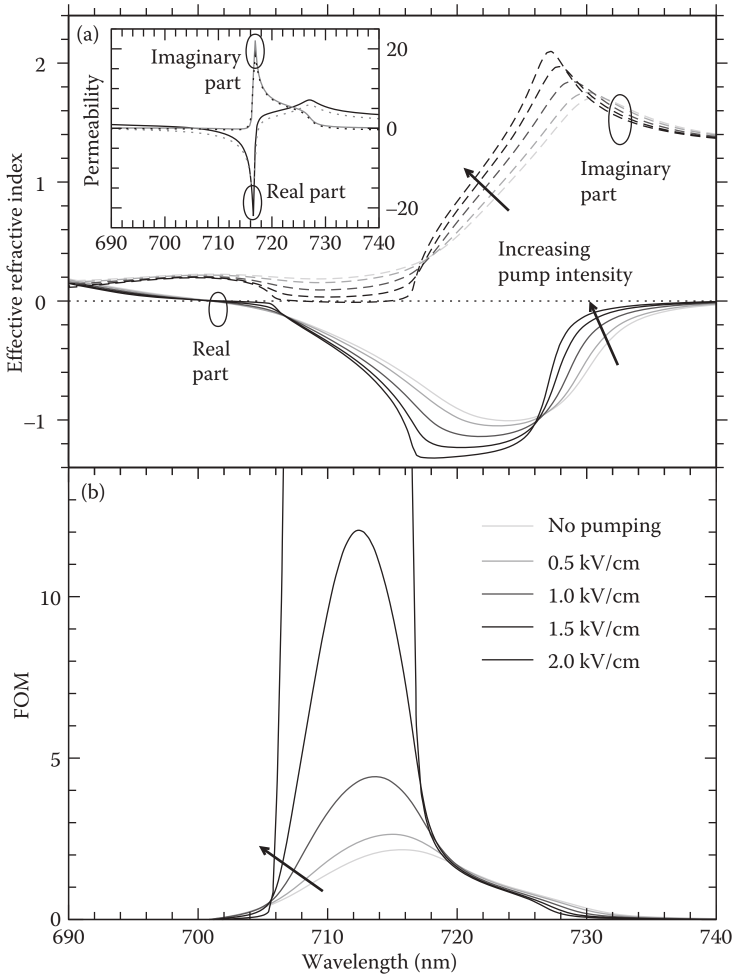


FIGURE 47.8 (a) Real and imaginary part of the retrieved effective refractive indices of the double-fishnet structure for different pump amplitudes. The peak electric field amplitude of the pump increases in steps of 0.5 kV/cm from no pumping (lightest) to a maximum of 2.0 kV/cm (darkest). The inset shows the real and imaginary part of the effective permeability (black and gray line, respectively) and the result of the Kramers–Kronig relation (dotted lines) for the highest peak electric field amplitude of 2.0 kV/cm . (b) The figures of merit (FOM) for the same pumping amplitudes. (Reproduced from Sebastian Wuestner et al., *Phys. Rev. Lett.*, 105,127401, 2010.)

47.7 Lasing Dynamics

In order to transition from loss compensation to lasing, photons emitted by the gain medium must feedback and stimulate on average more emission than photons that are lost on the round trip. Photons can be lost by escaping the cavity to radiation modes, absorbed incoherently by dissipative channels, or absorption by the gain medium.

Initially, for a cold cavity with no lasing mode field built up, an inverted gain system may spontaneously emit into cavity modes. This spontaneous emission may trigger further stimulated emission in a process known as *amplified spontaneous emission* (ASE), which can be added to the semiclassical FDTD description by including spatially resolved dynamic Langevin noise to the system, which accounts for the dissipative reservoirs feeding back stochastically on the system. The noise couples to the four-level system and induces incoherent transitions, which then become amplified, allowing for the triggering of the lasing regime [8].

As a device enters a lasing regime, characteristic relaxation oscillations can be seen where inversion initially builds up and spontaneous emission events are induced. The fields then are amplified as they stimulate further emission, growing the field in the mode coherently. From here on the emitted fields start to grow exponentially, and when they are of sufficient strength will deplete the inversion density. This reduction of inversion feeds back by decreasing the available gain, leading to a decrease in field energy as the energy in the field is lost to dissipative processes in the metal layers. The decrease in energy density allows for the inversion to rebuild. This continues in an oscillatory manner with the energy density lagging behind the inversion by 90°. The amplitude of the inversion and energy oscillations decrease with each cycle until a stable steady value for both is reached. In contrast, the control structure, albeit entering a regime of relaxation oscillations, is more erratic and the oscillations do not settle to a steady state. Instead, the oscillations continue with a factor of 4 between the peak energy density and the trough. An example of these relaxation oscillations in shown for the SL laser set in Figure 47.9 IN contrast to a “control structure” where the waveguide is not of high enough quality for the device to enter the lasing regime. Relaxation oscillations are accompanied by a characteristic decrease of the width of the power spectrum, from a wide peak in ASE to a sharp one in the lasing regime. This observation is usually taken as a crucial evidence for lasing in experimental setups.

47.8 Stopped-Light Lasing

Stopped-light lasing differs from traditional nanoplasmonic lasers in that there is no cavity with predefined modes. Whereas in the structures presented, light is trapped in a waveguide mode, it is free to propagate in one direction, albeit at ZGV. This means that there is a continuum of planewave modes for the gain to emit into, which get selected on the basis of the overlap with the gain media. The feedback in the system can be understood by considering the spatially resolved, cycle-averaged Poynting vector, energy density, and inversion that are plotted in Figure 47.10 for a structure with a gain width of 1000 nm. The energy is concentrated on metal-dielectric interfaces, strongest on the lower interface, and is localized around the gain medium despite there being no cavity along the horizontal direction.

The Poynting vector shows how energy circulates in the structure. There are four energy flux vortices, one in each corner of the structure; as described in the figure caption, these have energy moving out of the gain region into the metal layers and into it in dielectric layers. Considering the x component of the Poynting flux Figure 47.10a, the forward and backward flows are in exact balance. The balanced counter-propagation of energy in the negative-permittivity (metal) layers against the dielectric layer is the basis of the (subwavelength) feedback in the system.

The inversion is shown in Figure 47.10d; there are areas of spatial hole burning where the energy density is highest. The spatial modulation seen is explained when considering the field profile and its formation.

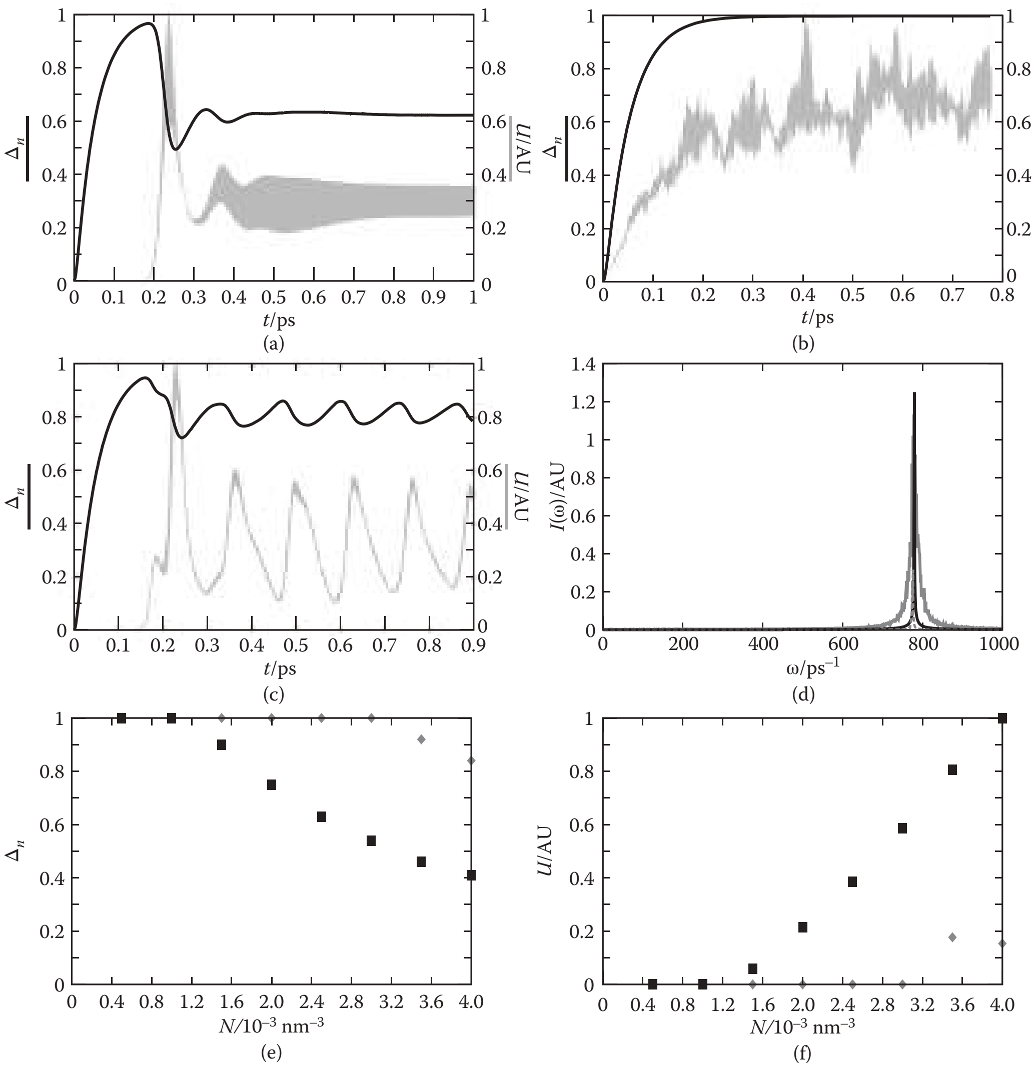
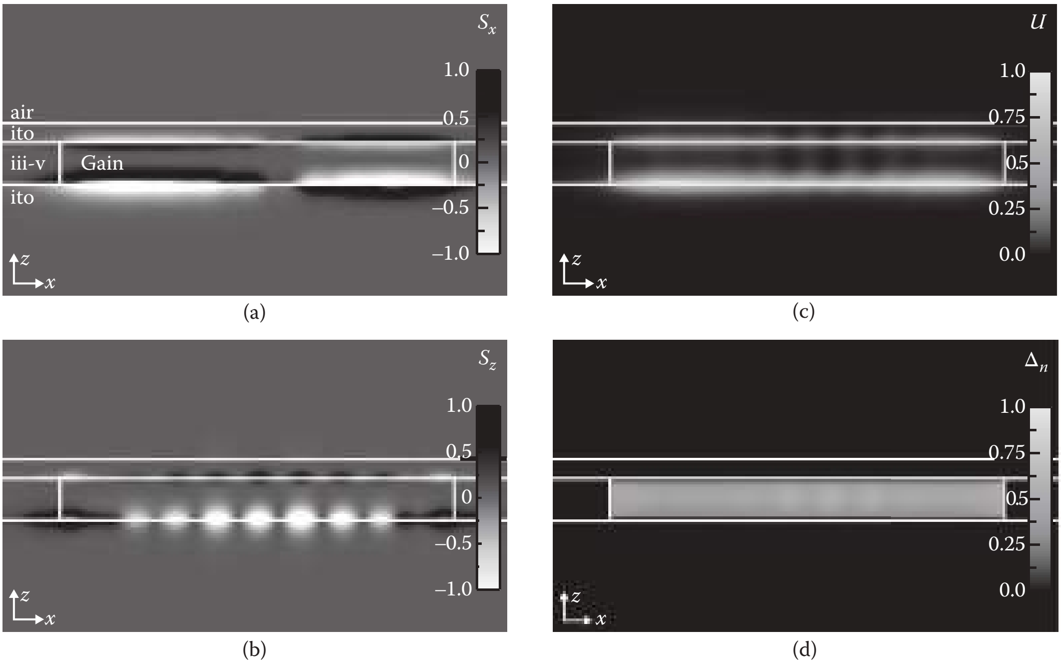


FIGURE 47.9 Lasing onset with increasing gain density. a, b, c. Energy density and mean inversion for (a) the stopped light structure with a gain density $N = 0.002$ nm^{-3}, (b) the stopped light structure with a gain density $N = 0.001$ nm^{-3}, and (c) the control structure with a gain density $N = 0.004$ nm^{-3}. (d) Power spectra for SL structure at $N = 0.002$ nm^{-3} (black), SL structure at $N = 0.001$ nm^{-3} (×5000 dashed gray), and control structure at $N = 0.004$ nm^{-3} (solid gray). (e) Steady-state inversion of SL (black) and control (gray) structures with varying gain density. (f) Steady-state cycle-averaged energy density with varying gain density.

47.9 Summary

In this chapter, we have explored techniques to design, optimize, and simulate a nanoplasmonic laser. We started by defining harmonic oscillator type Drude–Lorentz and Maxwell–Bloch models, a universal type of models for metals, semiconductors, and gain media. From a simple model of forces in time domain, a causal frequency-domain response function was derived. From here, the four-level system was introduced as two dipole transitions in the frequency domain and as a dynamic system with self-consistently evolving



(a) (c)

(b) (d)

FIGURE 47.10 Energy flux and density of the lasing mode. The lasing mode of a stopped-light laser structure with a width 1000 nm. (a,b) The x and z components of the cycle-averaged Poynting vector, respectively, with positive values in black and negative in white. Focusing on the bottom right lobe of the flux in panel a, energy flows into the gain region in the dielectric layers and out of it in the metal layer. For the same lobe in panel b, energy moves downward out of the gain region toward the center, and upward into the region at the edge. The counter-clockwise energy vortex in this corner is mirrored in the other corners. (c) The cycle-averaged energy density and (d) inversion plots. The inversion is the complement of the energy density, being highest in areas with low energy density and depleted where the mode sits.

electron occupation levels in the time domain. The TMM was next introduced as an exact frequency-domain tool to analyze planar structures. EA are suited to optimize systems that can be characterized by a range of numeric or discrete parameters and perform characterized by a fitness metric. The EA approach was introduced to and demonstrated in the case of the TMM and applied to designing stopped light laser structures. Next, FDTD, the time-domain numerical method, was discussed and used to simulate active hybrid metallo-dielectric structures with nonlinearities and time-dependant dispersion such as a lasing system. This allowed for the simulation of lasing dynamics where loss compensation, ASE, and relaxation oscillations could be captured. Finally, some specifics of the recently established stopped light laser principle were explored as an example system of a designed and optimized nanoplasmonic laser.

References

1. Jennifer A. Dionne, Luke A. Sweatlock, Matthew T. Sheldon, A. Paul Alivisatos, and Harry A. Atwater. Silicon-based plasmonics for on-chip photonics. *IEEE Journal of Selected Topics in Quantum Electronics*, 16(1):295–306, 2010.
2. Anatoly V. Zayats and Igor I. Smolyaninov. Near-field photonics: Surface plasmon polaritons and localized surface plasmons. *Journal of Optics A: Pure and Applied Optics*, 5(4):S16, 2003.
3. Dmitri K. Gramotnev and Sergey I. Bozhevolnyi. Plasmonics beyond the diffraction limit. *Nature Photonics*, 4(2):83–91, 2010.

4. Kevin F. MacDonald, Zsolt L. Samson, Mark I. Stockman, and Nikolay I. Zheludev. Ultrafast active plasmonics. *Nature Photonics*, 3(1):55–58, 2009.
5. Ortwin Hess, John B. Pendry, Stefan A. Maier, Rupert Oulton, Joachim M. Hamm, and Kosmas L. Tsakmakidis. Active nanoplasmonic metamaterials. *Nature Materials*, 11(7):573–584, 2012.
6. Sebastian Wuestner, Andreas Pusch, Kosmas L. Tsakmakidis, Joachim M. Hamm, and Ortwin Hess. Overcoming losses with gain in a negative refractive index meta-material. *Physical Review Letters*, 105(12):127401, 2010.
7. David J. Bergman and Mark I. Stockman. Surface plasmon amplification by stimulated emission of radiation: Quantum generation of coherent surface plasmons in nanosystems. *Physical Review Letters*, 90:27402, 2003.
8. Andreas Pusch, Sebastian Wuestner, Joachim M. Hamm, Kosmas L. Tsakmakidis, and Ortwin Hess. Coherent amplification and noise in gain-enhanced nanoplasmonic metamaterials: A Maxwell-Bloch Langevin approach. *ACS Nano*, 6(3):2420–2431, 2012.
9. Sebastian Wuestner, Joachim M. Hamm, Andreas Pusch, Fabian Renn, Kosmas L. Tsakmakidis, and Ortwin Hess. Control and dynamic competition of bright and dark lasing states in active nanoplasmonic metamaterials. *Physical Review B*, 85(20):201406, 2012.
10. Bumki Min, Eric Ostby, Volker Sorger, Erick Ulin-Avila, Lan Yang, Xiang Zhang, and Kerry Vahala. High-Q surface-plasmon-polariton whispering-gallery microcavity. *Nature*, 457(7228):455–458, 2009.
11. Tim W. Pickering, Joachim M. Hamm, A. Freddie Page, Sebastian Wuestner, and Ortwin Hess. Cavity-free plasmonic nanolasing enabled by dispersionless stopped light. *Nature Communications*, 5:4972, 2014.
12. William M. Spears. *Evolutionary Algorithms: The Role of Mutation and Recombination. Natural Computing Series*. Heidelberg: Springer Berlin Heidelberg, 2000.
13. Kane Yee. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Transactions on Antennas and Propagation*, 14:302–307, 1966.
14. Allen Taflove. *Computational Electrodynamics: The Finite-Difference Time-Domain Method. Antennas and Propagation Library*. Norwood, MA: Artech House, 1995.
15. Xingjie Ni, Satoshi Ishii, Mark D. Thoreson, Vladimir M. Shalaev, Seunghoon Han, Sangyoon Lee, and Alexander V. Kildishev. Loss-compensated and active hyperbolic metamaterials. *Optics Express*, 19(25):25242–25254, 2011.
16. Johannes Skaar. Fresnel equations and the refractive index of active media. *Physical Review E*, 73(2):26605, 2006.
17. Paul Kinsler and Martin McCall. Causality-based criteria for a negative refractive index must be used with care. *Physical Review Letters*, 101(16):1–4, 2008.
18. P. Sperber, W. Spangler, Bernd Meier, and Alfons Penzkofer. Experimental and theoretical investigation of tunable picosecond pulse generation in longitudinally pumped dye laser generators and amplifiers. *Optical and Quantum Electronics*, 20(5):395–431, 1988.

48

Quantum-Dot Nanolasers

48.1 Going Nano: A New Era of Small Lasers 627
48.2 Connecting to Earlier Work: Laser Rate Equations and the β Factor 628
48.3 Key Aspects of Nanolasers.............................. 631
48.4 Essential Physics I: Quantum-Optical Semiconductor Laser Theory 635
Dynamical Laser Equations • Pump Process for Continuous-Wave-Excited Lasers
48.5 Essential Physics II: Carrier Kinetics in QD Systems.............. 639
Carrier–Phonon Interaction • Coulomb Interaction
48.6 Connecting to Experiments: Transition to Lasing and the β Factor 646
48.7 Interemitter Coupling Effects and Superradiance in Nanolasers 647
48.8 Nonresonant Coupling and Lasing from Multiexciton States in Few-QD Systems.............................. 651
48.9 Outlook.............................. 653

Christopher Gies

Michael Lorke

Frank Jahnke

and

Weng W. Chow

48.1 Going Nano: A New Era of Small Lasers

The impact of semiconductor lasers on our daily lives is immense. They perform work every time someone gathers information for the internet, makes a telephone call, prints an article or pays for an item at a store. There are over 2.5 billion VCSELs[†] in use at our homes or offices, and roughly 100 million are produced each year to keep up with demand. VCSEL technology is now considered conventional, and research has progressed to smaller devices, modulated at higher speeds and producing output with greater spectral stability. This chapter focuses on two of the advances. One is the reduction of optical cavity volume by as much as two orders of magnitude, making a transition from microlasers (such as VCSELs) to nanolasers. Two is a change in active medium from quantum wells (QWs) to quantum dots (QDs), bringing about, e.g., devices operating with very few (tens) of emitters. These developments ushered in a new era for semiconductor device physics, one where quantum optical and many-body electron interaction effects dominate. The underlying motivation is the control of spontaneous emission [1]. The typical laser mitigates the randomness (noise) caused by spontaneous emission by overwhelming it with stimulated emission. Here, we are speaking of actually quieting the spontaneous emission noise, both spatially and temporally. In the former, we use nanocavities to inhibit spontaneous emission in undesirable directions. With the latter, we use very few QDs (ideally only one) to control the timing of photon emission, thereby improving photon statistics beyond the limit described by the Poisson distribution. Devices incorporating the two advances are being fabricated and experiments are being performed. The results are both promising and exciting from the device and physics aspects, respectively. They also lead to many new questions and renewed interest in some old ones: What is lasing and where is the threshold? Why is there not a phonon bottleneck? What is the homogeneous width of a QD transition [2–4]? What is the inhomogeneous broadening in my samples? Is there really thresholdless lasing [5–7]? Are QD lasers better than QW lasers [8]? Why do QD lasers not show modulation speeds as advertised? This chapter describes a theoretical framework capable of addressing all the abovementioned questions. The building blocks come from quantum electrodynamics (QED), which is the quantum theory of the interaction of light with matter, and carrier interactions described by many-body theory.

[†] Vertical cavity surface emitting lasers, cf. Chapter 34.

The fundamental concepts of lasing are often discussed in terms of rate equations for material excitations and cavity photons that may well be familiar to the reader. We begin this chapter with an excursion to show profit and limitations of laser rate equations for understanding the behavior of nanolasers, and at the same time establish a connection to more sophisticated semiconductor laser models. In Section 48.3, we give a short overview of the key aspects in modeling nanolasers and the characterization of their emission properties. In Section 48.4, we introduce a microscopic laser model for solid-state emitters that goes beyond the rate equations and that provides access to the quantum-statistical properties of the emission. A central aspect of device performance is determined by the efficiency of carrier scattering processes following excitation. Section 48.5 provides an overview of the underlying mechanisms of carrier–carrier and carrier-phonon interaction and explains methods to incorporate the dynamics into semiconductor laser models within different levels of approximation. The underlying formalism allows one to address a wide variety of effects present in semiconductor nanolasers, and we provide a few examples in Sections 48.6 and 48.7, where QD lasing and radiative coupling between individual emitters are discussed. Section 48.8 gives an outlook toward systems with only a few QDs inside the cavity, where the individual electronic many-particle properties of each emitter matter, and where device properties are dominated by quantum and correlation effects.

48.2 Connecting to Earlier Work: Laser Rate Equations and the β Factor

Lasing operation takes place when pumping of the active material provides sufficient gain to compensate the losses. It is thus determined by the interplay of the excitation and emission dynamics of the gain material, taking into account the feedback provided by the radiation field inside the resonator. A basic understanding of lasing action can be obtained from laser rate equations in the form of two coupled equations for the gain (in this case, the number of excitons N in the QD ensemble) and the intracavity photon number n:

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} N &= P - \gamma_l (n+1)N - \gamma_{\mathrm{nl}} N,\\ \frac{\mathrm{d}}{\mathrm{d}t} n &= -\kappa n + \gamma_l (n+1)N.\end{aligned} \tag{48.1}$$

Here, P and κ are the pump and photon loss rates, γ_l the spontaneous emission rate into the laser mode, γ_{nl} the rate of radiative recombination into other modes, and $\gamma_{\mathrm{sp}} = \gamma_l + \gamma_{\mathrm{nl}}$ the total spontaneous emission rate. The bracket $(n + 1)$ contains both the stimulated and spontaneous emission contributions. The ratio

$$\beta = \frac{\gamma_l}{\gamma_{\mathrm{sp}}} \tag{48.2}$$

defines the β-factor illustrated in Figure 48.1. It is a key characteristic quantity of any laser device as it defines device performance in the spontaneous-emission regime (i.e., below threshold) and, more importantly, the threshold current. The rate equations (Equation 48.1) are often expressed in terms of β:

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t} N &= P - \beta\gamma_{\mathrm{sp}}(n+1)N - (1-\beta)\gamma_{\mathrm{sp}} N, \\ \frac{\mathrm{d}}{\mathrm{d}t} n &= -\kappa n + \beta\gamma_{\mathrm{sp}}(n+1)N. \end{aligned} \tag{48.3}$$

In conventional lasers, $\beta \approx 10^{-5}$ [9], so that the majority of spontaneously emitted photons are lost. Once the system starts lasing at sufficiently high excitation, stimulated emission completely overwhelms spontaneous emission, so that the fraction of photons lost into nonlasing modes becomes insignificant. This behavior can be directly traced by accessing both emission channels separately. In micropillar lasers, the fundamental mode often used as laser mode has a far-field emission in vertical direction of the pillar, whereas losses occur laterally through the pillar side walls. The input–output curve (photon number versus pump rate) collected separately from both directions, together with a schematic of the experiment, is shown in Figure 48.2 [10]. As long as spontaneous emission is the dominating emission process, both emission intensities increase linearly with pump with a fixed ratio determined by the β factor. At higher excitation powers, stimulated emission from the laser mode sets in and photon emission from the laser mode increases nonlinearly and more strongly than the emission into loss channels. The resulting increase of the normalized ratio *r* of both emission channels therefore gives direct account of the onset of stimulated emission. Note that the S shape of the intensity curve for the lateral emission is due to stray light from the laser mode that is collected in lateral direction. Ideally, a clamped intensity is expected.

When continuous excitation is considered, the rate equations are solved in the steady state. Input–output curves obtained from the rate equations (Equations 48.1) for different values of β are shown in Figure 48.3. For $\beta \ll 1$, a sudden increase in emission intensity is visible that allows one to identify the lasing threshold. With increasing coupling efficiency β, the threshold shifts to lower pump rates, offering the prospect of devices with reduced threshold currents. At the same time, the intensity jump decreases until the "thresholdless" case is obtained for $\beta = 1$. In this regime an identification of the transition into lasing is impossible from the emission intensity alone. A characterization of the threshold requires going beyond the rate equation approximation to study statistical and coherence properties of the emission. One of the attractiveness of nanolasers is that they are capable of operating in a regime close to $\beta = 1$ due to their small effective mode volumes and large cavity lifetimes that cause strong funneling of spontaneously emitted photons into a single resonator mode [1]. At the same time, the photonic density of states is reduced away from the cavity-mode frequency, which further suppresses emission into nonlasing modes and thereby also

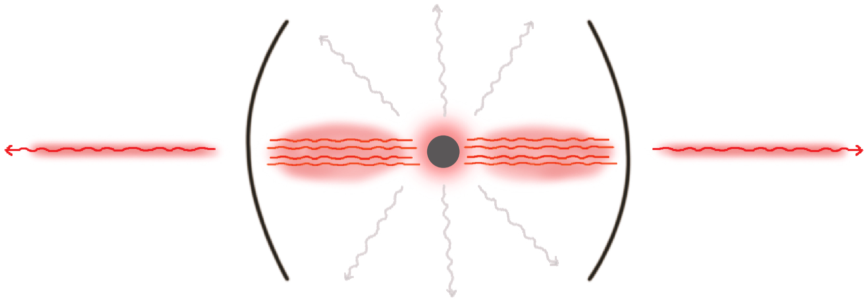


FIGURE 48.1 Spontaneous emission of the emitter into the laser mode (black lines) and into nonlasing modes (gray lines). The ratio of the spontaneous emission into the laser mode and the total spontaneous emission is described by the β factor.

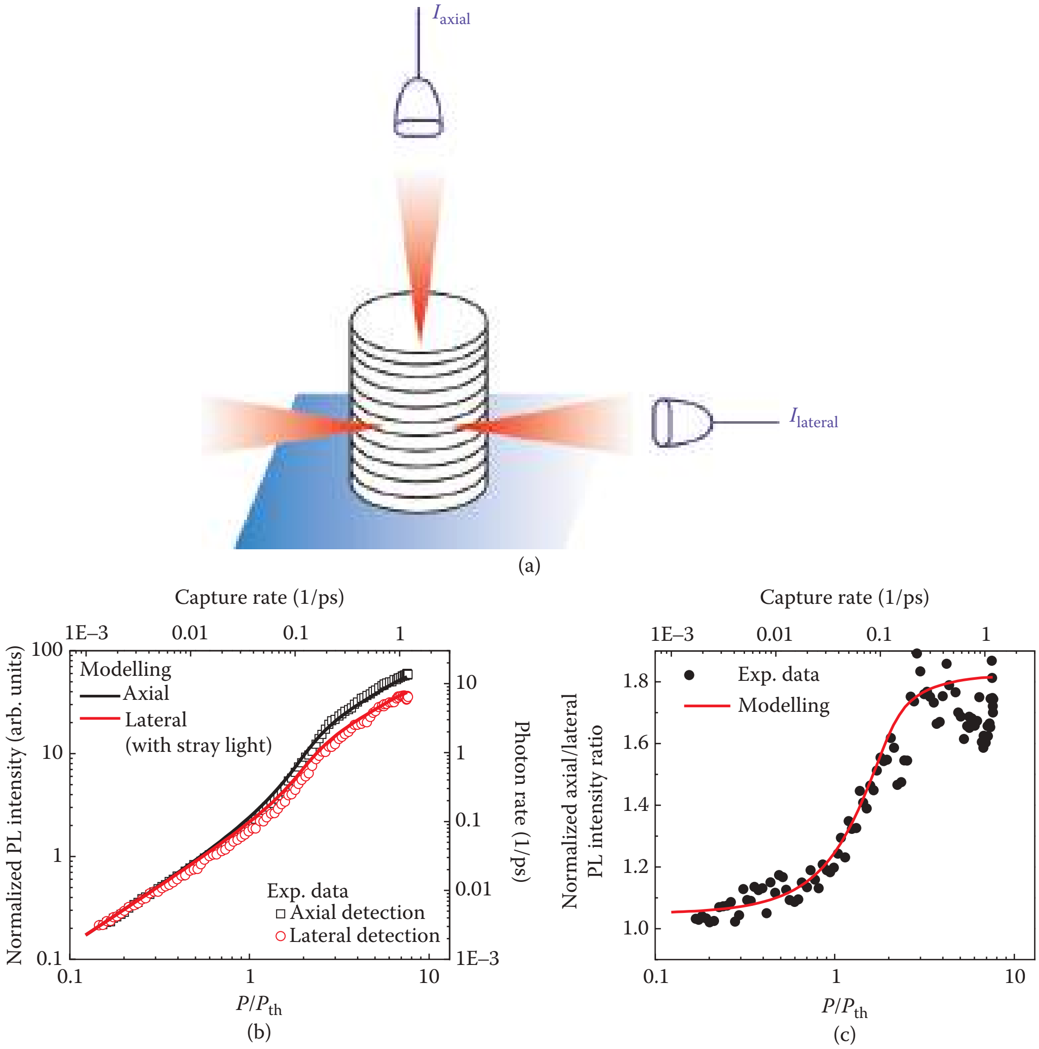


FIGURE 48.2 (a) Sketch of an experiment for detecting laser-mode emission in axial direction, and leaky mode emission in lateral direction separately. (b) Input–output characteristics of a quantum dot (QD) micropillar laser. Emission intensity into the laser mode (axial direction) and into nonlasing modes (lateral direction) are shown separately. (c) Their ratio r increases at the onset of stimulated emission, providing a direct visualization of the transition into the lasing regime. The results have been obtained from the semiconductor laser theory introduced in Section 48.4. Note that in lateral emission some scattered light from the axial emission is collected, which is responsible for the unclamped intensity behavior. [PL, photoluminescence. (Figures b and c adapted from Musiał, A. et al., *Physical Review B*, 91, 205310, 2015.)]

increases the β factor. While cavities operate at a specific narrow frequency window, structures that cause a broadband enhancement of spontaneous emission without using a cavity, such as photonic trumpets [11], are based on the very same concept.

The rate equations (Equation 48.1) have further drawbacks for modeling QD-based nanolasers. Most importantly, they do not provide information on laser field coherence and photon correlations. The rate equations are obtained by neglecting all correlations between light and matter degrees of freedom. Due to the enhanced interaction inside the microresonator, photon, and carrier-photon correlations have been demonstrated to play an important role in nanolasers, and this is where new approaches are required

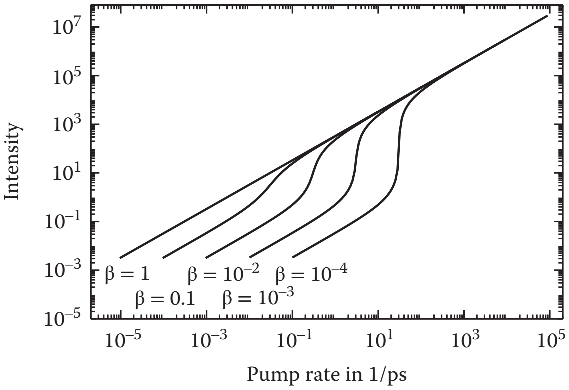


FIGURE 48.3 Laser input–output characteristics as obtained from rate equations. Shown are steady-state results at each pump value, parameterized by the β factor and the cavity loss rate. With increasing β, the threshold shifts to lower pumping and the laser threshold becomes increasingly hard to identify, and additional indicators beyond the reach of the rate-equation description are needed.

to understand and control these properties for new devices. In Section 48.4, we introduce a systematic approach to include semiconductor and correlation effects in extended laser models and discuss effects in nanolasers that set these systems apart from conventional laser devices.

48.3 Key Aspects of Nanolasers

Nanolasers use resonators that confine light in a volume of the light wavelength cube or smaller. The free spectral range, i.e., the energetic distance between resonator modes, grows with decreasing size of the cavity mode volume. Consequently, there is increased potential that the material gain overlaps only with one, or at most, a very few modes [12] of the resonator, as illustrated in the top right panel of Figure 48.4. This sets them apart from other lasers, such as edge-emitting lasers and the typical VCSELs [13,14], where optical modes lie more densely and emission is typically multimode (left panel). A nano-resonator may be a disk or pillar, employing distributed-Bragg reflectors [15], total internal reflection at semiconductor-air interfaces [16], or photonic crystal cavities [17].

It is a great advantage of small resonators that spontaneous emission enhancement $\propto Q/V$ can be used to strongly enhance the light–matter interaction and to favor emission into the single mode and thereby attain high values of the β factor that can approach unity. Here, V is the effective mode volume, and $Q = E/\Delta E$, with $E = \hbar\omega$ the mode energy and ΔE the mode linewidth, is the quality factor of the mode. The spontaneous-emission enhancement is direly needed to make up for the limited amount of gain material that can be placed inside the cavity. This so-called Purcell enhancement [18] can be directly observed in the temporal decay of the emission signal following short excitation. This is shown in Figure 48.5 [19] for micropillar nanolasers of different diameter and, therefore, different effective mode volume V.

As mentioned in Section 48.2, additional indicators are required to identify lasing in high-β devices due to the vanishing kink in the input–output curve. The most established method is to identify the statistical fluctuations of the emitted light. In the thermal regime below threshold, photons prefer to come in bunches. Lasing, on the other hand, originates from a coherent state of the light field, and photons arrive randomly in time. A transition from bunching to close to Poissonian emission is therefore a strong indicator for crossing the threshold to lasing. State-of-the-art streak-camera measurements give a direct visual account of this behavior: Figure 48.6 [20] shows temporally resolved single photon detection events for a system that has been excited by a short laser pulse. Counting statistics can be used to identify single photons, photon pairs, or bunches of more photons as indicated by the boxes. In practice, measurements are repeated thousands

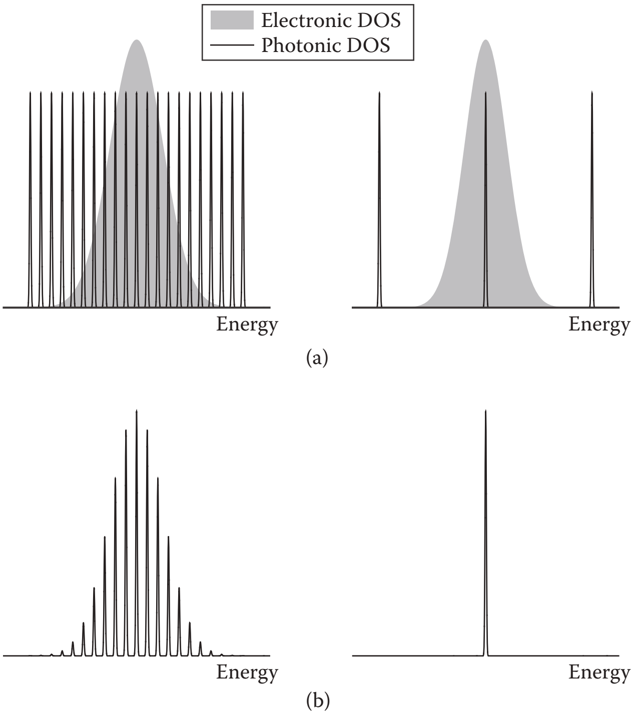


FIGURE 48.4 (a) Illustration of the photonic (black) and electronic (shaded) density of states of a typical laser (left) and a nanolaser (right) with the latter having a much larger free spectral range (energetic spacing between photon modes). The QD gain in the right panel is inhomogeneously broadened due to variations in composition or dimension, see Section 48.8. (b) The resulting spontaneous emission spectra are determined by the overlap of the photonic and electronic density of states. On the other hand, with the proper pillar, disk or photonic crystal design, a nanolaser can be single mode both below and above the lasing threshold. (DOS, density of states.)

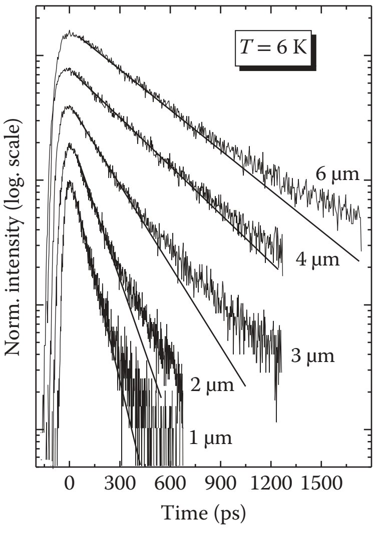


FIGURE 48.5 Time-resolved photoluminescence spectra following a *weak* excitation pulse. A decrease of the spontaneous emission time is observed with decreasing diameter of the micropillar cavities, reflecting the Purcell-enhancement $\propto Q/V$. The weak excitation ensures that the effect is not related to stimulated emission. (Reprinted from Schwab, M. et al., *Physical Review B*, 74, 045323, 2006. With permission.)

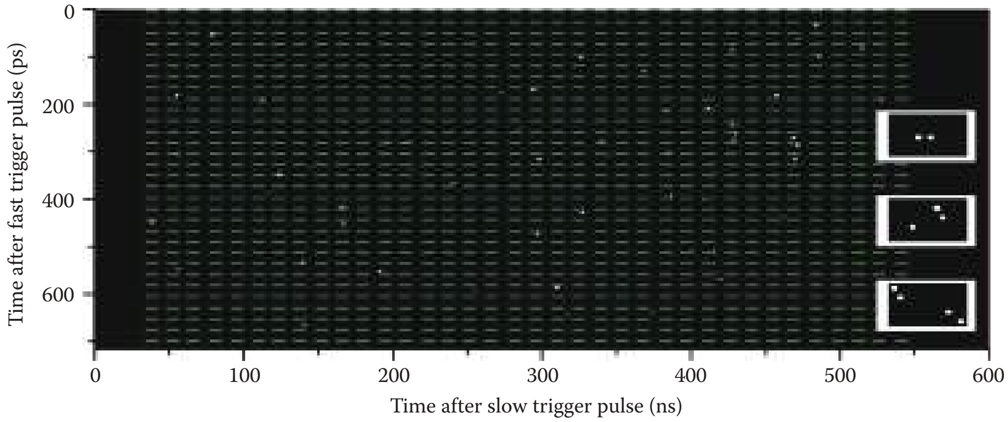


FIGURE 48.6 Streak-camera measurements provide impressive account of statistical fluctuations of the emitted light on the level of single photons. Single- and multiphoton emission events can be detected within a time-bin size of 2 ps (see magnification in the boxes) and provide insight into the quantum-mechanical state of the light field. To perform counting statistics in order to construct correlation functions like the ones shown in Figure 48.7, ten thousands of such measurements are performed. (Reprinted from Aßmann, M. et al., *Science* 325, 297–300, 2009. With permission).

of times. From the collected data, photon correlation functions $g^{(n)}$ can be obtained. They are of particular relevance, as they allow one to characterize the state of the quantum-mechanical light field as given by the distribution function p_n, which describes the *photon statistics* in terms of the probability of finding the system in the photon number state $|n\rangle$. Photon correlation functions are connected to the moments of the photon statistics:

$$\langle n^i \rangle = \sum_{k=0}^{\infty} k^i p_k. \tag{48.4}$$

As an example, the second-order photon-correlation function (sometimes referred to as photon autocorrelation function) is given by

$$g^{(2)} = \frac{\langle n^2 \rangle - \langle n \rangle}{\langle n \rangle^2}. \tag{48.5}$$

Correlation functions of order n take on a value of $n!$ if the light field is in a thermal state, and at all orders they possess a value of 1 for a perfectly coherent state [21]. For a micropillar QD nanolaser, this has been experimentally observed for $g^{(2)}, g^{(3)}$ and $g^{(4)}$ as function of pump rate, shown in Figure 48.7.

More commonly used than a streak camera is a Hanbury Brown and Twiss type measurement [22] schematically shown in Figure 48.8, where the bunching behavior of photons is quantified by the simultaneous registration of two photons at the two detector arms, providing access to the second-order correlation function $g^{(2)}$. For detectors, avalanche photodiodes (APDs) are used. Their time resolution is typically in the range of hundreds of picoseconds. Whether this resolution is sufficient depends on the type of measurement performed and the coherence time τ_{coh} of the emission signal. On the timescale of the latter, $g^{(2)}(\tau)$ goes to unity with respect to the delay time τ between two detection events, and the actual $g^{(2)}(0)$ at $\tau = 0$ can no longer be resolved if the detector's time resolution exceeds the coherence time, leading to the effect that $g^{(2)}(0)$ seems to take on a value of 1 even for thermal light [23].

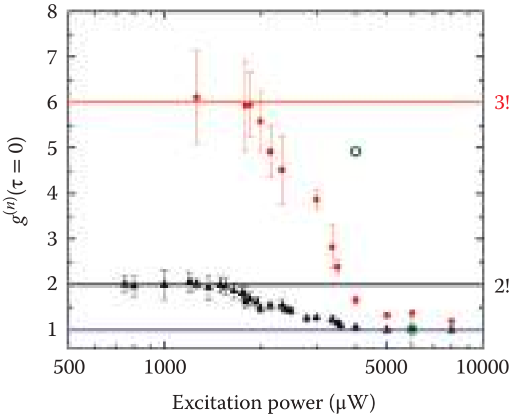


FIGURE 48.7 Photon autocorrelation functions $g^{(2)}(0)$ (triangles), $g^{(3)}(0)$ (dots), and $g^{(4)}(0)$ (open circles) as function of excitation power obtained from streak-camera measurements like the ones shown in the previous figure. For a typical LED to laser transition, all correlation functions show a transition from their thermal values $g^{(n)} = n!$ to that of a coherent field, $g^{(n)} = 1$. (Courtesy of Marc Aßmann, TU Dortmund.)

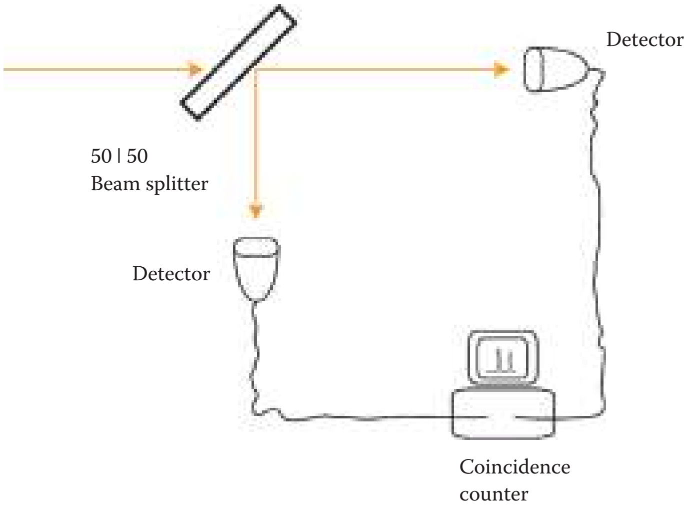


FIGURE 48.8 Illustration of a Hanbury Brown and Twiss measurement. A beam splitter distributes the emission between two detectors. Equal-time detection of two or more photons provides access to the second-order photon autocorrelation function $g^{(2)}(\tau = 0)$. A delay line in one arm can be used to measure $g^{(2)}(\tau \neq 0)$.

Another accessible indicator for coherent emission is a linewidth narrowing, or equivalently, an increase of the coherence time. Experimentally, the linewidth can be determined from spectra or Michelson interferometry measurements [21]. In the steady state $t = t_{ss}$, the coherence time τ_{coh} is related to the first-order coherence of the emission

$$\tau_{coh} = \int_{-\infty}^{\infty} d\tau \, |g^{(1)}(t_{ss}, \tau)|^2, \tag{48.6}$$

which requires the calculation of the two-time quantity in the integral by means of the quantum-regression theorem [24,25]. We will not go into more detail here and refer the reader to [26].

The criteria for lasing in nanolaser devices are a subject of much discussion [5,27]. Ideally, a combination of several criteria is used to identify the lasing regime and the threshold. These include the following:

1. Coherent emission approaching Poissonian photon statistics p_n.
2. A value close to 1 in the second-order photon correlation function $g^{(2)}(0)$. Note that this is a weaker criterion than 1., as $g^{(2)}(0)$ is only linked to the first two moments of the full photon distribution function p_n, c.f. Equation 48.5.
3. An increase in the coherence time at the onset of lasing, when phase-coherent stimulated emission begins to dominate over spontaneous emission.
4. Cavity-photon-induced feedback of the gain material providing gain instead of absorption.
5. A mean intracavity photon number above 1.
6. An S-shaped nonlinear region in the input–output characteristics at threshold.
7. Carrier dynamics developing a hole-burning effect in the carrier-population functions at the energy of the laser transition, cf. Figure 48.13.

All of these criteria have their justifications, but on their own, neither may suffice to conclusively identify lasing. For example, coherent emission (1) is a prerequisite of lasing, but it may be realized without amplification (4 and 5) or originate from an external coherent drive. An S-shaped jump (6) in the input–output characteristic can originate from a variety of effects, such as a transition between multiexciton configurations feeding the laser mode [28], background contributions [29], or spectral wandering. On its own it is not a clear indicator for lasing. Moreover, lasing can take place in the absence of a nonlinear emission regime, e.g., if spontaneous emission losses are small ($\beta \to 1$), or saturation effects set in before the S shape can fully develop, which is particularly relevant under pulsed excitation [30].

48.4 Essential Physics I: Quantum-Optical Semiconductor Laser Theory

Different approaches can be taken in formulating a laser theory, including semiclassical methods, where the radiation field is classical and the system dynamics under the influence of the light–matter interaction is described by Maxwell–Bloch equations [31,32]. In the following, we introduce an extended laser model that accounts for semiconductor-specific effects, such as carrier scattering and dephasing, and that provides access to the statistical properties of the emission. The model is obtained in the framework of a correlation-expansion technique for the quantized light field interacting with carriers occupying the discrete QD conduction- and valence states. The method is versatile, as it allows one to include various aspects of semiconductor and correlation effects on a required level in a systematic and consistent manner. At the same time, the rate equations are contained as limiting case, as has been shown in [33].

In contrast to semiclassical approaches, using a quantized light field introduces the possibility to investigate emission properties beyond the classical regime. In fact, nonclassical effects, such as antibunched light emission around the threshold region, have been demonstrated in QD nanolasers [34]. In semiclassical models, the electromagnetic field drives a coherent polarization $\langle v_\nu^\dagger c_\nu \rangle$ across the bandgap of the semiconductor medium, as illustrated in Figure 48.9. The operators c_ν ($c_\nu^\dagger$) annihilate (create) a conduction-band carrier in the state $|\nu\rangle$, the operators v_ν ($v_\nu^\dagger$) are the equivalent for valence carriers. The resulting dynamics is familiar from atomic two-level models, where a resonant field drives Rabi oscillations. In a fully quantum-mechanical approach, the electromagnetic field is described by modes that are occupied by field quanta, the photons. If the modes are energetically well separated, like in a nanolaser, where the cavity possesses a large free spectral range, it suffices to consider the interaction of the active medium with photons in a single mode q. The coherent polarization $\langle v_\nu^\dagger c_\nu \rangle$ is then replaced by a photon-assisted polarization $\langle b_q^\dagger v_\nu^\dagger c_\nu \rangle$ that describes a carrier deexcitation (or excitation via $\langle b_q c_\nu^\dagger v_\nu \rangle$) by simultaneous emission (absorption)

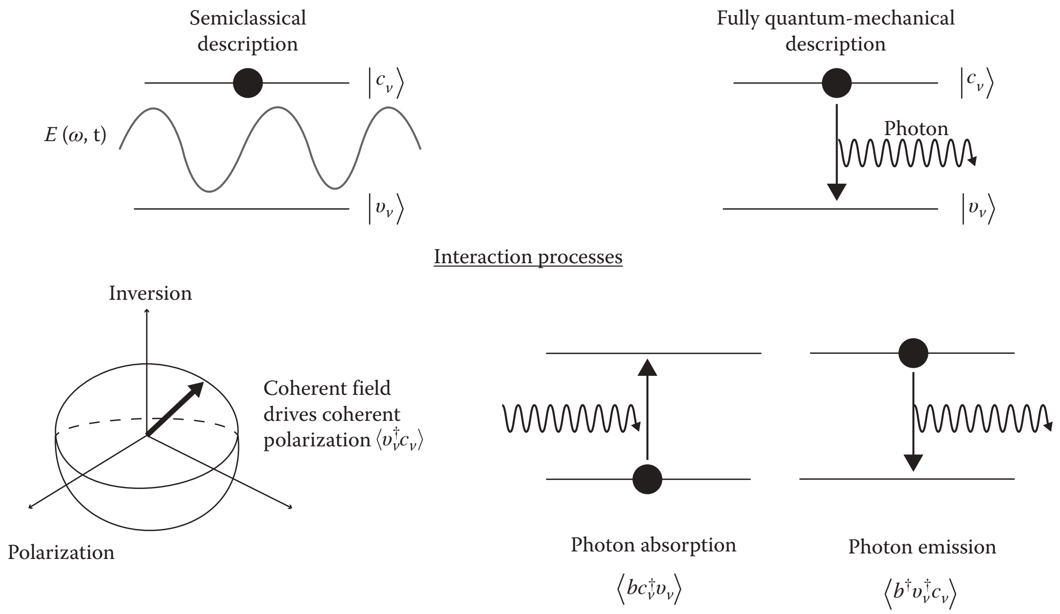

Semiclassical description
Fully quantum-mechanical description
$|c_\nu\rangle$
$E\,(\omega, \mathrm{t})$
$|v_\nu\rangle$
Photon
$|c_\nu\rangle$
$|v_\nu\rangle$
Interaction processes
Inversion
Coherent field drives coherent polarization $\langle v_\nu^\dagger c_\nu\rangle$
Polarization
Photon absorption
$\langle b c_\nu^\dagger v_\nu\rangle$
Photon emission
$\langle b^\dagger v_\nu^\dagger c_\nu\rangle$

FIGURE 48.9 Illustration of the fundamental light–matter interaction in a semiclassical (left) and fully quantum-mechanical description (right). Semiclassically, matter degrees of freedom are described quantum-mechanically (e.g., by the states $|c_\nu\rangle$ and $|v_\nu\rangle$). The interaction with a classical coherent field $E(\omega, t)$ drives a coherent polarization $\langle v_\nu^\dagger c_\nu\rangle$. The polarization and population dynamics of often described in terms of the rotation of a vector on a unit sphere (Bloch sphere), as illustrated in the bottom left. By quantizing the light field, field quanta (photons) are introduced, and the fundamental interaction processes become the electronic (de)excitations via the simultaneous absorption or emission of a photon (bottom right). The semiclassical coherent polarization is replaced by the photon-assisted polarization $\langle b^\dagger v_\nu^\dagger c_\nu\rangle$.

of a photon into the mode q; the Bose operators b_q $(b_q^\dagger)$ annihilate (create) a photon in mode q. The quantum-mechanical and semiclassical pictures are schematically compared in Figure 48.9.

The fully quantum-mechanical laser model is obtained by calculating the dynamical equations for carrier populations $f_\nu^c = \langle c_\nu^\dagger c_\nu\rangle$, $f_\nu^v = \langle v_\nu^\dagger v_\nu\rangle$ in the conduction and valence states of the QD, and the expectation value of the photon number operator in the laser mode q, for which we drop the index q in the following. These equations are obtained from Heisenberg's equations of motion for operators $A(t)$, $\dot{A} = i[H, A]/\hbar$. In the Hamiltonian, we consider the free contributions of QD single-particle states (first and second term in Equation 48.8) and the electromagnetic field (third term in Equation (48.8)), and the light–matter interaction,

$$H = H^{\text{free}} + H^{\text{LM}}, \tag{48.7}$$

with

$$H^{\text{free}} = \sum_\nu \varepsilon_\nu^c\, c_\nu^\dagger c_\nu + \sum_\nu \varepsilon_\nu^v\, v_\nu^\dagger v_\nu \;+ \hbar\omega\left(b^\dagger b + \frac{1}{2}\right) \tag{48.8}$$

and

$$H_{\text{LM}} = -i \sum_{\alpha\nu} \left(g_{\alpha\nu}\, c_\alpha^\dagger v_\nu b + g_{\alpha\nu}\, v_\alpha^\dagger c_\nu b\right) + \text{h.c.} \tag{48.9}$$

Here, $\hbar\omega$ is the mode energy, and $g_{\alpha\nu}$ is the light–matter coupling strength that is determined by the overlap of the wave functions of the QD single-particle states $|\alpha\rangle$ and $|\nu\rangle$ and the mode function of the confined electromagnetic field [33]. The sums run over the single-particle states of the individual QD emitters, and the index is to be understood to contain both, QD label and single-particle state. The interaction Hamiltonian contains the elementary processes of electron–hole recombination and generation via emission and absorption of a photon, including the ones schematically shown in the bottom right of Figure 48.9. Note that the counter-rotating terms in Equation (48.9) are often neglected in the rotating wave approximation.

48.4.1 Dynamical Laser Equations

For the dynamical evolution of the photon number $\langle b^\dagger b\rangle$ in the given mode and the electron and hole populations $f^e_\nu = \langle c^\dagger_\nu c_\nu\rangle, f^h_\nu = 1 - \langle v^\dagger_\nu v_\nu\rangle$, the contribution of the light–matter interaction H_{LM} in the Heisenberg equations of motion leads to [33]

$$\left(\hbar\frac{\mathrm{d}}{\mathrm{d}t}+\kappa\right)\langle b^\dagger b\rangle = 2\,\mathrm{Re}\sum_{\nu'}\left|g_{\nu'}\right|^2\langle b^\dagger v^\dagger_{\nu'} c_{\nu'}\rangle, \tag{48.10}$$

$$\hbar\frac{\mathrm{d}}{\mathrm{d}t}f^{e,h}_\nu = -2\,\mathrm{Re}\left|g_\nu\right|^2\langle b^\dagger v^\dagger_\nu c_\nu\rangle. \tag{48.11}$$

Here, we have scaled $\langle b^\dagger v^\dagger_\nu c_\nu\rangle \rightarrow g_{q\nu}\langle b^\dagger v^\dagger_\nu c_\nu\rangle$ to have the modulus of the coupling matrix elements appear and have further assumed that g is diagonal in the equal-envelope approximation. The finite lifetime of the cavity mode is introduced by considering a complex mode energy, $\tilde\omega = \omega - i\kappa$, where κ is directly connected to the Q-factor of the laser mode, $Q = \hbar\omega/2\kappa$. The dynamics of the photon number is determined by the abovementioned photon-assisted polarization $\langle b^\dagger v^\dagger_\nu c_\nu\rangle$ that describes the expectation value for a correlated event, where a photon in the mode q is created in connection with an interband transition of an electron from a conduction to a valence state. The sum over ν involves all possible interband transitions from various QDs.

An important component of the carrier dynamics is the recombination into nonlasing modes giving rise to the β factor discussed in Section 48.2. For the fraction $1-\beta$ of carriers emitting into nonlasing modes, a term proportional to the spontaneous emission rate can be added to the carrier population equations to account for the losses [33]:

$$\left.\frac{\mathrm{d}}{\mathrm{d}t}f^{e,h}_\nu\right|_{\mathrm{nl}} = \gamma_{\mathrm{sp}}(1-\beta)f^e_\nu f^h_\nu. \tag{48.12}$$

In contrast to the rate equations (Equation 48.3), where this contribution is typically accounted for by the β coefficient, the spontaneous emission losses in a semiconductor are determined by both electron and hole populations. The dynamical equation for the cavity-photon-assisted polarization is given by

$$\left(\hbar\frac{\mathrm{d}}{\mathrm{d}t}+\kappa+\Gamma+i\left(\tilde\varepsilon^e_\nu+\tilde\varepsilon^h_\nu-\hbar\omega\right)\right)\langle b^\dagger v^\dagger_\nu c_\nu\rangle = f^e_\nu f^h_\nu - \left(1-f^e_\nu-f^h_\nu\right)\langle b^\dagger b\rangle + \frac{1}{g_\nu}\sum_\alpha g_\alpha C^{\mathrm{x}}_{\alpha\nu\nu\alpha} + \delta\langle b^\dagger b c^\dagger_\nu c_\nu\rangle - \delta\langle b^\dagger b v^\dagger_\nu v_\nu\rangle. \tag{48.13}$$

The free evolution of $\langle b^\dagger v^\dagger_\nu c_\nu\rangle$ is determined by the detuning of the QD transitions from the cavity laser mode. This oscillatory term drops out if the transition is in perfect resonance with the mode. Only accounting for identical resonant QDs is a common and well-justified practice in ensembles of many (>100 emitters), as the summation over all QDs becomes a mere prefactor of the QD number, which is a significant simplification. In a semiconductor, the source term of spontaneous emission is described by an expectation value of four carrier operators $\langle c^\dagger_\alpha v_\alpha v^\dagger_\nu c_\nu\rangle$, see [35]. The Hartree–Fock factorization of this source term leads to $f^e_\nu f^h_\nu$, which appears as the first term on the right hand side of Equation 48.13. Corrections to this factorization are included in $C^{\mathrm{x}}_{\alpha'\nu\nu'\alpha} = \delta\langle c^\dagger_{\alpha'} v^\dagger_\nu c_{\nu'} v_\alpha \rangle$. Dephasing is what destroys a coherent polarization. In semiconductors, dephasing arises from carrier scattering processes, excitation of carriers, and from the interaction with lattice vibrations that are responsible for the homogeneous linewidth. These contributions are here summarized in a phenomenological constant Γ.

The correlation functions $\delta\langle b^\dagger_q b_q c^\dagger_\nu c_\nu \rangle$ and $\delta\langle b^\dagger_q b_q v^\dagger_\nu v_\nu \rangle$ introduce carrier-photon correlations that are neglected on the level of rate equations. Their time evolution is given by [33]

$$\left(\hbar \frac{\mathrm{d}}{\mathrm{d}t} + 2\kappa\right) \delta\langle b^\dagger b c^\dagger_\nu c_\nu \rangle = -2\,|g_\nu|^2\, \mathrm{Re}\Big[\delta\langle b^\dagger b^\dagger b v^\dagger_\nu c_\nu \rangle + \left(\langle b^\dagger b \rangle + f^e_\nu\right) \langle b^\dagger v^\dagger_\nu c_\nu \rangle\Big], \tag{48.14}$$

$$\left(\hbar \frac{\mathrm{d}}{\mathrm{d}t} + 2\kappa\right) \delta\langle b^\dagger b v^\dagger_\nu v_\nu \rangle = 2\,|g_\nu|^2\, \mathrm{Re}\Big[\delta\langle b^\dagger b^\dagger b v^\dagger_\nu c_\nu \rangle - +\left(\langle b^\dagger b \rangle + f^h_\nu\right) \langle b^\dagger v^\dagger_\nu c_\nu \rangle\Big] \tag{48.15}$$

which couples to

$$\begin{aligned}\left(\hbar \frac{\mathrm{d}}{\mathrm{d}t} + 3\kappa + \Gamma + i\left(\tilde{\varepsilon}^e_\nu + \tilde{\varepsilon}^h_\nu - \hbar\omega\right)\right) \delta\langle b^\dagger b^\dagger b v^\dagger_\nu c_\nu \rangle &= -2\,|g_\nu|^2\, \langle b^\dagger v^\dagger_\nu c_\nu \rangle^2 + 2\big[\langle b^\dagger b \rangle + f^h_\nu - f^e_\nu\big]\times \\ &\big[\delta\langle b^\dagger b c^\dagger_\nu c_\nu \rangle - \delta\langle b^\dagger b v^\dagger_\nu v_\nu \rangle\big] - (1 - f^e_\nu - f^h_\nu)\, \delta\langle b^\dagger b^\dagger b b \rangle\end{aligned} \tag{48.16}$$

and

$$\left(\hbar \frac{\mathrm{d}}{\mathrm{d}t} + 4\kappa\right) \delta\langle b^\dagger b^\dagger b b \rangle = 4\,|g_\nu|^2 \sum_{\nu'} \delta\langle b^\dagger b^\dagger b v^\dagger_{\nu'} c_{\nu'} \rangle. \tag{48.17}$$

On this level of approximation, we obtain a closed system of coupled equations that go beyond the rate equations by containing carrier-photon and photon–photon correlation functions that provide access to the second-order photon-correlation function

$$g^{(2)}(\tau = 0) = 2 + \frac{\delta\langle b^\dagger b^\dagger b b \rangle}{\langle b^\dagger b \rangle^2}. \tag{48.18}$$

48.4.2 Pump Process for Continuous-Wave-Excited Lasers

The earlier equations describe the light–matter interaction of a semiconductor system in terms of carrier populations and the polarization-analogue for the quantized light field, i.e., the photon-assisted polarization. An important component is yet missing, namely the interaction of charge carriers with the environment that allows for some sort of excitation mechanism. Pumping can be either done by current injection or by optical excitation. In both cases, electrons and holes are typically created in the continuum states of the wetting layer (WL) or barrier material, from where they are captured into the discrete QD states. This incoherent capture is one example of carrier scattering processes that, at this stage, are still missing from our description. While the quantitative modeling of these many-body processes is discussed in detail in Section 48.5, for continuous-wave (CW) excitation it is often sufficient to consider only the dynamics of carriers in the localized QD states. In its most simple form, two localized states are used for electrons and holes each, which closely resembles a four-level laser scheme. This choice is physically motivated by the electronic structure of typical QDs, which, depending on their size, consists of several confined states. We refer to the lowest (highest) and second-lowest (second highest) electron (hole) state as *s*- and *p*-states, respectively, as shown in Figure 48.10. In this reduced level scheme, the *p*-states are used for carrier pumping at rate *P*, which approximates the combined effect of carrier excitation in higher

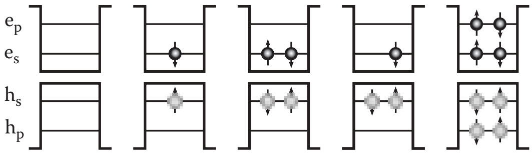


FIGURE 48.10 In a QD with several confined single-particle states for electrons and holes (the s- and p-states are shown), a multitude of many-particle configurations can be realized by occupying these states with electrons and holes. These many-particle configurations contain the neutral, charged, and multiexciton states, and dark states that differ in energy due to the Coulomb interaction between carriers.

states and the successive capture into the QD states:

$$\frac{d}{dt}f_p^e\Big|_{\text{Pump}} = P(1 - f_p^e - f_p^h) \tag{48.19}$$

$$\frac{d}{dt}f_p^h\Big|_{\text{Pump}} = P(1 - f_p^e - f_p^h). \tag{48.20}$$

The blocking factors in the brackets ensure that the occupation remains bounded by unity. From their generation in the QD p-states, a phenomenological set of equations distributes carriers into the QD s-states:

$$\frac{d}{dt}f_s^e\Big|_{\text{scatt}} = \gamma_{p\to s}^{r,e} f_p^e(1 - f_s^e) - \gamma_{s\to p}^{r,e} f_s^e(1 - f_p^e) \tag{48.21}$$

$$\frac{d}{dt}f_s^h\Big|_{\text{scatt}} = \gamma_{p\to s}^{r,h} f_p^h(1 - f_s^h) - \gamma_{s\to p}^{r,h} f_s^h(1 - f_p^h) \tag{48.22}$$

$$\frac{d}{dt}f_p^e\Big|_{\text{scatt}} = -\gamma_{p\to s}^{r,e} f_p^e(1 - f_s^e) + \gamma_{s\to p}^{r,e} f_s^e(1 - f_p^e) \tag{48.23}$$

$$\frac{d}{dt}f_p^h\Big|_{\text{scatt}} = -\gamma_{p\to s}^{r,h} f_p^h(1 - f_s^h) + \gamma_{s\to p}^{r,h} f_s^h(1 - f_p^h). \tag{48.24}$$

Equations 48.21 through 48.24 are motivated in the more general concepts of carrier kinetics in the following section. Together, Equations 48.10–48.24 serve as an accessible and complete model for QD-based lasers, and we discuss exemplary results in Section 48.6.

48.5 Essential Physics II: Carrier Kinetics in QD Systems

As we overviewed in Section 48.2, rate equations can provide important insights for the theoretical understanding of lasers. They can be used to describe input–output curves [36–39], and, combined with Langevin approaches [40,41] even the photon statistics and intensity noise.

An important reason why carrier dynamics is essential for the understanding of laser operation is the excitation mechanism. To see this, let us assume that carrier populations in lasers are given by Fermi–Dirac distribution functions (often a reasonable approximation). This implies that states with below-average kinetic energy are Pauli-blocked for the excitation mechanism. This is illustrated in Figure 48.11, where states with above-average kinetic energy are available for the excitation, while highly occupied states (shaded region, applied to 77K) are not. Therefore, it is important to understand how carriers, injected at the green-shaded area, end up in the spectral window from where they recombine. Accountable for this are two mechanisms that we discuss in this section. The carrier–carrier Coulomb interaction redistributes carriers between different states and leads to an equilibration towards a Fermi–Dirac distribution [42–46]. But as every carrier scattered downward in energy is partnered by a scattering partner that is scattered upward (right sketch in Figure 48.12), the Coulomb scattering is not able to dissipate energy.

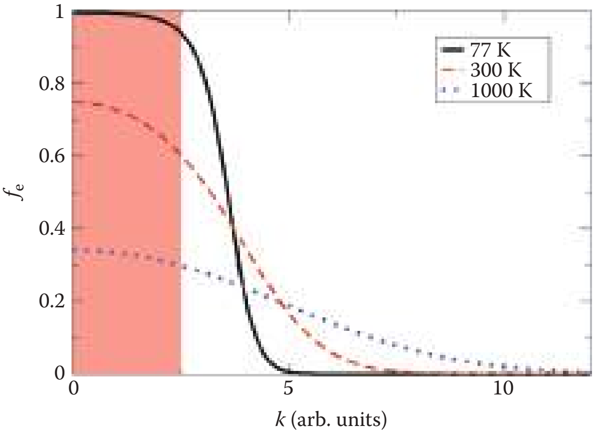


FIGURE 48.11 Fermi–Dirac distribution for electrons in a QW at different temperatures and a carrier density of $5 \times 10^{11}\,\text{cm}^{-2}$, the shaded background indicates the availability for carrier injection for the low-T distribution.

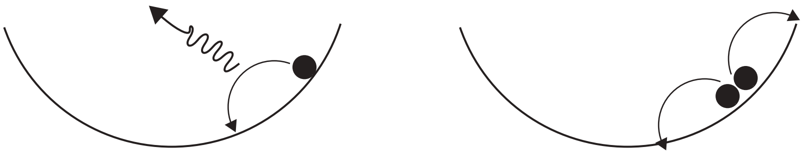


FIGURE 48.12 Conceptual sketch of carrier scattering processes. Left: carrier–phonon interaction, right: carrier–carrier Coulomb interaction.

This leads to a hot Fermi–Dirac function such as depicted in Figure 48.11, that has significantly different properties regarding inversion and gain. The cooling toward the lattice temperature is provided by the carrier-phonon interaction [47–51], that provides for the cooling by taking up the energy of the hot carriers and distributing it toward crystal lattice (left sketch in Figure 48.12). In a laser under CW-excitation conditions the energetic difference from excitation to emission energy also leads to heating and often the carrier-phonon interaction is not strong enough to provide complete cooling towards the lattice temperature. Therefore, a dynamical equilibrium is reached that may consist of Fermi–Dirac distributions with several 100K above the lattice temperature [52,53].

As already outlined in Section 48.4, it is generally necessary to set up equations of motion not for the total carrier density N but for individual carrier populations of different single particle states $f^c_\alpha = \langle c^\dagger_\alpha c_\alpha \rangle$ to describe these effects. Such so-called "kinetic equations" together with the laser equations, Equations 48.10 through 48.20, enable us to treat excitation kinetics and turn-on dynamics of semiconductor QD-based laser systems and will be the topic of this section.

The question of what timescales and by which processes the carriers approach (quasi-) equilibrium conditions (if at all) influences the design of laser devices, as the answer to these questions determines, e.g., the dynamical response and turn-on delay [53–57]. While rate equations can also describe the turn-on dynamics for conventional laser devices, where carrier dynamics takes place on a fs to ps timescale, whereas the laser turn-on happens on the order of 10 ns, the situation is very much different in nanocavities with large Purcell enhancement, i.e., the enhancement of the spontaneous emission rate due to the presence of the cavity, where the turn-on delay can be on the order of 10 ps [53,58] and thus on the same timescale as the carrier scattering. The use of population function both for QD and WL states also allows one to describe effects such as spectral hole burning, or the difference between optical and injection pumping, that are unavailable to a theoretical model that tracks only the dynamics of the total carrier density N.

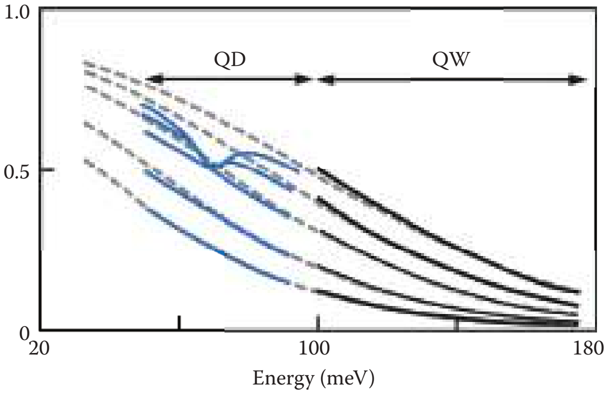


FIGURE 48.13 Carrier populations in a QD+WL system under laser operation for increasing pump intensity. The gray lines describe the population of QD carriers (subject to inhomogeneous broadening), the black lines show the populations of the WL states. For high pump rates the "hole" burnt into the population function by laser operation is clearly visible. The shaded-dashed lines are corresponding Fermi–Dirac functions. (From Chow, W. W. et al., *Light: Science & Applications*, 3, e201, 2014. With permission.)

An example of such carrier population functions are shown in Figure 48.13. The seemingly continuous features of the discrete QD single-particle state populations originate from the large number of emitters in the inhomogeneously broadened ensemble, cf. Section 48.8.

The simplest equation for such state-dependent carrier dynamics is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}f_\alpha = (1-f_\alpha)\Gamma_\alpha^{\mathrm{in}} - f_\alpha\Gamma_\alpha^{\mathrm{out}}, \tag{48.25}$$

where $\Gamma_\alpha^{\mathrm{in}}$ and $\Gamma_\alpha^{\mathrm{out}}$ give the in- (out-) scattering rates into (out of) the state α. A more detailed version of Eq. (48.25) has the structure

$$\frac{\mathrm{d}}{\mathrm{d}t}f_\alpha = \sum_\beta f_\beta(1-f_\alpha)\gamma_{\beta\to\alpha} - f_\alpha(1-f_\beta)\gamma_{\alpha\to\beta} \tag{48.26}$$

and describes that the population of state α is enlarged by scattering from β to α and reduced by scattering from α to β. The Pauli-blocking terms $f_\beta(1-f_\alpha)$ and $f_\alpha(1-f_\beta)$ occur as scattering has to be proportional to the occupancy f of the initial and to the non-occupancy $(1-f)$ of the final state. Under continuous excitation (CW laser), it may suffice to restrict the description to the s-shell and p-shell states of the QD and to use phenomenological scattering rates $\gamma_{\beta\to\alpha}^{e,h}$, in which case we recover Equation 48.21 through 48.24 in Section 48.4.

An often used approximation that is generally of relevance if not only QD states are discussed but also the continuous states of the WL or barrier are taken into consideration is the so-called relaxation time approximation

$$\frac{d}{dt}f_\alpha = \frac{f_\alpha - F_\alpha(N,T)}{\tau}, \tag{48.27}$$

which describes the relaxation of a nonequilibrium carrier population f_α toward a Fermi function $F_\alpha(N,T)$ at temperature T and carrier density N on a time-scale τ. While this approximation often works well, it also necessitates to determine the time-scale τ and the carrier temperature T. The last point is not trivial as in laser devices the carriers typically posses a higher temperature than the lattice [52,53]. To answer these questions and also to investigate situations where the relaxation time approximation (Equation 48.27) is not applicable we have to take the microscopic origin of the scattering process into account. In the following, we discuss the dominant physical mechanisms behind the carrier dynamics, which are carrier-phonon scattering and carrier–carrier Coulomb scattering.

48.5.1 Carrier–Phonon Interaction

As discussed earlier, carrier temperature is an important factor for laser operation as it controls the population inversion. Hence the optical gain depends strongly on the temperature of the carrier system [59,60]. As the carrier–carrier Coulomb interaction cannot dissipate energy, the source for thermalization of the carriers toward the lattice temperature is the interaction with phonons. To investigate this interaction mechanism, we first have to determine which phonon modes are responsible for the carrier dynamics. A main difference between different phonon branches is whether they are longitudinal or transversal in nature. One can show [61] that due to their ability to form polarizations, mainly longitudinal phonons couple to the carrier system and that for most situations the interaction with transversal phonons is negligible. This leaves us with two phonon branches that are most important for the carrier-phonon interaction. These are the longitudinal optical (LO) phonons and longitudinal acoustic (LA) phonons. The speed of propagation of an acoustic phonon, which is also the speed of sound in the lattice, is given by the slope of the acoustic dispersion relation, $\frac{\partial \omega_k}{\partial k}$. At low values of k, the dispersion relation is almost linear, and the speed of sound is approximately independent of the phonon wavenumber.

Even though we are interested in laser structures, where the active material is composed of nanostructures (QDs), it is a good approximation to take the phonon modes and interaction potentials of the surrounding bulk material. This is the case, as we are mostly interested in long-wavelength phonons where the phonon wavelength is much larger than the size of the nanostructure we are investigating.

The interaction of carriers with LO phonons can be described by the Fröhlich coupling [61]. As this coupling mechanism is Coulombic in nature, the corresponding interaction matrix elements drop off rapidly with increasing momentum transfer $\mathbf{q}$. Therefore a often used approximation consists of treating the LO phonons as dispersionless with constant frequency $\hbar\omega_{\mathbf{q}} = \hbar\omega_{\text{LO}}$ [62]. For acoustic phonons, the main interaction mechanisms are the deformation potential coupling and the piezoelectric coupling [62]. While these play an important role in the line broadening of QD lines in the low temperature and low carrier-density regime, they do not provide an efficient source for carrier scattering.

The interaction matrix elements for the interaction of carriers with LO-phonons are [60,63]

$$|M_{\alpha\beta}|^2 = \frac{M_{\text{LO}}^2}{e^2/\varepsilon_0} V_{\alpha\beta\alpha\beta}, \tag{48.28}$$

containing Coulomb matrix elements $V_{\alpha\beta\alpha\beta}$ (see Equation 48.33 and Reference [64]), the elementary charge e and the vacuum permittivity ε_0. The prefactor $M_{\text{LO}} = 4\pi\alpha \frac{\hbar}{\sqrt{2m}} (\hbar\omega_{\text{LO}})^{\frac{3}{2}}$ includes the polar coupling strength α and the reduced mass m. The matrix elements are given here in a way that explicitly shows the Coulombic nature of the carrier-LO-phonon coupling.

A first approach toward carrier scattering is given by the Boltzmann equation [40,61]. For carrier-phonon scattering, this is given by Equation 48.26 with

$$\gamma_{\beta\to\alpha} = \sum_q M_{\beta\alpha}(q)\Big((1+n_q)\delta(\epsilon_\beta - \epsilon_\alpha - \hbar\omega_{\text{LO}}) + n_q\delta(\epsilon_\beta - \epsilon_\alpha + \hbar\omega_{\text{LO}})\Big). \tag{48.29}$$

The terms proportional to n_q and $(1 + n_q)$ describe phonon absorption and emission processes. In QD systems, additional effects exist that are not described within the framework of a Boltzmann equation, but are closely connected to carrier scattering, as they stem from the same microscopic interaction mechanism. These are quasi-particles and non-Markovian effects. The main reason is that the carrier-phonon interaction leads to the formation of a new quasi-particle. The quasi-particle obtained by dressing the carriers with the carrier-phonon interaction—the polaron—describes the lattice distortion accompanying the electron in its motion. In the framework of Green's functions (GFs), polarons are described using the retarded GF Kadanoff–Baym equation [63]:

$$\left[i\hbar\frac{\partial}{\partial\tau} - e^{a}_{\alpha}\right] G^{\mathrm{a,R}}_{\alpha}(\tau) = \delta(\tau) + \int d\tau' \; \Sigma^{\mathrm{a,R}}_{\alpha}(\tau - \tau')\, G^{\mathrm{a,R}}_{\alpha}(\tau'). \tag{48.30}$$

The corresponding retarded selfenergy in random-phase approximation (RPA) [62] is given by

$$\Sigma^{\mathrm{a,R}}_{\alpha}(\tau) \;=\; i\hbar \sum_{\beta} \; |M_{\alpha\beta}|^2 G^{\mathrm{a,R}}_{\beta}(\tau)\, d^{<}(-\tau), \tag{48.31}$$

where the phonon propagators $d^{\gtrless}$ contain the phonon frequency and the phonon population [63]. The spectral function $\hat{G}_{\alpha}(\hbar\omega) = 2\mathrm{Im}(G^{\mathrm{a,R}}_{\alpha}(\hbar\omega))$, which follows from the retarded GF, can be seen as a generalization of the single particle energy, i.e., a free particle has a spectral function of the form $\hat{G}_{\alpha}(\hbar\omega) = \delta(\epsilon_{\alpha} - \hbar\omega)$. In contrast, the Polaronic spectral functions but also posses a finite spectral width that is connected to the lifetime of the respective quasi-particle.

Using the quasi-particles as described by the polaron retarded GF, we can formulate a quantum-kinetic equation for polaron scattering:

$$\begin{aligned}\frac{\partial f^{a}_{\alpha}(t)}{\partial t} &= 2\,\mathrm{Re} \sum_{\beta} \int_{-\infty}^{t} dt' \; |M_{\alpha\beta}|^2 G^{a,\mathrm{R}}_{\beta}(t,t') \left[G^{a,\mathrm{R}}_{\alpha}(t,t')\right]^{*} \\ &\quad * \left\{ \left[f^{a}_{\beta}(t')(1 - f^{a}_{\alpha}(t'))\right] d^{>}(t',t) - \left[f^{a}_{\alpha}(t')(1 - f^{a}_{\beta}(t'))\right] d^{<}(t',t) \right\},\end{aligned} \tag{48.32}$$

which contains scattering by phonon emission and absorption processes just as in the Boltzmann equation. Effects included here that go beyond the Boltzmann treatment are a weakening of the exact energy conservation due to the polaron properties (finite spectral width) and due to non-Markovian effects described by the explicit dependence of the time change of f^{a}_{α} at time t on all earlier times.

Results from such a quantum-kinetic treatment are shown in Figure 48.14. The QD level spacing is chosen to be 10% higher than the LO-phonon energy of 36meV. Therefore, a Boltzmann scattering equation predicts a complete inhibition of scattering due to the δ-functions in Equation (48.29). In contrast, the quantum-kinetic treatment shows that, due to quasi-particle effects (dotted line), scattering channels open up. The solid line shows the combined influence of quasi-particles and non-Markovian effects, that are caused by the time integrals in Equation (48.32). The inhibition of scattering as given by a Boltzmann equation, was predicted as a "phonon-bottleneck" early on [65,66], however, fast-carrier scattering was observed e.g., in Reference [67], showing the relevance of polaronic effects. For details see References [63,68,69].

48.5.2 Coulomb Interaction

The most important carrier scattering mechanism in the regime of high excitation densities, relevant for laser operation, is the carrier–carrier Coulomb interaction. It leads to a fast redistribution of carriers from

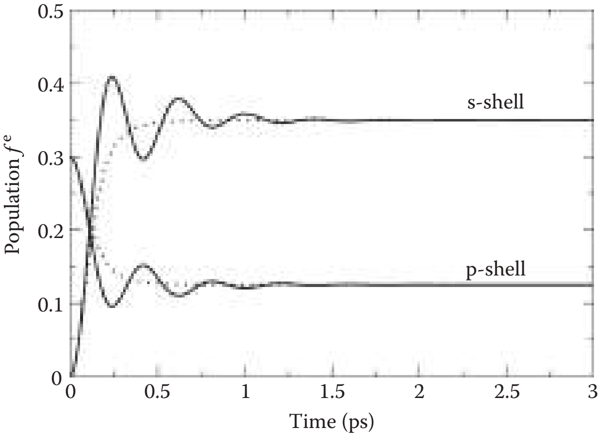


FIGURE 48.14 Redistribution of carrier occupations due to scattering of electrons from p-shell to s-shell in a QD via the carrier phonon interaction. In a Boltzmann-equation approach the occupations would remain constant due to a complete lack of scattering processes as described in the text (Reprinted from Seebeck, J. et al., *Physical Review B*, 71, 125327, 2005. With permission).

the spectral window where excitation takes place toward the spectral window from where they recombine. While Coulomb scattering cannot dissipate energy and thermalize the system (see Figure 48.12), it leads to equilibration toward a hot Fermi–Dirac function.

To determine the strength of the Coulomb interaction in nanostructures, we need to determine matrix elements of the Coulomb interaction in the single-particle basis given by, e.g., tight-binding descriptions of QDs. Using such single-particle wavefunctions, we can construct the Coulomb matrix elements

$$V_{\alpha\beta\gamma\delta} = \frac{1}{A}\sum_{\mathbf{q}} V_{\mathbf{q}} \left\langle \alpha \middle| e^{-i\mathbf{q}\cdot\mathbf{r}} \middle| \delta \right\rangle \left\langle \beta \middle| e^{+i\mathbf{q}\cdot\mathbf{r}} \middle| \gamma \right\rangle, \tag{48.33}$$

consisting of overlap integrals between single-particle wavefunctions and the Coulomb potential $V_{\mathbf{q}}$ [64]. A closer analysis of these Matrix elements also points towards a source for the efficiency of Coulomb scattering in nanostructures. Even though the available phase space of scattering partners is reduced significantly in QDs, compared to QW or bulk systems, this is partially balanced by enhanced interaction matrix elements due to the strong localization of the QD wave functions.

Like for the carrier-phonon scattering a first description of carrier–carrier scattering can be given by a Boltzmann-type equation

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}f_{\alpha}(t) = &\frac{2\pi}{\hbar}\sum_{\beta\gamma\delta}\left(\left|W_{\alpha\gamma\delta\beta}\right|^2 - W_{\alpha\gamma\delta\beta}W^*_{\alpha\gamma\beta\delta}\right)\delta(\epsilon_\alpha - \epsilon_\beta + \epsilon_\gamma - \epsilon_\delta)\\ &\Big[f_\beta(t)(1-f_\alpha(t))f_\delta(t)(1-f_\gamma(t)) - f_\alpha(t)(1-f_\beta(t))f_\gamma(t)(1-f_\delta(t))\Big],\end{aligned} \tag{48.34}$$

which contains screened interaction matrix elements $W_{\alpha\gamma\delta\beta}$ and an energy conserving δ-function. A feature of the carrier–carrier Coulomb scattering in QDs is that, while most scattering channels become extremely inefficient at low carrier density due to the population factors in Equation (48.34), some some scattering channels like electron–hole scattering are possible at very low carrier densities [70].

Results for the carrier–carrier Coulomb scattering in a QD laser under optical cw excitation are presented in Figure 48.15, where the carrier population functions is shown as a function of energy for different times. Discrete symbols denote the QD populations for s-shell and p-shell and the lines the continuous

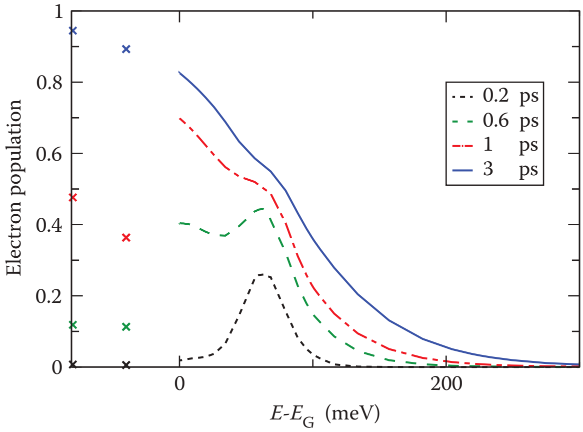


FIGURE 48.15 Carrier populations as a function of energy for electrons in a QD system under laser operation for different times. Symbols denote QD populations, lines WL populations. (Reprinted from Lorke, M. *Applied Physics Letters*, 99, 151110, 2011. With permission.)

WL states that are optically excited. After the initial Gaussian population profile has been excited by the pump, an ultra-fast redistribution is observed, that immediately starts to populate the QD states. It should be noted that, contrary to an often assumed situation [56,71], the timescales between the relaxation dynamics in the WL and the dynamics of the QD populations are not decoupled. On one hand, this is due to the efficiency of the initial in-scattering into the QD, as the QD states are initially empty. On the other hand, the relaxation within the WL is significantly slower than in a pure QW, as the capture into the QD states is most efficient for the WL states with low quasi-momenta. Therefore, these states are constantly depleted during the early stage of the kinetics, which slows down the relaxation of the WL distribution itself toward a quasi-equilibrium distribution [53] Furthermore, after the initial relaxation that lasts about 3 ps, a heating of the carrier population is observed on a timescale of 100 ps, as the pumping into the WL injects carriers at a higher energy than the emission energy.

These results also show (compared to those of the preceding section) that Coulomb and phonon scattering act on the same timescale in QD systems.

As for the phonons, quantum-kinetic effects lead to quasi-particle formation and non-Markovian effects. The corresponding quantum-kinetic equations can be written as

$$\frac{\partial f^{a}_{\alpha}(t)}{\partial t} = \frac{2}{\hbar}\mathrm{Re}\sum_{\beta\gamma\delta}\int_{-\infty}^{t} \mathrm{d}t'\left(\left|W_{\alpha\gamma\delta\beta}\right|^{2} - W_{\alpha\gamma\delta\beta}W^{*}_{\alpha\gamma\beta\delta}\right)G^{a,\mathrm{R}}_{\beta}(t,t')\left[G^{a,\mathrm{R}}_{\alpha}(t,t')\right]^{*} * G^{a,\mathrm{R}}_{\delta}(t,t')\left[G^{a,\mathrm{R}}_{\gamma}(t,t')\right]^{*}$$

$$* \left[f_{\beta}(t')(1-f_{\alpha}(t'))f_{\delta}(t')(1-f_{\gamma}(t')) - f_{\alpha}(t')(1-f_{\beta}(t'))f_{\gamma}(t')(1-f_{\delta}(t'))\right]. \qquad (48.35)$$

As for the Boltzmann-type equation for carrier scattering, this equation contains direct and exchange scattering, proportional to $\left|W_{\alpha\gamma\delta\beta}\right|^{2}$ and $W_{\alpha\gamma\delta\beta}W^{*}_{\alpha\gamma\beta\delta}$, respectively. The energy conserving δ-function is generalized by the product of retarded GFs just as for the carrier-phonon interaction and via the explicit time dependence on earlier times memory effects are included. While these equations have not been evaluated in detail due to their numerical complexity, it is expected that the included quasi-particle and non-Markovian effects would influence several properties of nanolasers. On one hand, large signal modulation and turn-on delay are expected to be influenced by non-Markovian effects. On the other hand, quasi-particle effects can, as for the carrier-phonon interaction, alter the scattering rates influencing also the modulation response and threshold current. In the photon-assisted polarization, Equation 48.13 exist analogous expressions to the scattering contributions given in Equation 48.35 that lead to excitation-enhanced dephasing

$$\frac{\mathrm{d}}{\mathrm{d}t}\langle b^{\dagger} v_{\alpha}^{\dagger} c_{\alpha}\rangle(t) = \int_{-\infty}^{t} \mathrm{d}t'(-\Gamma_{\alpha}^{\mathrm{DD}}(t,t'))\langle b^{\dagger} v_{\alpha}^{\dagger} c_{\alpha}\rangle(t') + \sum_{\beta} \Gamma_{\alpha\beta}^{\mathrm{OD}}(t,t')\langle b^{\dagger} v_{\beta}^{\dagger} c_{\beta}\rangle(t'), \qquad (48.36)$$

where Γ^{DD} and Γ^{OD} consist of interaction matrix elements W, retarded GFs $G^{a,\mathrm{R}}$ and population functions f_{α} in a similar way as they appear in Equation 48.35 for the carrier scattering. We have investigated the influence of quasi-particle and non-Markovian effects on the quasi-classical analogues of these quantities in detail to study optical gain spectra of QD systems [8,60,72–75] and we expect an influence of these contributions on the turn-on behavior and the modulation dynamics [38].

48.6 Connecting to Experiments: Transition to Lasing and the β Factor

As first example, we discuss characteristic input–output curves and statistical properties of a Purcell-enhanced QD nanolaser device. In Figure 48.16 results from the semiconductor laser model are shown for various values of the β factor and fixed number of QD emitters. For β = 1, all spontaneous emission is

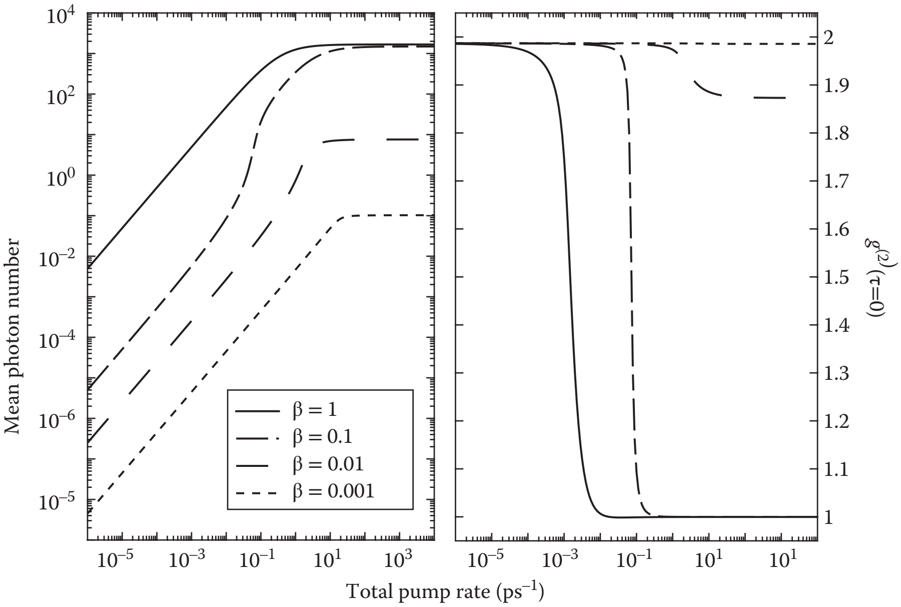


FIGURE 48.16 Input–output curves (left) and autocorrelation function $g^{(2)}(0)$ (right) for a nanolaser with 300 QD emitters. Curves are shown for different values of the β factor. For β = 1, a visible threshold is absent in the input–output curve, whereas $g^{(2)}(0)$ shows a clear transition from thermal to coherent emission. Note that the jump from below to above threshold is larger than in the results obtained from rate equations (Figure 48.3) and does here no longer scale with 1/β. For β = 0.1, a soft kink appears in the input–output curve. For smaller values of β, the limited number of QD emitters cannot provide sufficient gain to reach lasing before saturation is reached. A cavity-Q factor of 32,000 and a cavity-enhanced spontaneous emission rate of $\tau_{\mathrm{sp}} = 50$ ps has been used.

directed into the laser mode. In contrast to the rate-equation result that we have discussed in the context of Figure 48.3, saturation is visible at strong pumping. This effect is important in QD nanolasers, since the small size of the resonator naturally limits the amount of gain material that can be brought into the cavity. A monotonous increase of the output intensity with pumping, as given by the rate equations (Equation 48.1), would be unphysical. As a consequence, for smaller values of β, the gain can become insufficient to reach lasing before saturation sets in. In the example in Figure 48.16, this is visible for $\beta = 0.01$, where the threshold is reached, but the emission remains largely thermal with $g^{(2)}(0) \approx 1.9$, and for $\beta = 0.001$, where the QD emission saturates long before the laser threshold at a low mean intracavity photon number of 0.1. Access to the photon autocorrelation function $g^{(2)}(0)$ provides an important tool in judging the emission properties far beyond what can be inferred from the emission intensity alone and is a significant advancement over the rate equations.

Deeper insight into cavity-QED effects in nanolasers can be obtained from the emission and correlation *dynamics*. The top panel in Figure 48.17 shows time resolved the emission pulse (solid curves) and autocorrelation function $g^{(2)}(t, \tau = 0)$ (symbols) following picosecond-short optical excitation of the nanolaser. Also here the high time resolution has been realized by using a streak-camera setup. Following a sufficiently strong excitation pulse, the statistical properties of the emission pulse change as function of time from thermal to coherent and back to thermal. Coherent emission is reached around the peak region of the emission pulse and forms a plateau when the excitation level is above the threshold region. The data agrees well with time-dependent results of the laser model (bottom panel), which are readily available from the time evolution of the coupled differential equations presented in Section 48.4. In contrast to conventional measurements that integrate over the emission pulse, here it is possible to track the degree of coherence during the pulse and to provide an understanding of the coherence it contains. In their combination, state-of-the art measurements and microscopic models can reveal new aspects of solid-state light sources, reflecting the emission dynamics and the dynamics of quantum-mechanical correlations down to the few-photon level. The explicit treatment of dynamical properties of nanolasers, such as carrier dynamics, turn-on delay, and modulation response requires to go beyond the simple approach in Equations 48.21 through 48.24 for the carrier scattering and use the more involved methods of Section 48.5 due to the sensitivity of these quantities [55].

48.7 Interemitter Coupling Effects and Superradiance in Nanolasers

When several emitters couple to a common light field, the interaction introduces correlations between them. The eigenstates of the interacting system are then no longer given by those of the individual systems, but by so-called Dicke states of the collective system. In 1954, Dicke has shown [76] that from certain collective states, emission can be enhanced or suppressed, which is referred to as superradiance and subradiance. In particular, for the Dicke state of the half inverted system of N two-level emitters, the recombination rate is proportional to $N(N - 1)$ [77]. This quadratic behavior with emitter number has become one of the hallmarks of superradiance. Since its discovery by Dicke, superradiance has been extensively and continuously studied in a variety of systems. Most commonly, however, superradiance is associated with temporal modifications of the emission in spatially extended systems, such as clouds of atoms [78] or, when it comes to semiconductor systems, excitons in extended systems [79].

In an ensemble of QDs that are embedded in a microcavity, depending on the inhomogeneous broadening of the QDs and the cavity Q factor, a fraction of the QDs overlaps both spectrally and spatially with the mode. The common light field these resonant solid-state emitters are subjected to provides a mechanism to form inter-emitter correlations that can lead to sub- and superradiant effects in QD nanolasers. Laser models typically assume that emitters act individually and neglect such effects. In the following, we outline how to use the formalism introduced in Section 48.4 to formulate a laser theory that includes inter-emitter coupling in an approximative way and give a quantitative discussion of its effects. As it turns out, their

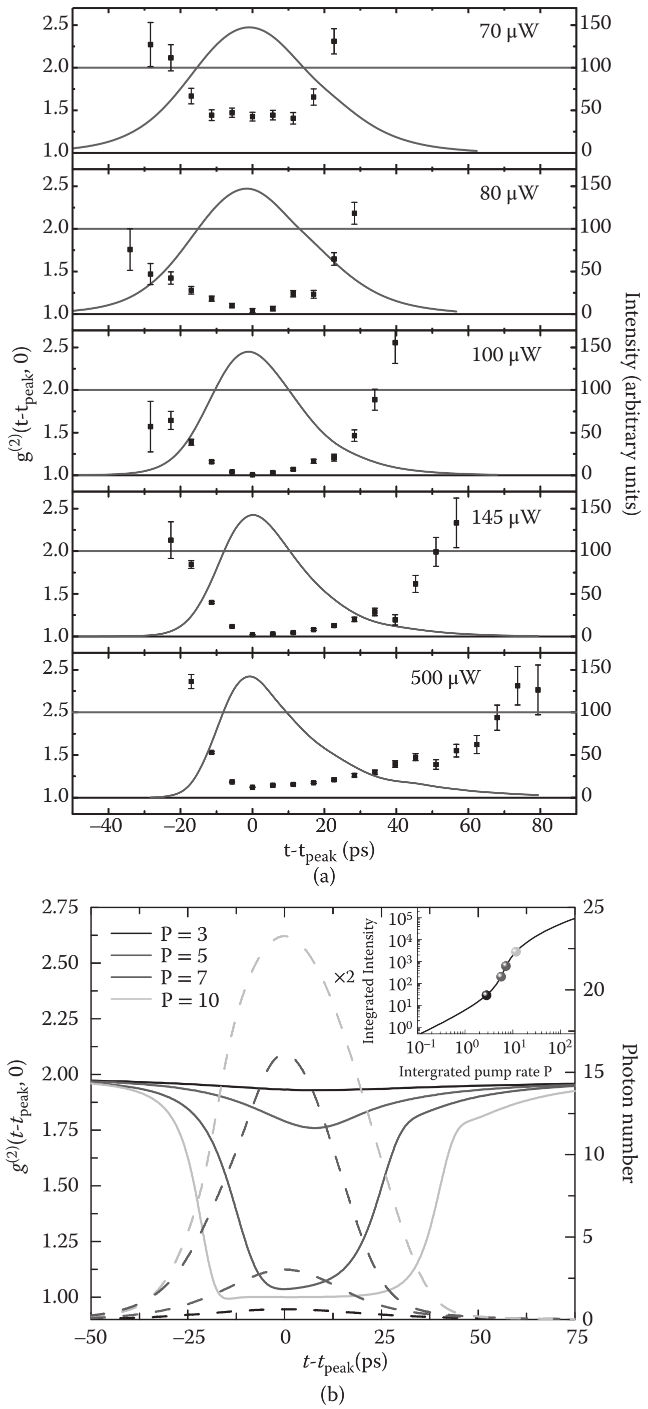


FIGURE 48.17 Time-resolved measurements (a) and theoretical results (b) for emission intensity and photon autocorrelation function following picos-second short optical excitation of a QD-micropillar laser. (Reprinted from Aßmann, M. et al., *Physical Review B* 81, 165314, 2010. With permission).

impact on both stationary and dynamical laser properties can be quite significant. A detailed description of the formalism is found in [80].

To this end, we introduce a formulation for operators that act on the electronic QD states in terms of the many-particle configurations $|i\rangle$, where i refers to configurations like ground state, exciton, trion, biexciton, various dark states, etc. These configuration states, some of which are shown in Figure 48.10, form the basis of the single-QD subspaces of the Hilbert space. Operators Q^{α}_{ij} are introduced that describe either transitions from configuration j to configuration i for $i \neq j$, or the probability that configuration i is realized for $i = j$ in QD α. This approach has advantages over descriptions that use single-particle operators c_{ν} and v_{ν}, like we did in Section 48.4. In doing so, many-electron configurations of one emitter are addressed by a single operator Q rather than many single-particle creation and annihilation operators. It further simplifies the systematic approximation scheme used to truncate the hierarchy of equations of motion with respect to correlations involving *different* QDs. Interemitter correlations are described by a hierarchy of expectation values $\langle Q^{\alpha}_{ij}Q^{\beta}_{kl}\rangle$, $\langle Q^{\alpha}_{ij}Q^{\beta}_{kl}Q^{\gamma}_{mn}\rangle, \ldots$ that involve transitions in QDs α and β, or QDs α, β and γ, etc. The cluster expansion approach facilitates a truncation at a certain order of interemitter coupling, e.g. on the level of pair-correlations between emitters given by $\langle Q^{\alpha}_{ij}Q^{\beta}_{kl}\rangle$ and illustrated in Figure 48.18. All quantum-mechanical expectation values containing electronic QD operators acting on more than two emitters are then neglected. On this level, which is the lowest order that contains interemitter coupling effects, a system of N identical QDs with four confined states each is described by about 300 coupled differential equations, also including photon correlations up to the fourth order providing access to the autocorrelation function $g^{(2)}(0)$. The generation of the equations of motion can be assisted by using computer algebra [80]. While the formalism may appear to involve a great deal of effort, it is highly efficient in the numerical evaluation, and computations for many emitters can typically be performed on a single workstation.

The impact of interemitter coupling on the steady-state properties of a continuously driven QD nanolaser can be seen in Figure 48.19. In the top panel, input–output curves are shown that have been obtained from two separate calculations, one including and one omitting QD–QD correlations. C_F in the bottom panel visualizes the difference in the output intensity I, i.e.,

$$C_F = \frac{I_{\text{rad. coupled QDs}}}{I_{\text{independent QDs}}} - 1. \tag{48.37}$$

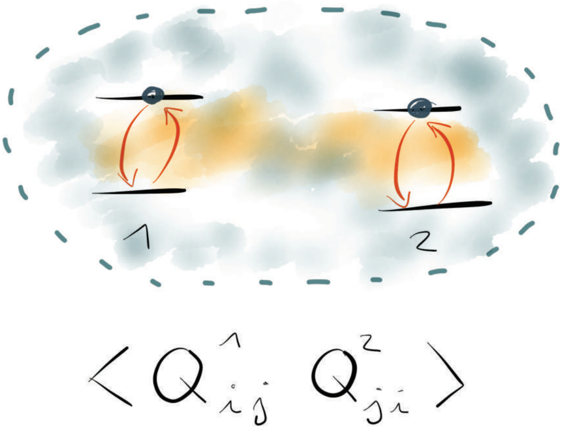


FIGURE 48.18 Illustration of photon-mediated inter-emitter dipole correlations between two emitters.

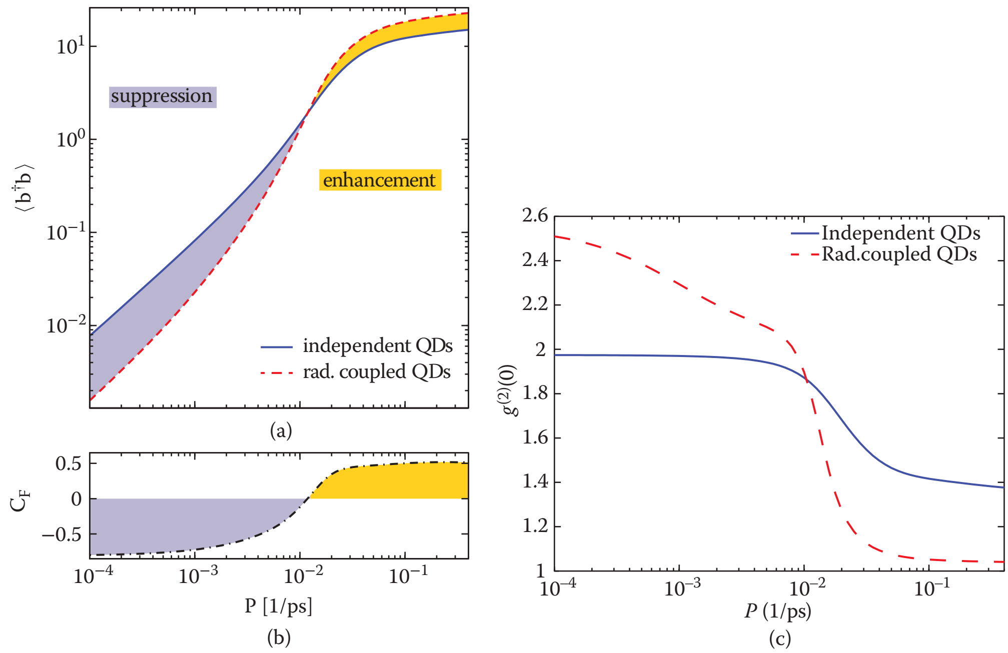


FIGURE 48.19 (a) Input–output curve for a QD nanolaser with typical specifications: 100 resonant emitters, cavity loss rate $\kappa = 0.05/\text{ps}$ (corresponding to $Q = 20,000$), carrier relaxation rate $\gamma_r = 0.05/\text{ps}$, and spontaneous losses into leaky modes $\gamma_{\text{spont}} = 0.01/\text{ps}$. (b) The cooperativity factor visualizes the difference caused by the radiative coupling. (Adapted from Leymann, H. A. M. et al., *Physical Review Applied*, 4, 044018, 2015. With permission.) (c) Corresponding autocorrelation functions. At low excitation, superthermal bunching is a signature of the subradiant emission regime. At high excitation, the threshold to lasing is not fully crossed if emitters act individually and radiative coupling is neglected. Dipole–dipole correlations between distant emitters create additional coherence in the superradiant regime.

The possibility to "switch" the coupling on an off is a particular advantage of the equation-of-motion method and allows to directly assess the impact of the radiative coupling. The following effects are observed:

1. In the low-excitation regime, the inter–emitter coupling creates a subradiant state with a reduced spontaneous emission rate and reduced photon output. Once the laser threshold is crossed, a superradiant state is formed which enhances the laser output *in addition* to stimulated emission provided by photons in the cavity. In sum, the effect on the input–output curve over the whole excitation regime is quite significant. In particular, in the given example, which has been calculated for a typical microcavity laser with 100 QD emitters as active medium, one would underestimate the true β factor by one order of magnitude when ignoring the radiative coupling
2. The input/output characteristics exhibit a different and untypical slope below threshold.
3. In the subradiant regime (low emission intensity), dipole correlations between pairs of QDs cause an increased probability to emit photons synchronously, leading to super-thermal bunching behavior in the statistical properties of the emission, as shown in the right panel of Figure 48.19.
4. Due to inter-emitter coupling, fewer QDs are required to reach the lasing threshold. In small systems, where achieving sufficient gain due to the limited number of emitters in the cavity has always been an issue, the presence of radiative coupling can explain why lasing is more easily reached than expected from conventional laser theories that do not include QD–QD correlations.

5. Coherent emission with $g^{(2)}(0) = 1$ is achieved at lower intracavity mean photon number in the presence of interemitter coupling. This may be interpreted as an increase in the "coherence per photon" due to the alignment of dipole correlations between emitters.

Interestingly, these theoretical predictions are hard to verify experimentally. A strong indicator for the presence of a dipole-correlated phase is the super-thermal photon bunching [81], which has recently been observed in time-resolved studies using a streak-camera setup in combination with microscopic theory [82]. In the same study, highly accelerated spontaneous emission and subradiant photon trapping have been demonstrated. In their sum, the observation of these criteria give a convincing account of the presence of strong inter-emitter coupling effects in a typical QD nanolaser.

48.8 Nonresonant Coupling and Lasing from Multiexciton States in Few-QD Systems

An inherent property of self-assembled QDs that are grown in the so-called *Stranski–Krastanov* mode [83] is inhomogeneous broadening. It refers to inhomogeneity in size, shape and material composition from one QD to the next, causing differences in the confinement potential and with it the electronic single-particle states. While the typical linewidth of QD transitions is tens to hundreds of μeV depending on temperature [2,84,85], the line of an ensemble of self-assembled QDs consists of the spectral lines of the individual emitters, which for many emitters then appears as a single line with a typical broadening of 20–50 meV. An illustrative demonstration of this behavior is given in Figure 48.20 [86]. Inhomogeneous broadening is generally seen as a weak point of self-assembled QDs, as it takes away some of the advantage of the narrow and well-defined transition energies that is characteristic for the single emitter. On the other hand, in cavity-QED it simplifies the task of creating spectral overlap between emitters and a cavity mode (cf. Figure 48.4) and compensates for spectral wandering, since the ensemble broadening is large compared to the cavity linewidth (typically 50–500 μeV). The influence of detuning of emitters in the ensemble and a cavity mode is greatly reduced by nonresonant coupling mechanisms discussed later. The influence of inhomogeneous broadening on QD laser properties has been studied in the framework of a QD laser theory in [27].

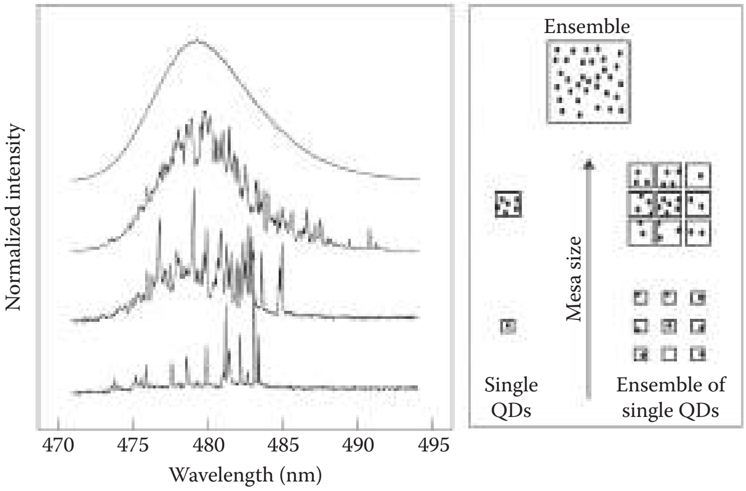
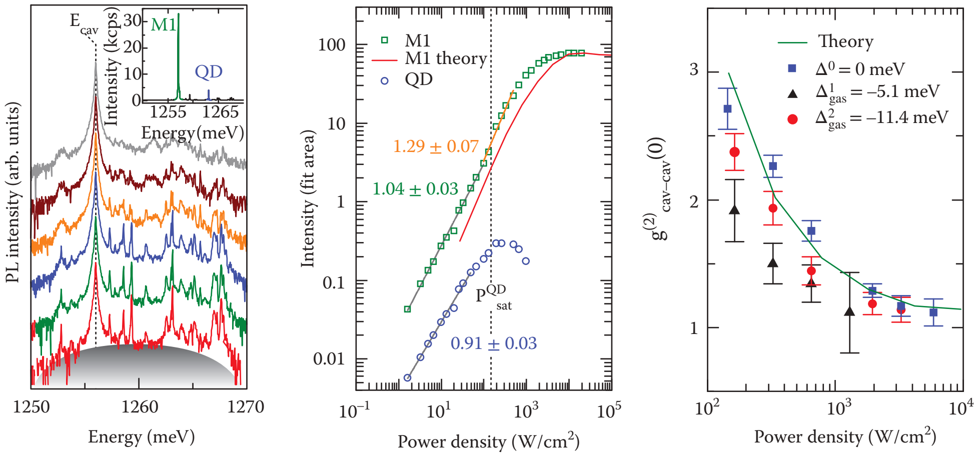


FIGURE 48.20 Experimental results illustrating inhomogeneous broadening in a quantum dot (QD) ensemble. With decreasing mesa size, the QD ensemble consists of fewer and fewer emitters (illustrated right). The broad inhomogenous emission spectrum for many emitters (top left) is revealed to consist of discrete lines as the number of emitters is reduced (bottom left). (Reprinted from Scheibner, M. et al., *Nature Physics*, 3, 106–110, 2007. With permission.)

In this section, we look into what happens if the resonator contains only a few emitters, and the gain spectrum consists of a multitude of separate sharp lines. Such a situation is depicted in the left panel of Figure 48.21 [87]. Then, the emission properties are determined by the interplay of various many-particle configurations of each QD emitter, and their contribution to the photon production depends on the relative spectral position to the cavity mode. While at first it seems contradictory that lasing can be achieved in such a system, strong emission enhancement can drive even a single emitter close to the regime of lasing. At the same time, nonresonant coupling mechanisms haven been identified to provide means for a detuned QD to emit photons into the cavity mode. Nonresonant cavity feeding plays a significant role in nanolasers with ensembles of emitters, but a quantification of its effect is nearly impossible to obtain in such a system. The few-emitter limit, however, offers the unique possibility to study non-resonant mode coupling in a highly controllable environment. In fact, the underlying physical process has long been elusive. Intensive research on this topic has identified three mechanisms that are responsible for the effect:

1. Accoustic phonons can bridge small energy gaps in the range of 1 meV between a detuned emitter resonance and the cavity mode [88–93]. The efficiency is determined by the imbalance between phonon-assisted cavity-photon emission, and the reverse process. Low temperature favors this asymmetry, because low phonon population makes the emission of a phonon more likely than its absorption. Therefore, phonon-mediated nonresonant coupling is more efficient for coupling blue-detuned emitter transitions to the mode [94].
2. At higher carrier densities, interactions with the quasi-continuous WL states provide means for Auger-like scattering processes that allow to bridge larger energy gaps of up to 10 meV [95–99]. While their efficiency is small in comparison to the ultrafast intraband relaxation [53,64,100–102], Auger-coupling is of central importance in the high-excitation regime of nanolasers.
3. QDs possess a rich electronic structure that gives rise to a multitude of multiexciton states, as illustrated in Figure 48.10. The Coulomb interaction energetically separates transitions from different multi-exciton states. This energetic separation, and the possibility of interband recombinations from higher confined states (e.g., the p shell) covers an even larger energetic window, as has been shown



(a) (b) (c)

FIGURE 48.21 **(a)** Spectra of a few-quantum dot (QD) photonic-crystal cavity laser for excitation energies ranging from 0.14 to 5.9 kW/cm^2. **(b)** Input–output curves at the cavity-mode energy (squares) and the QD exciton (circles), the latter showing saturation while the cavity-mode emission continues to increase. **(c)** Second-order photon correlation function at the cavity mode for three different detunings of the cavity-mode energy. Irrespective of the detuning between QD transitions and mode, lasing is achieved at sufficiently high excitation. PL, photoluminescence. (Reprinted from Lichtmannecker, S. et al., *Scienti. Rep.*, 7, 2017).

in [99,103,104]. Thus, even if the QD exciton line, which dominates the spectrum at low excitation, is detuned from the cavity mode, a higher transition, e.g., from the biexciton, may be in perfect resonance with the mode and lead to efficient photon production at higher excitation.

There are several approaches to include the non-resonant coupling of detuned emitters in a laser theory, which superficially all have the same effect of producing additional photons in the resonator mode. In their origin they differ fundamentally, which is reflected in the saturation behavior and in the statistical properties of the emission. On the one hand, a "direct" cavity feeding is often used in the literature [105–108], where photons are directly generated in the cavity mode by an inverse cavity loss term. Such a feeding mechanism, acting together with the cavity loss, gives rise to a thermal photonic state with a temperature defined by the cavity-pump-to-loss ratio. This overlooks the fact that the background photons are generated through exciting an active medium, and hence are expected to have a coherent component [89]. On the other hand, detuned emitters can be explicitly included in the calculation, and their coupling to a detuned cavity mode via phonon or Auger coupling is facilitated by an effective system-bath coupling (Lindblad) term.This approach has the advantage that properties of the gain medium are preserved. The resulting cavity feeding is strongly non-linear with pump rate and exhibits saturation if the number of detuned emitters is limited. Furthermore, the overall emission can become coherent if the background contributions drive the laser above the threshold [89].

As an example, the left part of Figure 48.21 [109] shows the spectrum of a few-QD photonic-crystal nanocavity laser. Only few discrete QD lines are found in the vicinity of the mode (M1). The input–output characteristics is that of a nanolaser with a slight kink, hinting at a β-factor close to unity (middle panel), and coherent emission with $g^{(2)}(0) = 1$ being reached (right panel). Remarkably, the performance of the device is insensitive to tuning of the cavity-mode energy, which can be realized by gas deposition on the photonic crystal cavity. Evidence for this is found in the autocorrelation function in the right panel, which shows a clear transition to lasing with $g^{(2)}(0) = 1$ *independently* of the exact cavity position. In a system with few discrete emission channels this provides strong evidence for the importance of nonresonant coupling effects. Similar results have been reported in [110], to which the authors referred to as "self-tuning gain" effect. Moreover, in the spontaneous emission regime the coupling of various multiexcitonic emission channels from different emitters inside the cavity leads to super-thermal photon bunching with $g^{(2)}(0) > 2$. A density-matrix theory including quantum-mechanical correlations between these transitions, shown as green curve in the right panel, is in good agreement with the measured data. Neither rate equations, nor conventional laser models can unravel the intricate quantum-mechanical nature that governs the laser properties of few-emitter cavity-QED systems.

48.9 Outlook

With present-day nanolaser devices, we are closer than ever to reaching fundamental limits of light–matter interaction in semiconductor systems. Electronic excitations can be generated and used on the level of single excitons [111,112]. Brought into interaction with a single mode of microresonator, semiconductor nanolasers operate at the transition to the regime of quantum optics, where cavity-QED effects, such as photon blockade, photon-antibunching, and vacuum Rabi oscillations can coexist with lasing in different excitation regimes of the same device [106,113,114]. This offers fascinating prospects both from a fundamental and an applied point of view.

Acknowledgments

The authors would like to thank Joachim Piprek for the kind invitation to write this chapter. We also thank Andreas Beuthner for his help in creating the schematic figures. Christopher Gies and Frank Jahnke acknowledge financial support from the German Science Foundation (DFG).

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC., a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA-0003525. WWC thanks the hospitality of the Technical University Berlin and travel support from the German Research Foundation via the collaborative research center 787.

References

1. Vahala, K. J. (2003). Optical microcavities. *Nature 424* (6950), 839–846.
2. Borri, P., W. Langbein, S. Schneider, U. Woggon, R. L. Sellin, D. Ouyang, and D. Bimberg (2002). Relaxation and dephasing of multiexcitons in semiconductor quantum dots. *Physical Review Letters 89* (18), 187401.
3. Borri, P., W. Langbein, U. Woggon, V. Stavarache, D. Reuter, and A. D. Wieck (2005). Exciton dephasing via phonon interactions in InAs quantum dots: Dependence on quantum confinement. *Physical Review B 71* (11), 115328.
4. Florian, M., A. Steinhoff, C. Gies, and F. Jahnke (2016). Scattering-induced dephasing of many-particle transitions in semiconductor quantum dots. *Applied Physics B 122* (1), 1–7.
5. Blood, P. (2013). The laser diode: 50 years on. *IEEE Journal of Selected Topics in Quantum Electronics 19* (4), 1503201–1503201.
6. Ning, C. (2013). What is laser threshold? *IEEE Journal of Selected Topics in Quantum Electronics 19* (4), 1503604–1503604.
7. Rice, P. R. and H. J. Carmichael (1994). Photon statistics of a cavity-QED laser: A comment on the laser–phase-transition analogy. *Physical Review A 50* (5), 4318–4329.
8. Chow, W. W., M. Lorke, and F. Jahnke (2011). Will quantum dots replace quantum wells as the active medium of choice in future semiconductor lasers? *IEEE Journal of Selected Topics in Quantum Electronics 17*, 1349–1355.
9. Yokoyama, H. and S. D. Brorson (1989). Rate equation analysis of microcavity lasers. *Journal of Applied Physical 66*, 4801.
10. Musiał, A., C. Hopfmann, T. Heindel, C. Gies, M. Florian, H. A. M. Leymann, A. Foerster, C. Schneider, F. Jahnke, S. Höfling, M. Kamp, and S. Reitzenstein (2015). Correlations between axial and lateral emission of coupled quantum dot micropillar cavities. *Physical Review B 91* (20), 205310.
11. Munsch, M., N. S. Malik, E. Dupuy, A. Delga, J. Bleuse, J.-M. Gérard, J. Claudon, N. Gregersen, and J. Mørk (2013). Dielectric GaAs antenna ensuring an efficient broadband coupling between an InAs quantum dot and a Gaussian optical beam. *Physical Review Letters 110* (17), 177402.
12. Leymann, H. A. M., C. Hopfmann, F. Albert, A. Foerster, M. Khanbekyan, C. Schneider, S. Höfling, A. Forchel, M. Kamp, J. Wiersig, and S. Reitzenstein (2013). Intensity fluctuations in bimodal micropillar lasers enhanced by quantum-dot gain competition. *Physical Review A 87* (5), 053819.
13. Agrawal, G. P. and N. K. Dutta (1993). *Semiconductor Laser*. Norwell, MA: Kluwer.
14. Chuang, S. L. (1995). *Physics of Optoelectronic Devices*. New York, NY: Wiley.
15. Lohmeyer, H., K. Sebald, C. Kruse, R. Kröger, J. Gutowski, D. Hommel, J. Wiersig, N. Baer, and F. Jahnke (2006). Confined optical modes in monolithic II-VI pillar microcavities. *Applied Physics Letters 88* (5), 051101.
16. McCall, S. L., A. F. J. Levi, R. E. Slusher, S. J. Pearton, and R. A. Logan (1992). Whispering gallery mode microdisk lasers. *Applied Physics Letters 60* (3), 289–291.
17. Joannopoulos, J. D., P. R. Villeneuve, and S. Fan (1997). Photonic crystals: Putting a new twist on light. *Nature 386* (6621), 143–149.
18. Purcell, E. M. (1946). Spontaneous emission probability at radio frequencies. *Physical Review 69*, 681.

19. Schwab, M., H. Kurtze, T. Auer, T. Berstermann, M. Bayer, J. Wiersig, N. Baer, C. Gies, F. Jahnke, J. P. Reithmaier, A. Forchel, M. Benyoucef, and P. Michler (2006). Radiative emission dynamics of quantum dots in a single cavity micropillar. *Physical Review B 74* (4), 045323.
20. Aßmann, M., F. Veit, M. Bayer, M. v. d. Poel, and J. M. Hvam (2009). Higher-order photon bunching in a semiconductor microcavity. *Science 325* (5938), 297–300. PMID: 19608912.
21. Loudon, R. (2000). *The Quantum Theory of Light*. Oxford, UK: Oxford University Press.
22. Hanbury Brown, R. and R. Q. Twiss (1956). Correlation between photons in two coherent beams of light. *Nature 177*, 27–29.
23. Ulrich, S. M., C. Gies, J. Wiersig, S. Reitzenstein, C. Hofmann, A. Löffer, A. Forchel, F. Jahnke, and P. Michler (2007). Photon statistics of semiconductor microcavity lasers. *Physical Review Letters 98*, 043906.
24. Gardiner, C. and P. Zoller (2004). *Quantum Noise A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods with Applications to Quantum Optics*. Heidelberg, Berlin, New York: Springer Verlag.
25. Gies, C., M. Florian, P. Gartner, and F. Jahnke (2012). Modelling single quantum dots in microcavities. In F. Jahnke (Ed.), *Quantum Optics with Semiconductor Nanostructures*. Philadelphia, PA: Woodhead Publishing Limited.
26. Ates, S., C. Gies, S. M. Ulrich, J. Wiersig, S. Reitzenstein, A. Löffler, A. Forchel, F. Jahnke, and P. Michler (2008). Influence of the spontaneous optical emission factor β on the first-order coherence of a semiconductor microcavity laser. *Physical Review B 78* (15), 155319.
27. Chow, W. W., F. Jahnke, and C. Gies (2014). Emission properties of nanolasers during the transition to lasing. *Light: Science & Applications 3* (8), e201.
28. Ritter, S., P. Gartner, C. Gies, and F. Jahnke (2010). Emission properties and photon statistics of a single quantum dot laser. *Optics Express 18* (10), 9909–9921.
29. Reitzenstein, S., C. Böckler, A. Bazhenov, A. Gorbunov, A. Löffer, M. Kamp, V. D. Kulakovskii, and A. Forchel (2008). Single quantum dot controlled lasing effects in high-q micropillar cavities. *Optics Express 16* (7), 4848–4857.
30. Gies, C., J. Wiersig, and F. Jahnke (2008). Output characteristics of pulsed and continuous-wave-excited quantum-dot microcavity lasers. *Physical Review Letters 101* (6), 067401.
31. Chow, W. and S. Koch (2005). Theory of semiconductor quantum-dot laser dynamics. *IEEE Journal of Quantum Electronics 41* (4), 495–505.
32. Kolarczik, M., N. Owschimikow, J. Korn, B. Lingnau, Y. Kaptan, D. Bimberg, E. Schöll, K. Lüdge, and U. Woggon (2013). Quantum coherence induces pulse shape modification in a semiconductor optical amplifier at room temperature. *Nature Communications 4*, 2953.
33. Gies, C., J. Wiersig, M. Lorke, and F. Jahnke (2007). Semiconductor model for quantum-dot-based microcavity lasers. *Physical Review A 75* (1), 013803.
34. Wiersig, J., C. Gies, F. Jahnke, M. Aßmann, T. Bestermann, M. Bayer, C. Kistner, S. Reitzenstein, S. H. A. Forchel, C. Kruse, J. Kalden, and D. Hommel (2009). Direct observation of correlations between individual photon emission events of a microcavity laser. *Nature 460*, 245.
35. Baer, N., C. Gies, J. Wiersig, and F. Jahnke (2006). Luminescence of a semiconductor quantum dot system. *European Physical Journal B 50*, 411.
36. Gregersen, N., T. Suhr, M. Lorke, and J. Mørk (2012). Quantum-dot nano-cavity lasers with Purcell-enhanced stimulated emission. *Applied Physics Letters 100* (13), 131107.
37. Lermer, M., N. Gregersen, M. Lorke, E. Schild, P. Gold, J. Mørk, C. Schneider, A. Forchel, S. Reitzenstein, S. Hing, and M. Kamp (2013). High beta lasing in micropillar cavities with adiabatic layer design. *Applied Physics Letters 102* (5), 052114.
38. Lorke, M., T. Suhr, N. Gregersen, and J. Mørk (2013). Theory of nanolaser devices: Rate equation analysis versus microscopic theory. *Physical Review B 87*, 205310.

39. Suhr, T., N. Gregersen, M. Lorke, and J. Mørk (2011). Modulation response of quantum dot nanolightemitting-diodes exploiting Purcell-enhanced spontaneous emission. *Applied Physics Letters 98* (21), 211109.
40. Chow, W. and S. Koch (1999). *Semiconductor-Laser Fundamentals* (1st ed.). Berlin: Springer-Verlag.
41. Coldren, L. and S. Corzine (1995). *Diode Lasers and Photonic Integrated Circuits*. New York, NY: Wiley.
42. Binder, R., H. Köhler, M. Bonitz, and N. Kwong (1997). Green's function description of momentum orientation relaxation of photoexcited electron plasmas in semiconductors. *Physical Review B 55*, 5110.
43. Bonitz, M., D. Kremp, D. Scott, R. Binder, W. Kraeft, and H. Köhler (1996). Numerical analysis of nonmarkovian effects in charge-carrier scattering: One-time kinetic equations. *Journal of Physics: Condensed Mattter 8*, 6057.
44. ElSayed, K., S. Schuster, H. Haug, G. Herzel, and K. Henneberger (1994). Subpicosecond plasmon response: Buildup of screening. *Physical Review B 49*, 7337.
45. Henneberger, K. (1988). Resonant laser excitation and electron-hole kinetics of a semiconductor: I. nonequilibrium green's function treatment and fundamental equations. *Physica A 150*, 419.
46. Jahnke, F. and S. W. Koch (1995). Many-body theory for semiconductor microcavity lasers. *Physical Review A 52*, 1712.
47. Bányai, L., P. Gartner, and H. Haug (1998). Self-consistent RPA retarded polaron Green function for quantum kinetics. *European Physical Journal B 1*, 209.
48. Gartner, P., L. Bányai, and H. Haug (1999). Two-time electron-LO-phonon quantum kinetics and the generalized Kadanoff-Baym approximation. *Physical Review B 60*, 14234.
49. Gartner, P., L. Bányai, and H. Haug (2002). Self-consistent RPA for the intermediate-coupling polaron. *Physical Review B 66*, 75205.
50. Hartmann, M. and W. Schäfer (1992). Real time approach to relaxation and dephasing processes in semiconductors. *Physical Status Solidi (B) 173*, 165.
51. Haug, H. (1992). Interband quantum kinetics with LO-phonon scattering in a laser-pulse-excited semiconductor I. Theory. *Physical Status Solidi (B) 173*, 139.
52. Jahnke, F. and S. W. Koch (1993). Theory of carrier heating through injection pumping and lasing in semiconductor microcavity lasers. *Optics Letters 18* (17), 1438–1440.
53. Lorke, M., T. R. Nielsen, and J. Mørk (2011). Switch-on dynamics of nanocavity laser devices. *Applied Physics Letters 99* (15), 151110.
54. Lingnau, B., K. Lüdge, E. Schöll, and W. W. Chow (2010). Many-body and nonequilibrium effects on relaxation oscillations in a quantum-dot microcavity laser. *Applied Physics Letters 97*, 111102.
55. Lorke, M., T. R. Nielsen, and J. Mørk (2010). Influence of carrier dynamics on the modulation bandwidth of quantum-dot based nanocavity devices. *Applied Physics Letters 97* (21), 211106.
56. Luedge, K., M. J. P. Bormann, E. Malic, P. Hoevel, M. Kuntz, D. Bimberg, A. Knorr, and E. Schoell (2008). Turn-on dynamics and modulation response in semiconductor quantum dot lasers. *Physical Review B 78*, 035316.
57. Luedge, K. and E. Scholl (2009). Quantum-dot lasers—Desynchronized nonlinear dynamics of electrons and holes. *IEEE Journal of Quantum Electronics 45* (11), 1396–1403.
58. Aßmann, M., F. Veit, M. Bayer, C. Gies, F. Jahnke, S. Reitzenstein, S. Höfling, L. Worschech, and A. Forchel (2010). Ultrafast tracking of second-order photon correlations in the emission of quantum-dot microresonator lasers. *Physical Review B 81* (16), 165314.
59. Lorke, M., W. Chow, T. R. Nielsen, J. Seebeck, P. Gartner, and F. Jahnke (2006). Anomaly in the excitation dependence of optical gain in semiconductor quantum dots. *Physical Review B 74*, 035334.
60. Lorke, M., T. Nielsen, J. Seebeck, P. Gartner, and F. Jahnke (2006). Influence of carrier-carrier and carrier-phonon correlations on optical absorption and gain in quantum-dot systems. *Physical Review B 73*, 85324.

61. Schäfer, W. and M. Wegener (2002). *Semiconductor Optics and Transport Phenomena* (1st ed.). Berlin: Springer-Verlag.
62. Mahan, G. (1990). *Many-Particle Physics*. New York, NY: Plenum Press.
63. Seebeck, J., T. Nielsen, P. Gartner, and F. Jahnke (2005). Polarons in semiconductor quantum dots and their role in the quantum kinetics of carrier relaxation. *Physical Review B 71*, 125327.
64. Nielsen, T. R., P. Gartner, and F. Jahnke (2004). Many-body theory of carrier capture and relaxation in semiconductor quantum-dot lasers. *Physical Review B 69* (23), 235314.
65. Benisty, H. (1995). Reduced electron-phonon relaxation rates in quantum-box systems: Theoretical analysis. *Physical Review B 51*, 13281.
66. Inoshita, T. and H. Sakaki (1996). Electron-phonon interaction and the so-called phonon bottleneck effect in semiconductor quantum dots. *Physica B 227*, 373.
67. Kurtze, H., J. Seebeck, P. Gartner, D. R. Yakovlev, D. Reuter, A. D. Wieck, M. Bayer, and F. Jahnke (2009). Carrier relaxation dynamics in self-assembled semiconductor quantum dots. *Physical Review B 80*, 235319.
68. Schuh, K., F. Jahnke, and M. Lorke (2011). Rapid adiabatic passage in quantum dots: Influence of scattering and dephasing. *Applied Physics Letters 99* (1), 011105.
69. Seebeck, J., T. Nielsen, P. Gartner, and F. Jahnke (2006). Quantum kinetic theory of phonon-assisted carrier transitions in nitride-based quantum-dot systems. *European Physical Journal B 49*, 167.
70. Seebeck, J., M. Lorke, P. Gartner, and F. Jahnke (2009). Carrier-carrier and carrier-phonon scattering in the low-density and low-temperature regime for resonantly pumped semiconductor quantum dots. *Physica Status Solidi (C) 6*, 488.
71. Uskov, A. V., C. Meuer, H. Schmeckebier, and D. Bimberg (2011). Auger capture induced carrier heating in quantum dot lasers and amplifiers. *Applied Physics Express 4* (2), 022202.
72. Goldmann, E., M. Lorke, T. Frauenheim, and F. Jahnke (2014). Negative differential gain in quantum dot systems: Interplay of structural properties and many-body effects. *Applied Physics Letters 104* (24), 242108.
73. Lorke, M., W. Chow, and F. Jahnke (2007). Reduction of optical gain in semiconductor quantum dots. *Proceedings of SPIE 6468*, 646818.
74. Lorke, M., F. Jahnke, and W. Chow (2007). Excitation dependence of gain and carrier induced refractive index changes in quantum dots. *Applied Physical Letters 90*, 51112.
75. Lorke, M., J. Seebeck, T. Nielsen, P. Gartner, and F. Jahnke (2006). Excitation dependence of the homogeneous linewidths in quantum dots. *Physical Status Solidi (C) 3*, 2393–2396.
76. Dicke, R. H. (1954). Coherence in spontaneous radiation processes. *Physical Review 93*, 99–110.
77. Mandel, L. and E. Wolf (1995). *Optical Coherence and Quantum Optics*. New York, NY: Cambridge University Press.
78. Bohnet, J. G., Z. Chen, J. M. Weiner, D. Meiser, M. J. Holland, and J. K. Thompson (2012). A steady-state superradiant laser with less than one intracavity photon. *Nature 484* (7392), 78–81.
79. Timothy Noe Ii, G., J.-H. Kim, J. Lee, Y. Wang, A. K. Wójcik, S. A. McGill, D. H. Reitze, A. A. Belyanin, and J. Kono (2012). Giant superfluorescent bursts from a semiconductor magneto-plasma. *Nature Physics 8* (3), 219–224.
80. Leymann, H. A. M., A. Foerster, F. Jahnke, J. Wiersig, and C. Gies (2015). Sub—And superradiance in nanolasers. *Physical Review Applied 4* (4), 044018.
81. Auffféves, A., D. Gerace, S. Portolan, A. Drezet, and M. F. Santos (2011). Few emitters in a cavity: From cooperative emission to individualization. *New Journal of Physics 13* (9), 093020.
82. Jahnke, F., C. Gies, M. Aßmann, M. Bayer, H.A.M. Leymann, A. Foerster, J. Wiersig, C. Schneider, M. Kamp, and S. Höfling (2016). Giant photon bunching, superradiant pulse emission and excitation trapping in quantum-dot nanolasers. *Nature Communications 7*, 11540.
83. Bimberg, D., M. Grundmann, and N. N. Ledentsov (1998). *Quantum Dot Heterostructures*. Hoboken, NJ: John Wiley & Sons.

84. Bayer, M. and A. Forchel (2002). Temperature dependence of the exciton homogeneous linewidth in In0.60Ga0.40As/GaAs self-assembled quantum dots. *Physical Review B 65* (4), 041308.
85. Moody, G., R. Singh, H. Li, I. A. Akimov, M. Bayer, D. Reuter, A. D. Wieck, and S. T. Cundiff (2013). Fifth-order nonlinear optical response of excitonic states in an InAs quantum dot ensemble measured with two-dimensional spectroscopy. *Physical Review B 87* (4), 045313.
86. Scheibner, M., T. Schmidt, L. Worschech, A. Forchel, G. Bacher, T. Passow, and D. Hommel (2007). Superradiance of quantum dots. *Nature Physics 3* (2), 106–110.
87. Lichtmannecker, S., M. Florian, T. Reichert, M. Blauth, M. Bichler, et al. (2017). A few-emitter solid-state multi-exciton laser. *Scientific Reports 7* (7420). doi: 10.1038/s41598-017-07097-9
88. Ates, S., S. M. Ulrich, A. Ulhaq, S. Reitzenstein, A. Löffler, S. Höfling, A. Forchel, and P. Michler (2009). Non-resonant dot—cavity coupling and its potential for resonant single-quantum-dot spectroscopy. *Nature Photonics 3* (12), 724–728.
89. Florian, M., P. Gartner, C. Gies, and F. Jahnke (2013). Phonon-mediated off-resonant coupling effects in semiconductor quantum-dot lasers. *New Journal of Physics 15* (3), 035019.
90. Hohenester, U., A. Laucht, M. Kaniber, N. Hauke, A. Neumann, A. Mohtashami, M. Seliger, M. Bichler, and J. J. Finley (2009). Phonon-assisted transitions from quantum dot excitons to cavity photons. *Physical Review B 80* (20), 201311.
91. Majumdar, A., E. D. Kim, Y. Gong, M. Bajcsy, and J. Vučković (2011). Phonon mediated off-resonant quantum dot–cavity coupling under resonant excitation of the quantum dot. *Physical Review B 84* (8), 085309.
92. Press, D., S. Götzinger, S. Reitzenstein, C. Hofmann, A. Löffler, M. Kamp, A. Forchel, and Y. Yamamoto (2007). Photon antibunching from a single quantum-dot-microcavity system in the strong coupling regime. *Physical Review Letters 98* (11), 117402.
93. Tarel, G. and V. Savona (2009). Emission spectrum of a quantum dot embedded in a nanocavity. *Physica Status Solidi (C) 6* (4), 902–905.
94. Roy, C. and S. Hughes (2011). Influence of electron–acoustic-phonon scattering on intensity power broadening in a coherently driven quantum-dot–cavity system. *Physical Review X 1* (2), 021009.
95. Chauvin, N., C. Zinoni, M. Francardi, A. Gerardino, L. Balet, B. Alloing, L. H. Li, and A. Fiore (2009). Controlling the charge environment of single quantum dots in a photonic-crystal cavity. *Physical Review B 80* (24), 241306.
96. Florian, M., P. Gartner, A. Steinhoff, C. Gies, and F. Jahnke (2014). Coulomb-assisted cavity feeding in nonresonant optical emission from a quantum dot. *Physical Review B 89* (16), 161302(R).
97. Karrai, K., R. J. Warburton, C. Schulhauser, A. Högele, B. Urbaszek, E. J. McGhee, A. O. Govorov, J. M. Garcia, B. D. Gerardot, and P. M. Petroff (2004). Hybridization of electronic states in quantum dots through photon emission. *Nature 427* (6970), 135–138.
98. Settnes, M., P. Kaer, A. Moelbjerg, and J. Mork (2013). Auger processes mediating the nonresonant optical emission from a semiconductor quantum dot embedded inside an optical cavity. *Physical Review Letters 111* (6), 067403.
99. Winger, M., T. Volz, G. Tarel, S. Portolan, A. Badolato, K. J. Hennessy, E. L. Hu, A. Beveratos, J. Finley, V. Savona, and A. Imamoğlu (2009). Explanation of photon correlations in the far-off-resonance optical emission from a quantum-dot–cavity system. *Physical Review Letters 103* (20), 207403.
100. Bockelmann, U. and T. Egeler (1992). Electron relaxation in quantum dots by means of auger processes. *Physical Review B 46*, 15574.
101. Efros, A. L., V. A. Kharchenko, and M. Rosen (1995). Breaking of the phonon bottleneck in nanometer quantum dots role of auger-like processes. *Solid State Communications 93*, 281.
102. Uskov, A. V., F. Adler, H. Schweizer, and M. H. Pilkuhn (1997). Auger carrier relaxation in self-assembled quantum dots by collisions with two-dimensional carriers. *Journal of Applied Physics 81*, 7895.

103. Dekel, E., D. Gershoni, E. Ehrenfreund, D. Spektor, J. M. Garcia, and P. M. Petroff (1998). Multiexciton spectroscopy of a single self-assembled quantum dot. *Physical Review Letters 80* (22), 4991–4994.
104. Laucht, A., M. Kaniber, A. Mohtashami, N. Hauke, M. Bichler, and J. J. Finley (2010). Temporal monitoring of nonresonant feeding of semiconductor nanocavity modes by quantum dot multiexciton transitions. *Physical Review B 81*, 241302.
105. del Valle, E., F. P. Laussy, and C. Tejedor (2009). Luminescence spectra of quantum dots in microcavities. II. Fermions. *Physical Review B 79*, 235326.
106. Nomura, M., N. Kumagai, S. Iwamoto, Y. Ota, and Y. Arakawa (2010). Laser oscillation in a strongly coupled single-quantum-dot–nanocavity system. *Nature Physics 6*, 279–283.
107. Valle, E. d. and F. P. Laussy (2011). Effective cavity pumping from weakly coupled quantum dots. *Superlattices and Microstructures 49* (3), 241–245.
108. Yao, P., P. K. Pathak, E. Illes, S. Hughes, S. Münch, S. Reitzenstein, P. Franeck, A. Löffer, T. Heindel, S. Höfling, L. Worschech, and A. Forchel (2010). Nonlinear photoluminescence spectra from a quantum-dot cavity system: Interplay of pump-induced stimulated emission and anharmonic cavity QED. *Physical Review B 81* (3), 033309.
109. Lichtmannecker, S., M. Florian, T. Reichert, M. Blauth, M. Bichler, F. Jahnke, J. J. Finley, C. Gies, and M. Kaniber (2016). A few-emitter solid-state multi-exciton laser. *ArXiv:1602.03998 [condmat].*
110. Strauf, S., K. Hennessy, M. T. Rakher, Y.-S. Choi, A. Badolato, L. C. Andreani, E. L. Hu, P. M. Petroff, and D. Bouwmeester (2006). Self-tuned quantum dot gain in photonic crystal lasers. *Physical Review Letters 96*, 127404.
111. Mlynek, J. A., A. A. Abdumalikov, C. Eichler, and A. Wallraff (2014). Observation of Dicke superradiance for two artificial atoms in a cavity with high decay rate. *Nature Communications 5*, 5186.
112. Müller, M., S. Bounouar, K. D. Jöns, M. Glässl, and P. Michler (2014). On-demand generation of indistinguishable polarization-entangled photon pairs. *Nature Photonics 8* (3), 224–228.
113. Muñoz, C. S., E. del Valle, A. G. Tudela, K. Müller, S. Lichtmannecker, M. Kaniber, C. Tejedor, J. J. Finley, and F. P. Laussy (2014). Emitters of n-photon bundles. *Nature Photonics 8* (7), 550–555.
114. Gies, C. et al., Phys. Rev. A **96**, 023806 (2017).

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

49

Nonlinear Dynamics in Quantum Photonic Structures

49.1 Introduction 662

49.2 Maxwell-Pseudospin Method 667

Pseudospin Equation • Dissipative Dynamics and Master Pseudospin Equation • Vector Maxwell Equations and Coupling to Pseudospin Equations • Coherent Maxwell-Pseudospin Equations in 2D • Numerical Methodology

49.3 Simulation of Nonlinear Dynamics in Optical Waveguides and Semiconductor Microcavities 674

One-Dimensional Model of a Two-Level System • Passive Cavity Properties • Active Nonlinear Optical Waveguide/Cavity Properties • Resonant Nonlinearities and Gain Saturation • Coherent Propagation Effects, Self-Localization and Pattern Formation in 2D Planar Nonlinear Optical Waveguides and Semiconductor Microcavities

49.4 Optical Orientation in Quantum Wells and QDs 691

Circularly Polarized Optical-Pulse Interactions with Two-Level Quantum Systems: Selective Spin Initialization and Readout • Circularly Polarized Optical Pulse Interactions with Degenerate Four-Level Systems

49.5 Modeling Coherent Spin Generation, Manipulation, and Readout in Charged QDs 699

Single-Charged QD Embedded in a Waveguide • Coherent Spin Manipulation through Hot-Charged Exciton States in QDs

49.6 Quantum Stochastic Formalism for Modeling Cavity-Emitter Systems 707

Langevin Formalism for Modeling Quantum-Optical Effects • Spontaneous Emission of a QD Ensemble in a Semiconductor Microcavity and Onset of Lasing

49.7 Modeling Optical Rotation in Chiral Carbon Nanotubes with Master Maxwell-Pseudospin Equations 712

Natural Optical Activity • Faraday Rotation and Magneto-Chiral Effects

49.8 Conclusion 724

Gabriela Slavcheva

and

Mirella Koleva

49.1 Introduction

Quantum optics unites physical optics and the quantum field theory of light. The foundations of the latter were laid by Planck, who succeeded to explain theoretically the black-body radiation spectrum by postulating that the energy of a harmonic oscillator is quantized. In 1927, Dirac developed this theory by solving the long-standing problem of the wave-particle duality and unifying the wave and particle aspects of light by quantization of the electromagnetic radiation. Thus, according to Dirac's single-particle interpretation, the light emitted from a spontaneously emitting source with low photon density is described as single-photon emission events.

A fundamental consequence of light-field quantization is the appearance of vacuum (zero-point energy) fluctuations, or quantum noise, which has no classical analogue and stems from the Heisenberg uncertainty relations. However, some of the greatest successes of quantum theory—such as Planck's black-body spectrum and Einstein's theory of the photoelectric effect—do not actually require a full quantum-mechanical treatment and can instead be explained using semiclassical theory, whereby matter is quantized but light remains a classical wave. Only in the 1960s, motivated by the development of lasers, did researchers start to examine effects which truly required a quantum theory of light. For instance, light quantization is required to explain spontaneous emission, irreversible decay, and the origin of the laser spectral linewidth. The demand to develop further the quantum theory of light in order to explain these effects was met by the pioneering work of Glauber [1,2].

Vacuum quantum fluctuations are necessary to account for phenomena, such as the Lamb shift: a radiative correction to the energy spectrum of hydrogen, as well as for the attractive force between two uncharged conductive plates in a vacuum placed a few nanometres apart, known as the Casimir effect. Some examples of effects which the semiclassical approach fails to explain are quantum beat phenomena in a three-level atom, two-photon interferometry and production of "entangled" states of light, resonance fluorescence (quantum jumps) in the emission of atoms, and the so-called "nonclassical" states of light.

The classification of light states as classical and nonclassical stems from the statistical properties of radiation and, more specifically, from the second-order coherent properties of the light emitter. A crucial step in elucidating these properties was the correlation experiment of Hanbury Brown and Twiss [3], which was designed to measure the coincidence of individual photons at two detectors placed equidistantly from a half-silvered mirror splitting a beam of oncoming low-intensity light. Using the second-order correlation function, $g^{(2)}(\tau)$, quantifying the synchronicity with which photons arrive at the detectors, it was then possible to discern the manner in which different types of light sources emitted light. This method revealed that classical light sources—such as a thermal black-body radiation source or a lamp—exhibited Bose–Einstein statistics manifested by photon "bunching", i.e., an increased likelihood of the photons to arrive at the detectors at the same time, as though they had "bunched together" while traveling from the source to the detectors. On the other hand, a stable coherent light source (e.g., a laser) exhibits Poissonian statistics, which is manifested by a constant unity second-order correlation function, $g^{(2)}(\tau)$, meaning that photons are not intercorrelated and are equally likely to impinge on the detectors with a delay of any length with respect to each other. However, most interestingly of all, nonclassical states of light are described by sub-Poissonian statistics (a diminished likelihood of coincident arrival) called "antibunching": the photons are "diluted" as they travel and eventually emitted one by one. Antibunching is characteristic of resonant fluorescence from single atoms, molecules, or semicondictor nanocrystals and quantum dots (QDs), and is a signature of a single-photon emitter. For perfectly antibunched light, the second-order correlation function vanishes for coincident detections. This is irreconcilable with the classical wave optics and is considered a hallmark of pure quantum behavior.

Photon-number (Fock) states are states of light containing a specific number of photons. The single-photon state is a special case of a Fock state and is a squeezed-light state [4], a quantum state of light resulting from the Heisenberg relations with, e.g., reduced amplitude fluctuations at the expense of enhanced fluctuations of the phase. An ideal amplitude-squeezed source delivers a stream of photons at regular time intervals, which is the precise realization of a single-photon source [5].

Another quantum-mechanical property is entanglement. Quantum systems consisting of interacting subsystems become highly correlated and their individual constituents become "entangled". The global quantum state of an entangled system—for e.g., one consisting of two particles (e.g., a photon pair)—remains correlated even when the two particles are separated by large distances from each other and unable to interact or communicate in any known way. The most famous example of an entangled state is the Einstein–Podolsky–Rosen–Bohm [6] state, whereby two spin 1/2 particles reside in a state with zero total spin angular momentum: $|\Psi\rangle = \frac{1}{\sqrt{2}}(|\uparrow\rangle \otimes |\downarrow\rangle - |\downarrow\rangle \otimes |\uparrow\rangle)$. Notably, this state cannot be represented as a simple product of the pure states of the individual particles.

Many of these quantum effects have been observed only in carefully controlled atomic systems; however, with the advent of quantum photonics (the applied device-oriented field of quantum optics), recent advances in semiconductor technology have enabled the fabrication of solid-state devices that emit single photons, opening new avenues for nonclassical light state generation. For instance, the use of radiative cascades from QD excited states is a principal method for generation of quantum entanglement for applications of QDs as polarization-entangled sources. The exchange interaction and the fine structure splitting of the energy levels involved in radiative cascades from QD excited states are of fundamental importance, since they determine the polarization and entanglement of the emitted photons.

Quantum photonics has attracted great interest in recent years. This is due to its potential to revolutionize science and day-to-day life alike through enabling the implementation of faster algorithms, secure transmission of information, vast increase in data storage, and execution of more accurate measurements. One of the major challenges of modeling quantum photonic devices is working out a way of efficient simulation and control of realistic, "open" quantum systems (i.e., allowing for energy to flow in and out of the system) and devices for applications in quantum technologies. For optical quantum information processing to advance beyond demonstration experiments, the development of on-chip capabilities will be essential. Solid-state qubit architectures have emerged as a most promising physical implementation of quantum photonic circuits, taking advantage of the state-of-the-art semiconductor technologies. Reliable, integrated sources of nonclassical light, as well as realization of high-fidelity on-chip readout of solid-state qubits, are a critical requirement for next-generation quantum technologies [7].

Recently, significant progress has been made toward the achievement of high-efficiency solid-state single-photon sources [8–12]. Among these sources, QDs are particularly promising since they can be embedded in devices enabling electrical injection while maintaining efficient light emission. Furthermore, the dot doping—or the number and type of the charge carriers within the dot—can be controlled by applying electrical bias in a Schottky-type structure. A quantum bit (qubit) is a linear superposition of basis states $|0\rangle$ (ground level) and $|1\rangle$ (excited level) in a two-level system, given by: $|\Psi\rangle = \alpha|0\rangle + \beta|1\rangle$, with α and β being complex numbers such that $\alpha^2 + \beta^2 = 1$. In a classical computer, the information is encoded in digital bit states 0 and 1. By contrast, in a quantum computer, the state of each bit is permitted to be any quantum-mechanical state of a qubit (manipulated using qubit rotations, also known as gate operations), leading to a massive parallelism and an exponential speed-up of the quantum computations.

The $|0\rangle$ and $|1\rangle$ states of a qubit could be encoded in the spin-up ($|\uparrow\rangle$) and spin-down ($|\downarrow\rangle$) states of a charge carrier. The spin-up and spin-down states of a single electron or hole confined in a charged semiconductor QD are a particularly attractive qubit candidate [13]. This arises from several advantages intrinsic to this quantum system. The first advantage is that the atomic-like electronic structure suppresses coupling of the spin to the solid-state quantum environment and thus impedes the spin decoherence processes. This leads to a very long spin coherence time (denoted T_2)—that is, the time within which the spin loses its phase information—and thus a long-lived quantum state, which is required for performing a large number of spin manipulations. Another important advantage is optical convertibility—i.e., the ability to switch through transfer of angular momentum between a "flying" qubit encoded in a photon and the stationary matter qubit of the spin confined in a QD. This enables the transport of qubits over large distances and facilitates the realization of scalable solid-state qubit architectures for optical quantum computing. Furthermore, this promising platform allows integrating the operations of optical initialization, manipulation, and readout of a spin qubit on a chip. By employing a combination of optical orientation [14] and quantum coherent control [15] techniques, using ultrashort polarized light pulses one can address individual carrier spins in semiconductor QDs, initialize and manipulate them coherently through optically excited states (charged excitons), and detect the resulting spin qubit [16]. For quantum technologies applications, control of both QD position and emission energy is crucial. For instance, deterministic site control of QD location is indispensable for multiqubit operations and scalability. Therefore, new simulation tools are needed to determine the allowed tolerances in the dot positions and emission energies to ensure observation of the desired quantum-optical effects.

Quantum coherent control represents a universal approach for predictable manipulation of the properties of quantum systems, such as atoms and molecules. This technique has most recently been applied to QD solid-state systems. Coherent control in semiconductor nanostructures allows for coherent manipulation of the carrier wavefunctions on a time scale shorter than typical dephasing times. This is a prerequisite for successful implementation of ultrafast optical switching and quantum information processing. Quantum coherent control requires use of ultrashort pulses considerably shorter than the characteristic relaxation times in matter ($\tau \ll T_1, T_2$). This is equivalent to the photon-dipole coupling rate exceeding all dissipative rates in the system, a general condition for achieving the strong-coupling regime. The ultrashort pulses are usually characterized by high field amplitudes and, consequently, lead to nonlinear optical effects, such as coherent pulse propagation and self-induced transparency (SIT) [17]. The phenomenon of SIT is observable above a critical power threshold for a given pulse width: a high-intensity, ultrashort pulse propagating through a medium composed of an ensemble of resonant quantum two-level absorbers, whose relaxation times greatly exceed the pulse duration, travels at a reduced speed and unchanged shape with anomalously low energy loss. The absorbers are driven into the excited state by absorbing ultrashort pulse energy; by reradiating this energy into the pulse, they return to the ground state. Thus, the optical energy is carried through the medium not by the electromagnetic field, but by a coupled light-matter polariton wave. The polariton is a mixed light-matter quasiparticle resulting from the strong coupling of the optical wave to the medium's polarization. As a result, the pulse travels as a solitary wave, known as a SIT-soliton. This soliton is localized both in space and time, in contrast with the well-known nonlinear optics solitons which result from the interplay between nonlinearity and the medium's dispersion and/or diffraction. The condition of SIT is predicted by the remarkable pulse-area theorem (PAT) [18] which establishes a general criterion for stable, ultrashort pulse propagation in attenuating or amplifying media based on an integral quantity—namely, the pulse area. This phenomenon can be preserved to a great extent in solid-state systems—e.g., semiconductors—and has been experimentally demonstrated by picosecond pulse propagation in QD waveguides [19].

When the laser pulse's temporal width becomes comparable with the optical period (i.e., when the pulse envelope contains only a few optical cycles of the carrier wave), a transition to a qualitatively new "extreme nonlinear" regime of strong light-matter interactions is induced in which the electric field itself, rather than intensity envelope, drives the interaction. The majority of the current analytical approaches are based on the standard, slowly varying amplitude approximation (SVEA) and rotating-wave approximation (RWA). However, a number of works have demonstrated the limitations of these approximations, and new phenomena have been predicted on the basis of a non-perturbative approach (see, e.g., [20], and references therein). This requires development of new, non-perturbative theoretical and computational methods beyond the above approximations as shown in Section 49.2, which could account properly for the ultrafast carrier-wave dynamics.

The earlier high-intensity (i.e., high photon density) limit allows to achieve the strong-coupling regime in free space. However, in the opposite, low-intensity (i.e., few-photon) limit, the light-matter coupling of a single photon to a single atom in free space is weak, and despite attempts by several groups [21–23] to improve it, the experimental realization of high-fidelity excited-state preparation remains challenging. In order to improve the performance of single-photon sources, it is necessary to incorporate the emitters in cavities. The cavity helps to select a single mode out of the infinite number of radiation field modes coupled to the emitter. As the emission is preferentially directed into a single mode, owing to the Purcell effect the spontaneous emission lifetime is reduced for emitters located at the antinode of the electric-field intensity in cavities with a small mode volume and high quality factor. Faster emission is desired because it makes the single-photon source more deterministic and reliable.

If the exciton-light coupling in a semiconductor microcavity is weak compared to all dissipative rates in the system, irreversible decay occurs—that is, energy cannot be alternately passed from light to matter. Depending on whether the quantum wells inside the semiconductor microcavity are situated near the node or antinode of the excitatory electric field, either an inhibited or enhanced spontaneous emission takes place. The cavity then operates in the weak coupling regime, whereby the dynamics of the light-matter interaction is dominated by incoherent processes and damping. Conversely, in the strong coupling limit, no irreversible decay takes place. Instead, a sequence of coherent reabsorption and reemission processes occurs and the energy oscillates between the exciton and photon modes, a phenomenon known as optical Rabi oscillations. The spectral signature of strong coupling is an anticrossing between the dot line (i.e., the spectral linewidth of the QD) and the cavity mode. The eigenmodes of the system are mixed entangled light-matter states, giving rise to quasiparticles known as microcavity exciton–polaritons which exhibit bosonic properties. Considerable scientific effort worldwide has been invested in exploration and control of a strongly coupled single QD in a cavity. The latter is considered a solid-state analogue of the atom-cavity system for which many quantum-optical effects have been demonstrated. Some examples include cavity field quantization [24] and photon blockade [25]. Embedding QDs in confined geometries—such as planar [26], micropillar [27], microdisk [28], microsphere [29], nanowire [30], or photonic crystal nanocavities [31]—provides a means for controlling the light-matter interactions. However, unlike isolated atoms, QDs are embedded in a solid-state environment, so confined charge carriers may experience additional interactions with phonons and other neighboring states and charges. These interactions may result in a departure from the behavior of an idealized two-level quantum system, but can also be exploited as additional control measures.

Cavity quantum electrodynamics (CQED), traditionally applied to atom-cavity systems, has gained new momentum after recent reports of quantum-optical experiments in solid-state systems demonstrating nonclassical light behavior, such as nonclassical photon statistics, nonlocality, and quantum entanglement [32,33]. These cutting-edge studies put to the test the very foundations of quantum mechanics and open up new avenues for on-demand generation and exploitation of various nonclassical states of light. So far, close to perfect single-photon interaction with a QD has been achieved in a cavity under the strong coupling regime [34,35].

A major goal of the simulation of quantum-photonic devices is to find an optimum design for combining photonic and solid-state devices on a chip. Integrated optical technologies for quantum computing based on linear optics [36] have been under active development in the recent decade. However, nonlinear optical interactions "on a chip" offer new, unexplored functionalities, and may play a key role in future technologies. In this respect, a major theoretical and computational challenge is the development of new, tractable models of the nonlinear quantum dynamics in realistic devices with complex geometries in multiple dimensions. This chapter reviews the evolution of the original methodology into our fully non-perturbative, mesoscopic, Maxwell's curl-pseudospin equation-based model of the optically induced spatiotemporal dynamics in quantum-photonic devices. This chapter contains further details about its numerical implementation and some concrete applications.

As we shall show in Section 49.2, our theoretical and computational approach differs substantially from the methods described in Chapter 46 of this book. In that chapter, a number of different approaches for calculation of the spontaneous emission have been employed. In the weak-coupling regime, a classical dipole model is developed whereby the dyadic Green's function is calculated and the spontaneous emission rate is found from Fermi's Golden Rule. Analytical expressions for Green's function are obtainable only for a limited number of device geometries, and more generally one has to resort to numerical methods (e.g., the finite-element method) to compute the electric field. This entails expanding the fields in the eigenmodes in different layers of the specific geometry under consideration. Another approach for modeling cavity-emitter systems is CQED, which is based on the Jaynes–Cummings Hamiltonian, describing the interaction of a two-level system strongly coupled to a single mode of an electromagnetic field. Two approximations are made: (1) for a very low Q cavity, the optical field operators can be replaced with material excitation operators; the dissipation in the system (spontaneous emission rate and dephasing) is taken into account through Lindblad operators. Using the quantum regression theorem that is valid for Markovian processes, an analytical expression for the correlation function is obtained; (2) for a coupled cavity-emitter system, a more complicated master equation is written with an additional dissipation term describing the cavity photon loss. However, for losses weaker than the cavity-emitter coupling, the cavity can be adiabatically eliminated and this results in a master equation describing the emitter only but with modified spontaneous emission rates. By contrast, we use the full master equation for the system density matrix which is valid for all regimes: weak, intermediate, and strong coupling. In addition, in our case the longitudinal and transverse dissipation are taken into account not through Lindblad operators, but by considering the dynamics of population transfer and decoherence between the discrete levels. Another difference between our model presented in Section 49.2 and the model of Gregersen et al. is that in our case, the cavity-emitter (Rabi) coupling is not a model parameter, but is calculated from the initial pulse amplitude, duration, and dipole moment of the specific optical transition, as in the actual experiment. Furthermore, the cavity loss is not a phenomenological parameter either, as it is calculated numerically through the perfectly transparent boundary conditions of the specific cavity geometry. The master equation is solved self-consistently with the vector Maxwell equations for the pulse propagation.

Our approach, predominantly using quantum-optical description based on the Liouville equation of motion for the density matrix, is quite distinctive from the many-body theoretical methods applied in Chapter 48. For instance, the developed by the aforementioned authors quantum-optical semiconductor laser theory beyond the semiclassical approach is based on a correlation-expansion and truncation technique for the quantized light field interacting with carriers in the single-particle picture (electrons and holes) occupying the discrete QD levels. Dynamical equations for the carrier populations in the conduction and valence states of the QD, for the expectation value of the photon number operator of the laser mode and the photon-assisted polarization are derived using Heisenberg operator equations of motion for the creation and annihilation operators. These coupled equations contain carrier–photon and photon–photon correlations, providing access to the second-order correlation function, $g^{(2)}$. The interaction with the environment is accounted through quantum kinetic equations, beyond the Boltzmann equation approach, for the carrier dynamics due to carrier–phonon and carrier-Coulomb interactions. Many-body theory is applied to calculate the corresponding matrix elements. In the former case, effects such as polaron scattering and non-Markovian dynamics are shown for QD systems. This is a very thorough many-body quantized electromagnetic field approach; however, it does not take into account the specific macroscopic boundary conditions of the particular device geometry considered, as the light-matter interaction is described by a single laser mode coupled to the quantum system. An extension for multiple emitters coupled to a single common light mode is proposed, taking advantage of the many-particle picture description of the QD states, such as excitons, trions, multiexcitons, etc., to describe interemitter correlations and radiative coupling. This is somewhat similar to our approach in that it considers many-particle states (e.g., excitons); however, our approach is extended in that it also considers multiple cavity electromagnetic field modes. The cluster expansion approach in Chapter 48 by Gies et al. is in essence a generation of equations of motion for the photon correlations up to a given order of truncation of the expansion and thus providing access to $g^{(2)}$. The dissipation phenomena within this approach—such as the carrier–phonon and carrier–carrier Coulomb scattering—are calculated from a microscopic many-body model (compared to the Lindblad terms and our approach, where they are taken as phenomenological parameters), which is an advantage, as the theory is self-consistent.

In Section 49.2, we will give an overview of the strategy we have used to construct a model for describing coherent light-matter interactions in semiconductor nanostructures.

49.2 Maxwell-Pseudospin Method

49.2.1 Pseudospin Equation

In what follows, we will consider either resonant or near-resonant laser optical field interactions with atom-like discrete-level quantum systems, such as QDs. The optical field we consider is monochromatic and (near-)resonant with a given transition between a pair of atomic levels, thus effectively mimicking a two-level atom. As the two-level atom is conceptually equivalent to a half-spin particle in a magnetic field, the dynamics of the laser-atom interaction is governed by the same Bloch spin vector equation of motion [37] developed for describing nuclear magnetic resonance and the precession of a classical gyromagnet (or its quantum analogue, the spin vector, **S**) in a constant magnetic field, **B**:

$$\frac{d\mathbf{S}}{dt} = f\mathbf{B} \times \mathbf{S}. \tag{49.1}$$

Here f guarantees the constant length of the spin vector, thus confining the spin dynamics to the surface of a Bloch sphere. (Note, however, that this equation is valid only for systems with equally spaced energy levels). The formal analogy between a spin and a nonspin (pseudospin) system was pointed out for the first time by Feynman and coworkers [38], showing that when coherent processes are involved in a two-level system, it is sufficient to consider a real three-vector, rather than the complex probability amplitudes in the Schrödinger equation or the complex density-matrix elements [37]. This real-vector representation provides an elegant and intuitive geometrical framework for comprehending the system dynamics in terms of rotations of a real state vector in the Hilbert space. Furthermore, in contrast to the wavefunction formalism, it describes both pure and mixed quantum states. Attempts to extend the formalism analogously for (N > 2)-level quantum systems fail to preserve the simple vector form of the basic equation of motion for the real state pseudospin (coherence, or Bloch) vector. The only known exception to this is the special case of excitation of a spin-J system in constant magnetic field (thus yielding a $(N = 2J + 1)$-level system), due to the equal spacing of the energy levels in this system [39]. Preserving this simple form of the vector equation has a number of advantages: (1) it accounts for the intrinsic symmetry of the underlying Hilbert space of the system and therefore is an exact description, independent of the strength, number, or time dependence of the external forces acting on the system; (2) similar to the two-level system, the dynamical evolution can be illustrated as a rotation in the real physical space of a real coherence vector.

The case of a three-level system is of particular interest, since it provides a useful framework for studying such phenomena as two-photon coherence, resonance Raman scattering, three-level echoes, three-level superradiance, coherent multistep photoionization/photodissociation, coherent population trapping, and electromagnetically induced transparency (EIT). A solution to this problem has been given by Elgin [40] by invoking the invariance of the state vector under rotations of the SU(3) transformation group. The dynamical evolution of a three-level system can be expressed in terms of an eight-dimensional real coherence vector, taking advantage of the group-theory and Gell-Mann's SU(3) generators, which had been developed for the quark triplet in the domain of high-energy physics [41].

A general solution for an N-level quantum system with arbitrary level spacing has been provided by Hioe and Eberly [42]. The starting point is the dynamical evolution of a quantum system described by the Liouville equation for the density operator in the Schrödinger picture

$$i\hbar\frac{\partial\hat{\rho}}{\partial t} = \left[\hat{H}, \hat{\rho}\right], \tag{49.2}$$

where $\hat{H}$ is the total system Hamiltonian:

$$\hat{H} = \hat{H}_0 + \hat{H}_{\text{int}}(t), \tag{49.3}$$

with $\hat{H}_0$ being the unperturbed Hamiltonian of an N-level system: a diagonal matrix with eigenenergies of each level, $\hbar\omega_k$ $(k = 1, \ldots, N)$, along the main diagonal, and $\hat{H}_{int}$ is a time-dependent perturbation (not necessarily small).

Hioe and Eberly showed that the presence of unitary group generators in the time evolution of an N-level quantum system permits to describe the evolution in terms of the rotations of a real coherence vector. Expanding the system Hamiltonian and the density matrix operators in terms of λ-generators of the SU(N) Lie algebra, $\hat{\lambda}_k$, they derived the pseudospin equation of motion in the Heisenberg picture, depicting the time evolution of the real pseudospin vector, **S**, as a generalized rotation in $N^2 - 1$ Hilbert space:

$$\dot{S}_i = f_{ijk}\gamma_j S_k, \;\; i, j, k = 1, 2, \ldots, N^2 - 1, \tag{49.4}$$

where the dot stands for the time derivative, γ_j is the torque vector, and $f_{ijk} = \frac{1}{4}i\left[Tr\left(\hat{\lambda}_i\hat{\lambda}_k\hat{\lambda}_j\right) - Tr\left(\hat{\lambda}_i\hat{\lambda}_j\hat{\lambda}_k\right)\right]$, $i, j, k = 1, 2, \ldots, N^2 - 1$, is the fully antisymmetric tensor of the structure constants of the SU(N) group, whose asymmetry guarantees the constant length of the pseudospin vector, $|\mathbf{S}|$, in the $(N^2 - 1)$-dimensional Hilbert space. The coherence vector is expressed as the trace of the product of the density and the SU(N) group generators, whereas the torque vector is expressed as the trace of the product of the system Hamiltonian operators and the SU(N) group generators

$$S_j(t) = Tr\left(\hat{\rho}(t).\hat{\lambda}_j\right); \;\; \gamma_j(t) = \frac{1}{\hbar}Tr\left(\hat{H}(t).\hat{\lambda}_j\right). \tag{49.5}$$

The SU(N) group generators obey the group property

$$\left[\hat{\lambda}_j, \hat{\lambda}_k\right] = 2if_{jkl}\hat{\lambda}_l \tag{49.6}$$

and orthogonality relations, $Tr\left(\hat{\lambda}_j.\hat{\lambda}_k\right) = \delta_{jk}$, however their choice is not unique.

For the simplest case of a two-level system, $N = 2$, the λ-generators are simply the Pauli matrices $\hat{\sigma}_x$, $\hat{\sigma}_y$, and $\hat{\sigma}_z$. This is obtained from Equation 49.4 when f_{ijk} is replaced by the Levi-Civita symbol, ε_{ijk}. For a dipole-coupling interaction $\hat{H}_{\rm int} = e\mathbf{E}.\hat{\boldsymbol{r}}$, where **r** is the local displacement operator and e is the electron charge, and a linearly polarized dipole optical transition ($\mathbf{E} \,||\, \mathbf{r}$ with an optical selection rule for the z-axis projection of the total angular momentum, $\Delta J_z = 0$) the optical Bloch equations are easily recovered:

$$\dot{\mathbf{S}} = \boldsymbol{\gamma} \times \mathbf{S} \;\; \begin{aligned} \dot{S}_1 &= \omega_0 S_2 \\ \dot{S}_2 &= -\omega_0 S_1 + 2\Omega_R S_3 \\ \dot{S}_3 &= -2\Omega_R S_2, \end{aligned} \tag{49.7}$$

where S_i $(i = 1, 2, 3)$ is the real pseudospin three-vector, and the torque vector is given by $\boldsymbol{\gamma} = \left(2\Omega_{\rm R}, 0, \omega_0\right)$ from Equation 49.5. Here we have defined the Rabi frequency, $\Omega_{\rm R} = \frac{\wp}{\hbar}E$, with $\wp = \langle i|e\hat{r}|j\rangle$, transition dipole matrix element of $i \rightarrow j$, and ω_0 is the atomic transition resonant frequency. In case of near-resonant excitation, ω_0 should be replaced by $\Delta\omega = \omega_0 - \omega$, where ω is the excitation frequency [43].

In the general case of an electric field vector noncollinear with the local displacement (dipole moment) vector, the optical transition is excited by circularly or elliptically polarized light obeying the selection rule $\Delta J_z = \pm 1$, where the sign "+/–" corresponds to right/left circularly polarized light. The dipole matrix elements are complex vectors yielding complex Rabi frequencies along the x- and y-direction in the plane determined by **E** and **r** (or equivalently, along the dipole moment vector): $\Omega_x = \frac{\wp}{\hbar}E_x$; $\Omega_y = \frac{\wp}{\hbar}E_y$. The torque vector is given by $\boldsymbol{\gamma} = \left(\Omega_x, \Omega_y, \omega_0\right)$ and the resulting pseudospin equations read (see also [37])

$$\dot{\mathbf{S}} = \boldsymbol{\gamma} \times \mathbf{S} \quad \begin{aligned} \dot{S}_1 &= -\omega_0 S_2 - \Omega_y S_3 \\ \dot{S}_2 &= \omega_0 S_1 + \Omega_x S_3 \\ \dot{S}_3 &= \Omega_y S_1 - \Omega_x S_2 \,. \end{aligned} \tag{49.8}$$

The relationship between the real-state coherence vector components and the density-matrix components for the general N-level case is given by Equation 49.5, and for a two-level system reads

$$\begin{aligned} S_1 &= \hat{\rho}_{12} + \hat{\rho}_{21} = 2\mathrm{Re}\left(\hat{\rho}_{12}\right) \\ S_2 &= i\left(\hat{\rho}_{12} - \hat{\rho}_{21}\right) = -2\mathrm{Im}\left(\hat{\rho}_{12}\right) = 2\mathrm{Im}\left(\hat{\rho}_{21}\right) \\ S_3 &= \hat{\rho}_{22} - \hat{\rho}_{21} \,. \end{aligned} \tag{49.9}$$

The physical meaning of the first two components of the pseudospin vector are dispersive (in-phase) and absorptive (in-quadrature) parts of the dipole polarization, and the third component is the population difference between the two levels. Since the level occupation probability is conserved, i.e., $\rho_{11} + \rho_{22} = 1$, it can be easily shown that the length of the coherence vector is unity: $|S|^2 = S_1^2 + S_2^2 + S_3^2 = 1$, and the evolution of the pseudospin is "inscribed" by the pseudospin vector on the surface of the Bloch sphere.

Up until now, we have only considered coherent processes; however, in an open quantum system there is always energy dissipation. For a two-level system, it is easy to define phenomenological longitudinal relaxation time (also known as population relaxation time), T_1, and transverse relaxation time (also known as polarization-decay or dephasing time), T_2, leading to the well-known optical Bloch equations for a two-level system with damping for e.g., linearly polarized transitions:

$$\begin{aligned} \frac{\partial S_1}{\partial t} &= \omega_o S_2 - \frac{1}{T_2} S_1 \\ \frac{\partial S_2}{\partial t} &= -\omega_o S_1 + 2\Omega_x S_3 - \frac{1}{T_2} S_2 \\ \frac{\partial S_3}{\partial t} &= -2\Omega_x S_2 - \frac{1}{T_1}\left(S_3 - S_{30}\right), \end{aligned} \tag{49.10}$$

where $S_{30} = \pm 1$ is the initial population profile; "–" and "+" correspond, respectively, to all of the population in the ground and the excited state. Upon optical excitation, the former describes absorption, while the latter describes optical gain. However, it is not so easy to generalize the dissipation dynamics for a system with an arbitrary number of levels (N-level system), because each pair of levels within this system is characterized by their individual population transfer rates and dephasing rates. We show how to do this generalization next. The time evolution of the two-level quantum system is described by the generalized rotation of the real three-vector, **S**, in the three-dimensional Hilbert space; at all times the tip of **S** is just touching the surface of the Bloch sphere, shown in Figure 49.1.

49.2.2 Dissipative Dynamics and Master Pseudospin Equation

To include energy relaxation and decoherence processes in the system dynamics, we adopt an unconventional approach. We introduce longitudinal relaxation in the Liouville equation for the density matrix by considering first the diagonal population components of the Liouville equation, ρ_{ii}, and taking into account population transfer between all allowed dipole optical transitions in the system with corresponding rates.

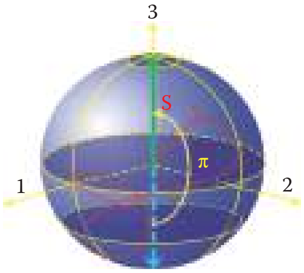


FIGURE 49.1 Geometrical representation of the time evolution of a two-level quantum system by generalized rotation of the real state pseudospin vector, **S**, in the 3D Hilbert space, drawing a trajectory on the $|\mathbf{S}| = const.$ sphere surface. A one-qubit rotation occurs when the vector is rotated through π from state 0 (spin down, dashed arrow) to state 1 (spin up, solid arrow along $\hat{3}$-axis), or vice versa.

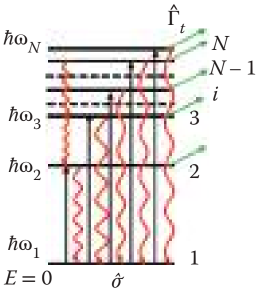


FIGURE 49.2 Schematic of an N-level quantum system with energy zero at the ground state. [Vertical solid arrows, coherent optical excitation; wavy lines, population relaxation $\hat{\sigma}$ (spontaneous emission); tilted arrows, dephasing $\hat{\Gamma}_t$ (decoherence).]

For each Liouville equation we can construct a longitudinal damping matrix of rank N: $\hat{\Gamma}_i$, $i = 1, \ldots, N$, whose elements are the population transfer rates between all pairs of levels. This allows us to introduce a diagonal matrix, $\hat{\sigma} = diag(Tr(\hat{\Gamma}_i.\hat{\rho}))$, which accounts for the longitudinal population relaxation. Similarly, we introduce an off-diagonal matrix, $\hat{\Gamma}_t$, of rank N with zero main diagonal components to account for the dephasing rates of all levels (Figure 49.2). Adding the relaxation and dephasing terms in the Liouville equation thus leads to the following master equation:

$$\frac{\partial\hat{\rho}}{\partial t} = \frac{1}{\hbar}\left[\hat{\rho}, \hat{H}\right] + \hat{\sigma} - \hat{\Gamma}_t.\hat{\rho}. \tag{49.11}$$

Using Equation 49.5, we can now rewrite the density-matrix equations in terms of the pseudospin vector, **S** [44]:

$$\frac{\partial S_j}{\partial t} = \begin{cases} f_{jkl}\Gamma_k S_l + \frac{1}{2}Tr\left(\hat{\sigma}.\hat{\lambda}_j\right) - \frac{1}{T_j}\left(S_j - S_{je}\right), & j = 1, 2, \ldots N\,(\mathrm{N}-1) \\ f_{jkl}\Gamma_k S_l + \frac{1}{2}Tr\left(\hat{\sigma}.\hat{\lambda}_j\right), & j = N\,(N-1)+1, \ldots, N^2-1, \end{cases} \tag{49.12}$$

where we have defined an equilibrium coherence vector, $\mathbf{S_E} = (S_{1e}, S_{2e}, \ldots, S_{(N^2-1)e})$, and $T_j, j = 1, 2\ldots,$ N^2-1 are phenomenologically introduced nonuniform decay times describing the relaxation of the components of the real state vector towards their equilibrium values, S_{je}. Because of the incoherent processes involved in maintaining the system at a definite level of excitation, the equilibrium system polarization is zero. This means that the subset of $\mathbf{S}_\mathbf{E}$ containing the polarization components of the equilibrium coherence vector is zero, i.e., $\mathbf{S}_\mathbf{E}^* = (S_{1e}, S_{2e}, \ldots, S_{N(N-1)e}) = 0$. The second term in the second line of the Equation 49.12 containing the trace of the density matrix through $\hat{\sigma}$ can be rewritten in terms of nonuniform relaxation times, $T_j = T_{N(N-1)+1}, \ldots, T_{N^2-1}$, describing the relaxation of the population components of the real coherence vector toward their equilibrium values. Therefore, the total equilibrium coherence vector is of the form: $S_E = (0, 0, \ldots, 0, S_{N(N-1)+1}, \ldots, S_{N^2-1})$. We shall refer to Equation 49.12 as the master pseudospin equations.

49.2.3 Vector Maxwell Equations and Coupling to Pseudospin Equations

The optical wave propagation is described by the vector Maxwell equations:

$$\begin{aligned} \frac{\partial \mathbf{H}}{\partial t} &= -\frac{1}{\mu}\nabla \times \mathbf{E} \\ \frac{\partial \mathbf{E}}{\partial t} &= \frac{1}{\varepsilon}\nabla \times \mathbf{H} - \frac{1}{\varepsilon}\frac{\partial \mathbf{P}}{\partial t}, \end{aligned} \tag{49.13}$$

where **P** is the macroscopic medium polarization vector given by

$$\mathbf{P} = -eN_a Tr\left(\hat{\rho}.\hat{\boldsymbol{r}}\right), \tag{49.14}$$

with N_a signifying the density of the ensemble of resonant dipoles in the medium. Starting with a phenomenologically constructed dipole-coupling interaction Hamiltonian with coherent Rabi frequency terms describing optical transitions between pairs of levels, one can decompose the dipole moment operator, and represent it as a linear combination of λ-generators, whose coefficients are the corresponding Cartesian dipole moment vector components. Then, using Equation 49.14, we can find a relationship between the macroscopic polarization components, $P_{x,y,z}$, and the coherence vector components, S_j. To illustrate this procedure, suppose we have a transverse electromagnetic wave propagating along the z-axis exciting a given $i \rightarrow j\,(i < j)$ transition in an N-level system. Let the wave be circularly polarized in the plane perpendicular to z, with in-plane electric-field components E_x and E_y. Then, the (i, j) term of the interaction Hamiltonian will contain a complex coherent term $\hbar(\Omega_x \pm i\Omega_y)$, where the sign "+/−" depends on the helicity of the pulse, σ^+ or σ^-. (The sign convention is such that "+" and "−" correspond to, respectively, clockwise and counterclockwise rotation of the electric-field vector, when viewed from a direction opposite to that of the propagation of the pulse.) Note that, for the case of a linearly polarized excitation, the Rabi frequency will be real. The in-plane dipole moment vector can be decomposed into two Cartesian components: $e\mathbf{r} = e(\hat{r}_x\vec{e}_x + \hat{r}_y\vec{e}_y)$, where $\vec{e}_x$ and $\vec{e}_y$ are the unit vectors along the x- and y-axes. The coefficients $\hat{r}_x$ and $\hat{r}_y$ are $\hat{\lambda}$-generators; $\hat{r}_x$ contains unity at positions (i, j) and (j, i), whereas $\hat{r}_y$ contains the imaginary number, $-i$, at position (i, j) and i at position (j, i):

$$\mathbf{r} = r_0\left\{\begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 1 & \cdot & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & 1 & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 \end{pmatrix}\vec{e}_x + \begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & -i & \cdot & 0 \\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & i & 0 & \cdot & \cdot & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 \end{pmatrix}\vec{e}_y\right\}, \tag{49.15}$$

where r_0 is the dipole length scale. The following general formula can be used to obtain the corresponding $\hat{\lambda}$-generators, depending on the number of discrete levels of the system, N:

$$\begin{aligned}\hat{r}_x &= \hat{\lambda}_{j-i+\sum\limits_{d=1}^{i-1}(\mathrm{N}-d)}\\ \hat{r}_y &= \hat{\lambda}_{j-i+\sum\limits_{d=1}^{i-1}(\mathrm{N}-d)+\frac{\mathrm{N}(\mathrm{N}-1)}{2}}.\end{aligned} \tag{49.16}$$

In fact, this linear combination reproduces the phenomenologically constructed dipole coupling interaction Hamiltonian, $\hat{H}_{int} = -e\mathbf{r}.\mathbf{E}$, when the dipole moment and the decomposed electric-field vector ($\vec{E} = E_x\vec{e}_x + E_y\vec{e}_y$) are substituted in

$$\hat{H}_{\mathrm{int}} = \hbar\begin{pmatrix} 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ 0 & 0 & 0 & \cdot & \cdot & \Omega_x - i\Omega_y & \cdot & 0\\ 0 & 0 & 0 & \cdot & \cdot & \cdot & \cdot & 0\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ 0 & \cdot & \Omega_x + i\Omega_y & 0 & \cdot & \cdot & \cdot & 0\\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot & \cdot\\ 0 & 0 & 0 & \cdot & \cdot & 0 & 0 & 0 \end{pmatrix}. \tag{49.17}$$

The macroscopic polarization vector components, P_x and P_y, are then obtained from Equation 49.14, giving

$$\begin{aligned}P_x &= -\wp N_a Tr(\hat{\rho}.\hat{r}_x)\\ P_y &= -\wp N_a Tr(\hat{\rho}.\hat{r}_y).\end{aligned} \tag{49.18}$$

Now we can write down the vector Maxwell equations for a circularly polarized optical wave exciting the $i \to j$ transition of an N-level system, and thus inducing dipole polarizations P_x and P_y along the x- and y-directions, respectively:

$$\begin{aligned}\frac{\partial H_x(z,t)}{\partial t} &= \frac{1}{\mu}\frac{\partial E_y(z,t)}{\partial z}\\ \frac{\partial H_y(z,t)}{\partial t} &= -\frac{1}{\mu}\frac{\partial E_x(z,t)}{\partial z}\\ \frac{\partial E_x(z,t)}{\partial t} &= -\frac{1}{\varepsilon}\frac{\partial H_y(z,t)}{\partial z} - \frac{1}{\varepsilon}\frac{\partial P_x(z,t)}{\partial t}\\ \frac{\partial E_y(z,t)}{\partial t} &= \frac{1}{\varepsilon}\frac{\partial H_x(z,t)}{\partial z} - \frac{1}{\varepsilon}\frac{\partial P_y(z,t)}{\partial t}.\end{aligned} \tag{49.19}$$

For the special case of a transverse optical wave linearly polarized along x and propagating in the z-direction, the above system (Equation 49.19) is reduced to

$$\begin{aligned}\frac{\partial H_y(z,t)}{\partial t} &= -\frac{1}{\mu}\frac{\partial E_x(z,t)}{\partial z}\\ \frac{\partial E_x(z,t)}{\partial t} &= -\frac{1}{\varepsilon}\frac{\partial H_y(z,t)}{\partial z} - \frac{1}{\varepsilon}\frac{\partial P_x(z,t)}{\partial t}.\end{aligned} \tag{49.20}$$

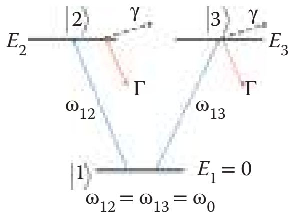


FIGURE 49.3 Energy-level scheme of a V-type three-level system. $\omega_{jk} = (E_j - E_k)/\hbar$, with energy zero chosen at the ground state $|1\rangle$. (Γ, population relaxation rate; γ, dephasing rate.)

We solve Maxwell's curl equations, (Equations 49.19 or 49.20), and the pseudospin equations (Equation 49.12) self-consistently, having established a link between them via the medium's polarization (Equation 49.18). We shall refer to this coupled system of first-order differential equations as the coherent Maxwell's curl-pseudospin model.

49.2.4 Coherent Maxwell-Pseudospin Equations in 2D

Description of nonlinear pulse propagation, interaction, and localization phenomena in multiple dimensions requires dipole coupling to at least a 2D resonant medium, whereby the electric field of a single-photon excitation is coupled to at least two distinct dipole optical transitions. We have previously shown that as a minimum requirement, it is sufficient to consider a degenerate three-level ensemble of dipoles in which two of the allowed dipole optical transitions are excited by each of the two components of the electric-field vector in the waveguide plane [45]. The pseudospin geometrical picture thus allows for adequate modeling of the interaction of an ultrashort laser pulse with a medium in two spatial dimensions. Therefore, we consider a V-type degenerate three-level system of resonant dipoles shown in Figure 49.3, for which a linearly polarized monochromatic electromagnetic wave induces polarizations (dipoles) in two orthogonal directions in the waveguide plane. Two cases arise: a transverse electric (TE) wave, for which $\boldsymbol{E} = (E_x, 0, 0)$ and $\mathbf{H} = (0, H_y, H_z)$, and transverse magnetic (TM) wave, for which $\mathbf{E} = (0, E_y, E_z)$ and $\mathbf{H} = (H_x, 0, 0)$. The respective vector Maxwell equations are

$$\begin{aligned} \frac{\partial H_y}{\partial t} &= -\frac{1}{\mu}\frac{\partial E_x}{\partial z} \\ \frac{\partial H_z}{\partial t} &= \frac{1}{\mu}\frac{\partial E_x}{\partial y} \\ \frac{\partial E_x}{\partial t} &= \frac{1}{\varepsilon}\left(\frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z}\right) - \frac{1}{\varepsilon}\frac{\partial P_x}{\partial t} \end{aligned} \tag{49.21}$$

for the TE wave and

$$\begin{aligned} \frac{\partial H_x}{\partial t} &= -\frac{1}{\mu}\frac{\partial E_z}{\partial y} + \frac{1}{\mu}\frac{\partial E_y}{\partial z} \\ \frac{\partial E_y}{\partial t} &= \frac{1}{\varepsilon}\frac{\partial H_x}{\partial z} - \frac{1}{\varepsilon}\frac{\partial P_y}{\partial t} \\ \frac{\partial E_z}{\partial t} &= -\frac{1}{\varepsilon}\frac{\partial H_x}{\partial y} - \frac{1}{\varepsilon}\frac{\partial P_z}{\partial t} \end{aligned} \tag{49.22}$$

for the TM wave. We can phenomenologically construct the system Hamiltonian corresponding to each of the cases above:

$$\hat{H}_{TE}(t)=\begin{pmatrix}0 & \Omega_x & 0\\ \Omega_x & \omega_0 & 0\\ 0 & 0 & \omega_0\end{pmatrix};\ \hat{H}_{\mathrm{TM}}(t)=\begin{pmatrix}0 & \Omega_y & \Omega_z\\ \Omega_y & \omega_0 & 0\\ \Omega_z & 0 & \omega_0\end{pmatrix}. \tag{49.23}$$

The master pseudospin equations (Equation 49.12) for a three-level system with $i, j, k = 1, 2, ..., 8$ are coupled to Maxwell's equations above through the macroscopic polarization, Equation 49.14. It thus becomes possible to derive the following relationships between the polarization components and the pseudospin vector components for the case of a TE wave:

$$P_x = -\wp N_a S_1, \tag{49.24}$$

and for a TM wave:

$$\begin{aligned} P_y &= -\wp N_a S_1\\ P_z &= -\wp N_a S_3. \end{aligned} \tag{49.25}$$

49.2.5 Numerical Methodology

The Maxwell-pseudospin system is discretized on a Yee grid in 1D (modified Yee grid in 2D, see Figure 49.4) and solved directly in the time domain by the finite-difference time-domain (FDTD) method [46]. At each time step, a solution to all equations is obtained using a predictor-corrector iterative scheme, proven to be very efficient in solving simultaneously a large number of first-order differential equations [47]. The time evolution of a discrete multilevel system under an external perturbation can be viewed as a Goursat-type initial boundary value problem, which is well posed when the time history of the initial electric field is given along, e.g., the left (lower) boundary of the simulation domain in 1D (2D). We impose analytical absorbing boundary conditions at the simulation domain boundaries, based on the Engquist–Majda [48] one-way wave equations discretized by the Mur finite-difference scheme [49] using a two-term Taylor-series approximation [46]. To ensure numerical stability, the time step and the spatial discretization are chosen to satisfy the Courant stability criterion: $\frac{c\Delta t}{\Delta z} \leq 1$ in 1D and $\frac{c\Delta t}{\left(\frac{1}{(\Delta y)^2}+\frac{1}{(\Delta z)^2}\right)^{-1/2}} \leq 1$ in 2D [46]. The FDTD numerical technique has numerous advantages over other methods, such as the beam propagation method (BPM). For instance, modeling structures with rapid longitudinal variations (on the scale of the wavelength, λ) of, e.g., refractive index or energy-level population, is impossible to model with BPM, because it relies on the paraxial approximation which in this case is invalid. Moreover, FDTD can handle strong coupling between forward- and backward-propagating fields and can be used to describe general, nonharmonic electromagnetic fields. It can also be used to simulate ultrashort (sub-picosecond) pulses in situations where the pulse bandwidth is significant compared to material resonances.

49.3 Simulation of Nonlinear Dynamics in Optical Waveguides and Semiconductor Microcavities

In this section, we apply the vector Maxwell-pseudospin formalism to the problem of nonlinear optical pulse propagation in planar waveguides and microcavities with embedded quantum systems. We provide a solution in both 1D and 2D. As a first step, we perform validation studies against well-established theoretical and computational methods.

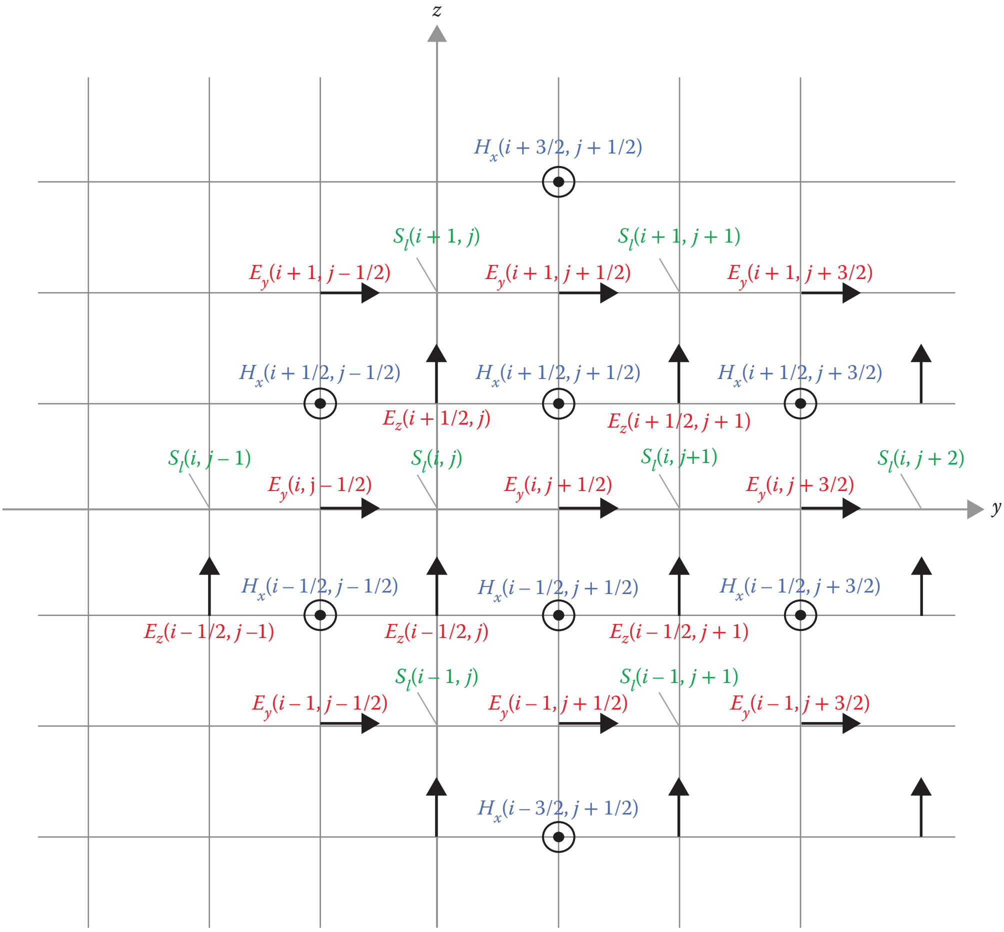

z
$H_x(i + 3/2, j + 1/2)$
$S_l(i + 1, j)$
$S_l(i + 1, j + 1)$
$E_y(i + 1, j - 1/2)$
$E_y(i + 1, j + 1/2)$
$E_y(i + 1, j + 3/2)$
$H_x(i + 1/2, j - 1/2)$
$H_x(i + 1/2, j + 1/2)$
$H_x(i + 1/2, j + 3/2)$
$E_z(i + 1/2, j)$
$E_z(i + 1/2, j + 1)$
$S_l(i, j - 1)$
$S_l(i, j)$
$S_l(i, j+1)$
$S_l(i, j + 2)$
$E_y(i, j - 1/2)$
$E_y(i, j + 1/2)$
$E_y(i, j + 3/2)$
y
$H_x(i - 1/2, j - 1/2)$
$H_x(i - 1/2, j + 1/2)$
$H_x(i - 1/2, j + 3/2)$
$E_z(i - 1/2, j -1)$
$E_z(i - 1/2, j)$
$E_z(i - 1/2, j + 1)$
$S_l(i - 1, j)$
$S_l(i - 1, j + 1)$
$E_y(i - 1, j - 1/2)$
$E_y(i - 1, j + 1/2)$
$E_y(i - 1, j + 3/2)$
$H_x(i - 3/2, j + 1/2)$

FIGURE 49.4 Modified Yee grid for discretization of the 2D Maxwell-pseudospin system for the purpose of carrying out numerical modeling. The grid is staggered, meaning that the electric and magnetic fields are calculated at every other node. This schematic is for a TM wave: the electric-field vector is in the y–z plane (arrows); the magnetic-field vector is pointing out of the page toward the reader (circle with dot). The electric- and magnetic-field components are spatially separated by $\Delta y/2$ and $\Delta z/2$ in the y-z plane, and temporally by $\Delta t/2$, with the electric-field components situated in the middle of the cells' edges, and magnetic-field components in the center of the cell [46]. The magnetic-field equation is solved at the spatial steps $(i + 1/2)\Delta z$ and $(j + 1/2)\Delta y$ over time step $(k + 1/2)\Delta t$. The electric-field components, E_y and E_z, are solved at the spatial steps $((j + 1/2)\Delta y, i\Delta z)$ and $(j\Delta y, (i + 1/2)\Delta z)$ over time step $k\Delta t$. Quantum-system variables, S_l, are assigned to the empty nodes in the 2D Yee grid, while the electric-field component values at these empty mesh points are estimated by averaging over the electric fields located at the nearest neighbors in the plane.

49.3.1 One-Dimensional Model of a Two-Level System

Consider a two-level quantum system, such as fundamental heavy-hole transitions in a QD. The coherent Maxwell-pseudospin equations describing the interaction of an optical wave propagating along the z-axis and linearly polarized in a transverse direction along x are given by Equations 49.10, 49.20, and 49.24:

$$\frac{\partial H_y}{\partial t} = -\frac{1}{\mu}\frac{\partial E_x}{\partial z}$$

$$\begin{aligned}
\frac{\partial E_x}{\partial t} &= -\frac{1}{\varepsilon}\frac{\partial H_y}{\partial z} - \frac{\wp N_a}{\varepsilon T_2}S_1 + \frac{\wp N_a \omega_0}{\varepsilon}S_2 \\
\frac{\partial S_1}{\partial t} &= -\frac{1}{T_2}S_1 + \omega_0 S_2 \\
\frac{\partial S_2}{\partial t} &= -\omega_0 S_1 - \frac{1}{T_2}S_2 + 2\frac{\wp}{\hbar}E_x S_3 \\
\frac{\partial S_3}{\partial t} &= -2\frac{\wp}{\hbar}E_x S_2 - \frac{1}{T_1}\left(S_3 - S_{30}\right).
\end{aligned} \tag{49.26}$$

We apply an excitation pulse of duration T_{p} with either a hyperbolic secant or Gaussian envelope and central frequency, ω_0, in resonance with the two-level transition at the boundary $z = 0$ of the simulation domain. Its time dependence is thus

$$E_x\left(z = 0, t\right) = \begin{cases} E_0 \operatorname{sech}\left(10\frac{t-\frac{T_{\mathrm{p}}}{2}}{\frac{T_{\mathrm{p}}}{2}}\right)\sin\left(\omega_0 t\right) \\ E_0 \exp\left[\left(\frac{t-t_0}{t_d}\right)^2\right]\sin\left(\omega_0 t\right), \end{cases} \tag{49.27}$$

where E_0 is the initial pulse amplitude, t_0 is the time moment at which the Gaussian pulse is initially centered, and t_d is the characteristic $1/e$ Gaussian decay.

49.3.2 Passive Cavity Properties

To validate our model, we construct a simple, symmetric test Bragg microcavity containing 5 $\lambda/4$ GaAs/$\mathrm{Al_{0.1}Ga_{0.9}As}$/AlAs layers, with λ being 1.4 μm (Figure 49.5f) and compute the passive (cold-)cavity modes and the electric-field standing-wave profile within the cavity. In order to compute the resonant cavity modes, we launch a short broad-band pulse of duration $\leq$ 100 fs from the left boundary of the structure and sample in time the electric-field evolution at the output on the right. This time trace is Fourier-transformed and normalized with respect to the Fourier transform of the input trace, thus yielding the transmission spectrum of the microcavity. The resonant frequency is extracted from the peak in the stop band of the spectrum (Figure 49.5d). Note that reducing the time step allows to resolve higher-order cavity modes. To obtain the electric-field standing-wave profile within the cavity, a continuous sine wave at the resonance frequency (obtained in the previous step) is pumped from the left boundary, and the time evolution of the intracavity field is monitored. The field amplitude increases in time and eventually reaches a steady value. The results shown in Figure 49.5f are compared with the cavity modes and field distributions computed by the transfer matrix method (TMM).

Using the above FDTD-based methodology, we will subsequently investigate passive cavity properties of more complex cavities. For example, we calculate the resonant cavity modes and standing-wave profile of a Coldren-type microcavity structure designed at $\lambda = 1.3$ μm and consisting of 10 GaAs/$\mathrm{Al_{0.9}Ga_{0.1}As}$ pairs for the bottom distributed Bragg reflector (DBR) mirror, followed by an AlAs supply layer, a tuning $\mathrm{Al_{0.7}Ga_{0.3}As}$ layer, a GaAs $\lambda/2$ cavity and 15 GaAs/$\mathrm{Al_{0.9}Ga_{0.1}As}$ pairs for the top DBR mirror. The simulation results for the resonant cavity modes, namely the transmission spectrum as a function of λ and the electric-field standing-wave profile along the structure, are displayed in Figure 49.6.

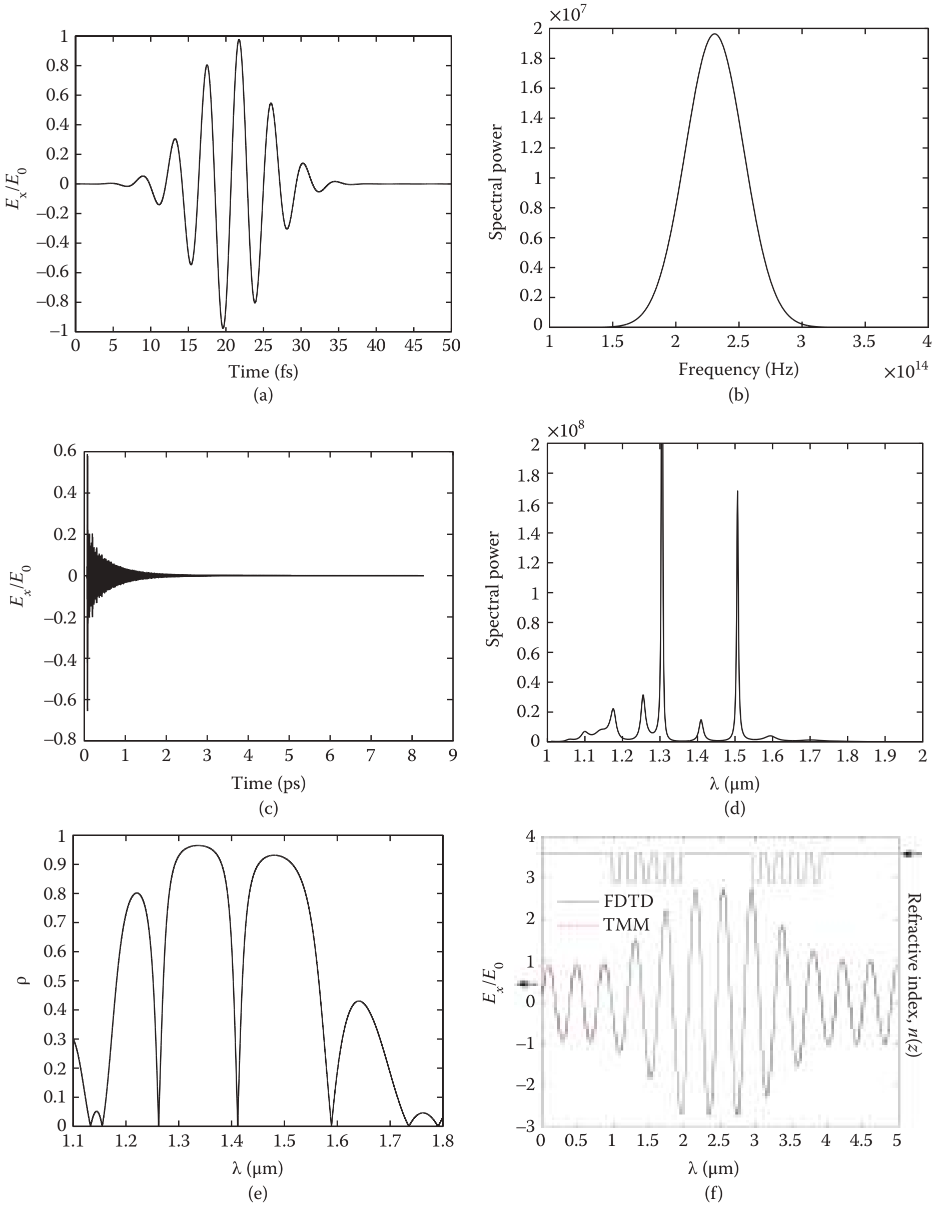
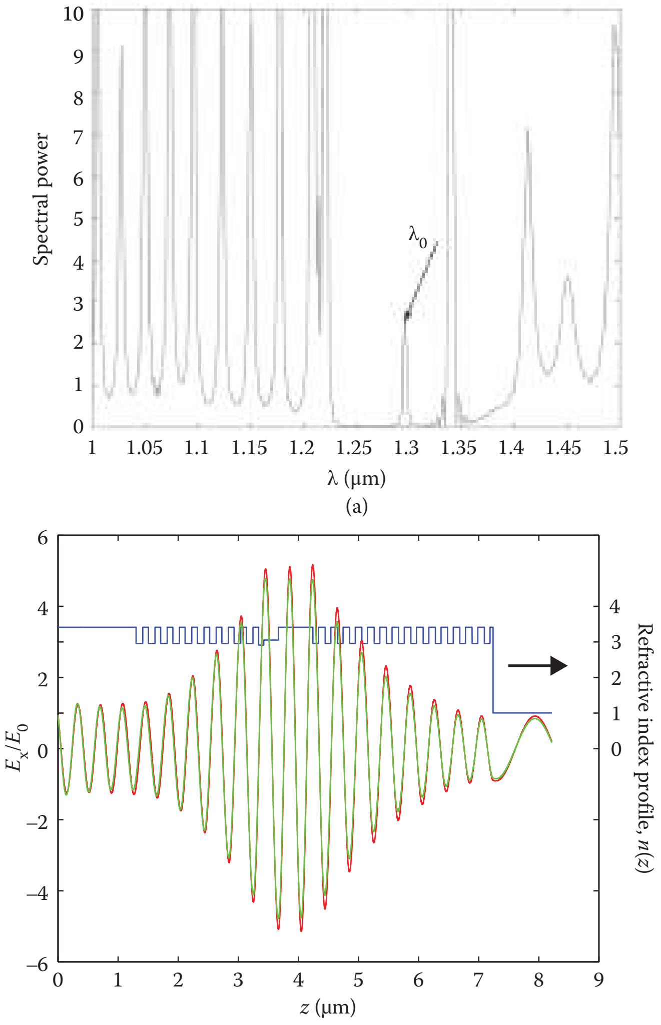


FIGURE 49.5 (a) Initial ultrashort pulse ($T_p \leq 100$ fs). (b) Fourier spectrum of the initial broad-band pulse. (c) Time trace of $E_x(t)$ at the output boundary of the structure. (d) Transmission spectrum as a function of wavelength, λ, computed as the discrete Fast Fourier transform (FFT) of the output time trace in (c), normalized with respect to the input FFT spectrum (b). The resonant cavity wavelength, $\lambda_0 \approx 1.4\,\mu$m, is extracted from the peak in the DBR cavity stop band. (e) Reflectivity spectrum as a function of wavelength computed by the TMM: the resonant cavity wavelength, $\lambda_0 \approx 1.4\,\mu$m is obtained from the dip in the reflectivity spectrum within the DBR cavity stop band. (f) Comparison of the standing-wave profile within the cavity, as obtained by the TMM (dashed curve) and FDTD (solid black curve) methods. The FDTD-inferred instantaneous field standing-wave profile is computed with continuous sine-wave excitation propagating from the left boundary at the cavity resonance frequency equivalent to $\lambda_{\text{pump}} = 1.4116\,\mu$m, obtained from (d) after a sufficiently large number of time steps 2×10^6 (each of duration $\Delta t = 3.33 \times 10^{-3}$ fs) with 31,300 spatial grid points (each spatial step of $\Delta z = 1$ nm), Courant number= 0.75. The scale on the right corresponds to the refractive-index profile of the DBR cavity (top).



(a)

(b)

FIGURE 49.6 (a) FFT of the time trace of the signal detected at the output boundary of a Coldren-type microcavity structure, revealing the cavity's (nonnormalized) transmission spectrum. The resonant cavity mode is indicated with λ_0, and corresponds to approximately 1.3 μm. (b) Standing-wave profile after 6 million (dashed curve) and 12 million (solid curve) time steps superimposed on the refractive-index profile of the microcavity (scale on right-hand side). The time step in the simulation was $\Delta t = 3.33 \times 10^{-3}$ fs and the spatial step was $\Delta z = 1$ nm. The amplitude of the standing wave increases as the number of time steps is increased, and eventually reaches a steady maximum value (solid curve). The amplitude is augmented with time due to the build-up of the field caused by multiple reflections at the cavity mirrors.

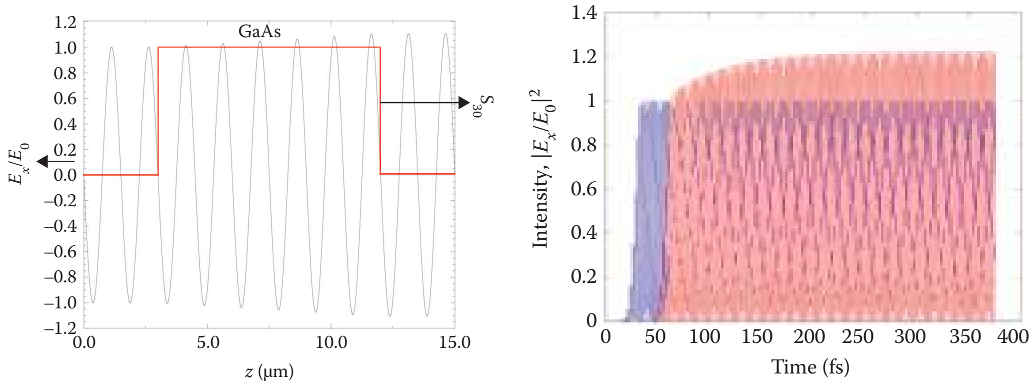
49.3.3 Active Nonlinear Optical Waveguide/Cavity Properties

We have performed gain validation studies of our 1D semiclassical Maxwell-pseudospin model. This was done by establishing the ability of our model to recover the linear small-signal (unsaturated) gain and nonlinear (saturated) gain of a homogeneously broadened (HB) two-level system, and comparing the result against that of the density-matrix theory, which served as a benchmark (see, e.g., [50]). The test structure used for these validation studies was a 15 μm-long simulation domain consisting of a GaAs slab waveguide with a gain region length of $L_g = 9\ \mu\text{m}$, refractive index $n_{\text{GaAs}} = 3.59$, and pumping wavelength of 1.5 µm corresponding to the resonant transition frequency of a two-level system sandwiched between two free-space regions (Figure 49.7a). The GaAs slab was represented by an initially inverted ($S_{30} = 1$) two-level medium with resonant dipole density $N_a = 1 \times 10^{24}\ \text{m}^{-3}$, dipole transition matrix element, $\wp = 1 \times 10^{-29}\ \text{C} \cdot \text{m}$ and population relaxation and dephasing times, $T_1 = 100$ ps and $T_2 = 50$ fs, respectively. We then applied a switch-on pulse: a continuous, sinusoidal electric-field pulse, with pulse carrier frequency ω_0 in resonance with the two-level system, which is smoothly increased in amplitude from 0 up to unit amplitude ($E_0 = 1\ \text{V} \cdot \text{m}^{-1}$) within five periods:

$$E_x(t) = E_0 \sin\left(\omega_0 t\right) \begin{cases} 0, & \text{for } t < 0 \\ \left(1 - x^2\right)^4, & \text{for } 0 \le t \le 5T_\text{p} \\ 1, & \text{for } t > 5T_\text{p}, \end{cases} \tag{49.28}$$

with $x = \left(t - 5T_\text{p}\right)/5T_\text{p}$, and T_p being the period of one cycle of the sinusoid. The electric-field amplitude is clearly amplified upon the passage of the sinusoidal pulse across the gain region, thus demonstrating linear small-signal gain (Figure 49.7a). In the time domain, the application of an initial sinusoidal pulse with unit amplitude is amplified at the output end of the structure (Figure 49.7b). Let us compute the complex propagation factor between two points in the gain medium, z_1 and z_2, separated by one dielectric wavelength, $l = z_2 - z_1 = \lambda_0/n$:

$$e^{ik_c(z_2 - z_1)} = \frac{E_x\left(z_2, \omega\right)}{E_x\left(z_1, \omega\right)}, \tag{49.29}$$



(a)

(b)

FIGURE 49.7 (a) Slight amplification of a sinusoidal electric-field switch-on pulse upon its passage through an initially inverted two-level GaAs slab medium of refractive index 3.59 after 1×10^6 time steps (each of duration 1×10^{-4} fs) with Courant stability parameter 0.75. The initial population profile across the structure is shown with a thick solid line (top-hat shape). The spatial discretization in this simulation was 1 Å. (b) Intensity evolution in time of the pulse at the input boundary of the simulated structure (fast-oscillating curve on the left) and the amplified time trace of the pulse at the output boundary (time-delayed curve on the right). The fast oscillations are at the resonant pulse carrier frequency, ω_0.

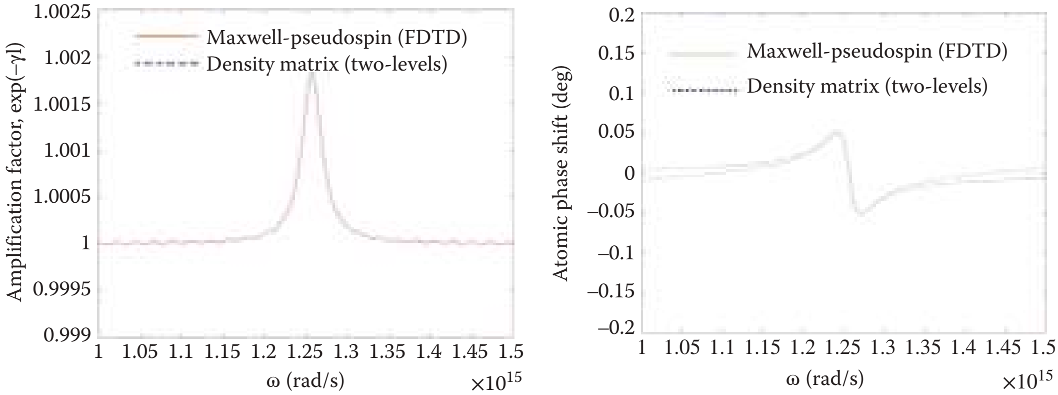
where $E_x(z_i, \omega)$ $i = 1, 2$, are the Fourier transforms of the time traces of the electric field sampled at the two points, and $k_c = \beta + i\gamma$ is the complex wavevector with β-phase shift and γ-gain/absorption coefficient, depending on its sign (+/−). Thus, substituting these back into the left-hand side of Equation 49.29, one can define the amplification and phase factors as $e^{-\gamma.l}$ and $e^{\beta.l}$, respectively. For the amplitude gain coefficient, the steady-state solutions of the density-matrix equations for a two-level system (see [50]) read:

$$\gamma^{ampl}(\omega) = -\frac{\omega_o}{c_o}\frac{N_a \wp^2}{2\varepsilon_o \hbar n_{\mathrm{GaAs}}} g(\omega), \tag{49.30}$$

where n_{GaAs} is the GaAs refractive index, c_0 is the speed of light, ε_0 is the dielectric permeability in vacuum, and the normalized Lorentzian lineshape function is given by: $g(\omega) = \frac{2T_2}{1+(\omega-\omega_o)^2 T_2^2}$, with normalization condition: $\int_{-\infty}^{\infty} g(\omega)d\omega = 1$. The "atomic" phase shift of the optical wave induced by the two-level system is given by

$$\Delta\varphi_{\mathrm{at}} = \Delta k.l = \frac{\omega}{c_o}\frac{N_a \wp^2 \left(\omega - \omega_o\right) T_2}{4\varepsilon_o \hbar n_{\mathrm{GaAs}}} g(\omega) = \beta l - n_{\mathrm{GaAs}}\frac{\omega}{c_o} l. \tag{49.31}$$

We can now compare the FDTD simulation results obtained with our Maxwell-pseudospin model (Equation 49.26) with the above analytic expressions (Equations 49.30 and 49.31); the results from the two methods are plotted in Figure 49.8. Our model is in remarkably good agreement with the density matrix-calculated gain/absorption coefficients (a) and phase shift (b) in a HB two-level system.



(a) (b)

FIGURE 49.8 (a) Amplification spectrum of a switch-on pulse with carrier wave of angular frequency, ω, passing through a GaAs slab waveguide. The spectrum was calculated using two independent methods: by applying the analytical expressions of the density-matrix method to a HB two-level system (Equation 49.30) (dashed curve), and by finding the FDTD solution of the Maxwell-pseudospin set of equations (Equation 49.26) (solid curve). The ω corresponding to the amplification peak is resonant with the optical transition of the two-level quantum system making up the waveguide and is denoted ω_0. Note that the ripple effect observed in the solid curve is an artefact of the FDTD numerical method. (b) ω-dependence of the phase shift (in degrees) of the switch-on pulse emerging from the waveguide. The phase-shift spectrum was inferred independently from the density-matrix method (Equation 49.31) (dashed curve) and the Maxwell-pseudospin method (Equation 49.26) (solid curve). FDTD, finite-difference time-domain.

49.3.4 Resonant Nonlinearities and Gain Saturation

In Section 49.3.3, we employed the analytical steady-state density-matrix equations to perform a successful validation of our Maxwell-pseudospin model against the linear small-signal gain in a HB two-level system. In what follows, we shall show that our model also describes correctly the nonlinear gain dynamics of a HB two-level saturable gain medium. We keep the model structure, source excitation (i.e., input pulse) shape, and material parameters the same as in Section 49.3.3, so that the saturation condition at resonance is satisfied:

$$\frac{\wp^2 E_0^2 T_1 T_2}{\hbar^2} > 1. \tag{49.32}$$

We apply the same source excitation (Equation 49.28) but vary the amplitude, E_0, running separate simulations for different amplitudes, ranging from $E_0 = 5 \times 10^6\ \mathrm{V \cdot m^{-1}}$ to $E_0 = 1 \times 10^8\ \mathrm{V \cdot m^{-1}}$. The nonlinear gain dynamics is shown in Figure 49.9 [51]. The gain peaks and levels off after some time, due to the depletion of the population residing on the upper level. The higher the initial excitation, the faster the gain reaches the steady-state field amplitude value. We proceed to demonstrate quantitative agreement between density-matrix theory of saturation of a two-level system and our Maxwell-pseudospin model. We choose the initial electric-field amplitude $E_0 = 5 \times 10^7\ \mathrm{V \cdot m^{-1}}$, $T_1 = 10\,\mathrm{ps}$, and $T_2 = 10$ fs, again satisfying the resonant saturation condition (Equation 49.32). The saturation value of the population inversion was calculated in [50] to be

$$S_{3\mathrm{sat}} = \frac{1}{1 + \left(\frac{E_0}{E_{s0}}\right)^2}, \quad E_{s0} = \frac{\hbar}{\wp\sqrt{T_1 T_2}}, \tag{49.33}$$

giving a value $S_{3\mathrm{sat}} = 0.3$ for the chosen parameter set. We can calculate the unsaturated intensity gain coefficient from the maximum intensity using $\gamma_{\mathrm{int}}^{\mathrm{unsat}} = \frac{ln(I_{\mathrm{max}}/I_o)}{2L_g}$. By definition, the saturated intensity is the intensity of the signal passing through the laser medium for which the gain coefficient is reduced down to one-half of its original value [52]:

$$\gamma_{\mathrm{int}}^{\mathrm{sat}} = \frac{\gamma_{\mathrm{int}}^{\mathrm{unsat}}}{1 + I/I_{\mathrm{sat}}}. \tag{49.34}$$

Given that the normalized saturated field intensity is $I_{\mathrm{sat}}/I_0 = exp(\frac{\gamma^{\mathrm{unsat}}}{2} L_g)$ [51], I_{sat}/I_0 is calculated to be 1.01. In order to make quantitative comparison with our model, we run the code with a large number of time steps (4.8×10^7, of 1×10^{-4} fs duration each) to ensure that the field and population dynamics have reached stationary values. In Figure 49.10a, we plot the final instantaneous electric-field distribution and population profile at subsequent time steps, showing the population relaxing to a steady-state value. In the time domain, the time trace at the waveguide output, (Figure 49.10b), and the population inside the gain region (Figure 49.10c), are plotted. It is easy to check that the steady-state values inferred from our model are in excellent agreement with the theoretically computed values. With this we conclude our corroboration of the ability of our model to reproduce correctly the gain saturation dynamics of a HB two-level quantum system.

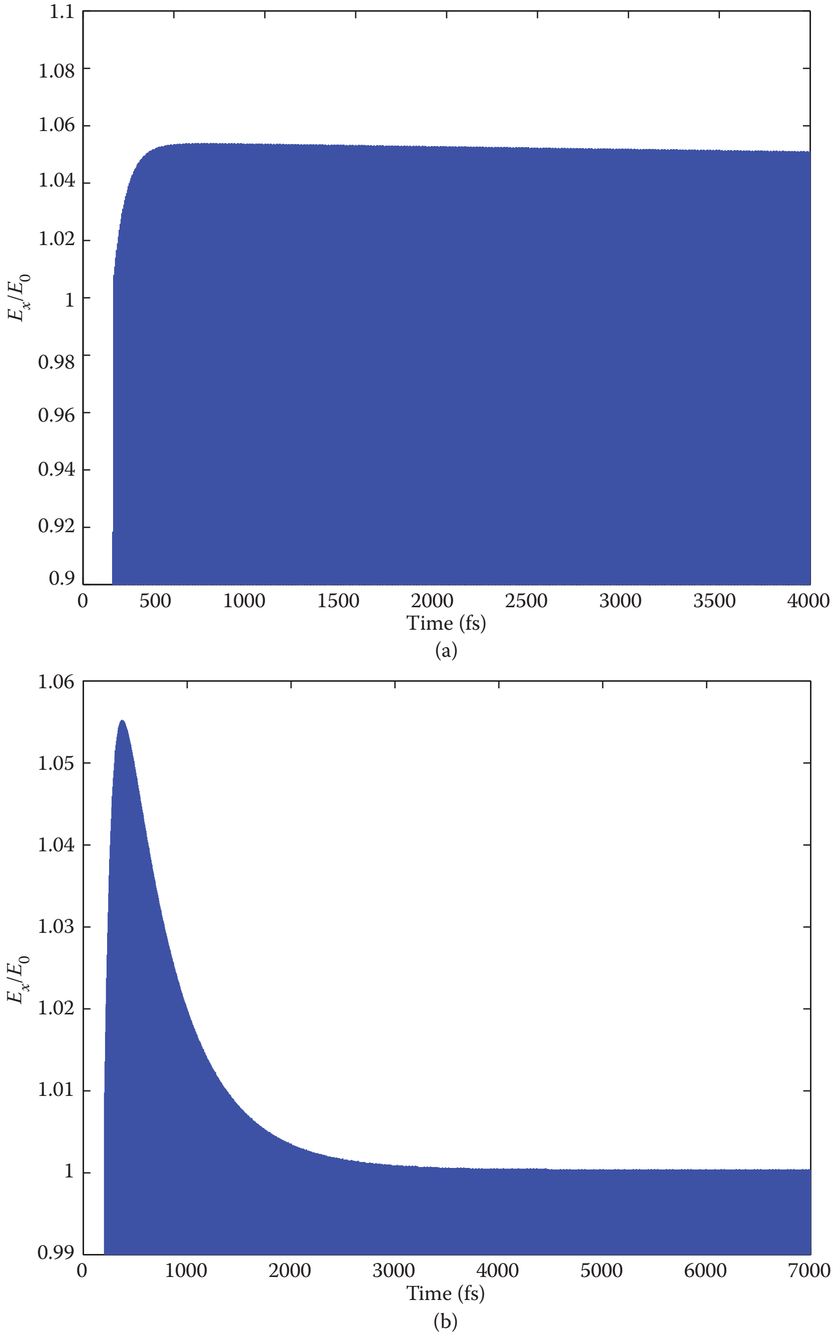


FIGURE 49.9 Gain saturation dynamics of a HB two-level quantum system excited by a source electric field with different amplitude: (a) $E_0 = 5 \times 10^6\ \mathrm{V \cdot m^{-1}}$; (b) $E_0 = 5 \times 10^7\ \mathrm{V \cdot m^{-1}}$. The time trace of the amplified electric field is sampled at the output of the simulation domain. In both cases, the amplitude of the oscillations (with frequency ω_0) stabilizes after some time. Note that after peaking, the gain in (a) continues to drop approximately linearly at a relatively slow rate over the entire time span shown on the graph, while the gain in (b) reaches a steady-state value after ~4 ps. (G. M. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2004 IEEE.)

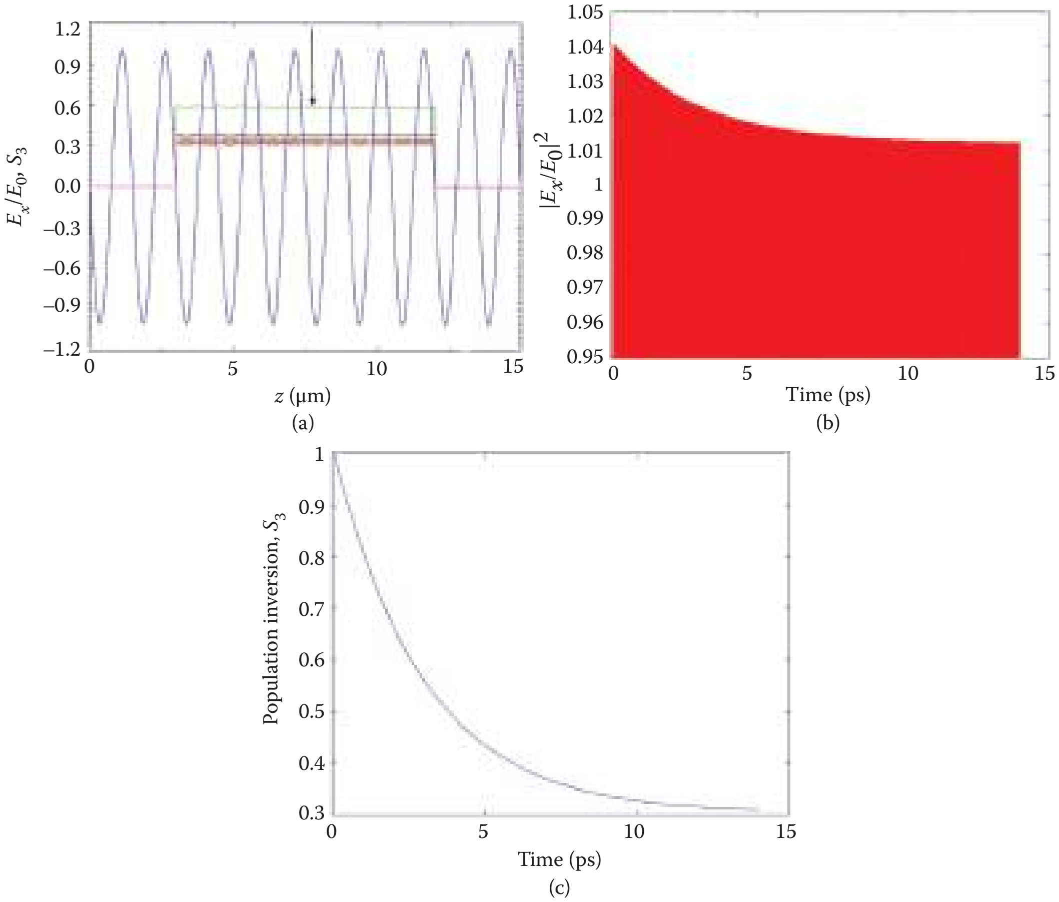


FIGURE 49.10 (a) Instantaneous electric-field amplitude profile across the waveguide (sinusoidal curve), obtained after 4.8×10^7 time steps (corresponding to a total time of ~4.8 ps), and the population inversion profile within the active region at several subsequent time steps (top-hat function with decreasing amplitude). The step in z is $\sim1.8\times10^{-10}$ m and the time step is $\sim1\times10^{-4}$ fs. The population is initially completely inverted: $S_{30} = 1$, but reduces over time (downward-pointing arrow), eventually converging to the theoretically obtained population steady-state value of 0.3. (b) Time evolution of the normalized field intensity showing convergence toward the theoretical value of 1.01. (c) Population relaxation over time, tending toward the theoretically predicted value of 0.3. (G. M. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2004 IEEE.)

49.3.5 Coherent Propagation Effects, Self-Localization and Pattern Formation in 2D Planar Nonlinear Optical Waveguides and Semiconductor Microcavities

49.3.5.1 2D Nonlinear Optical Waveguides

We employ our 2D Maxwell-pseudospin model (see Section 49.2.4), based on self-consistent solution of the time evolution equations of a degenerate three-level quantum system, (Equation 49.4), for $j, k, l = 1, 2, ..., 8$ and Maxwell's curl equations for a TE-guided mode, Equation 49.21 (Equation 49.22 for a TM equivalent), propagating in a planar parallel-plate mirror optical waveguide, or planar semiconductor microcavity [45], of the type shown in Figure 49.11.

It is easy to show that by setting $S_1 = S_1, S_4 = -S_2, S_7 = S_3$, the TE case is equivalent to a two-level system case (Equation 49.7) and the equations for the remaining coherence vector components are decoupled

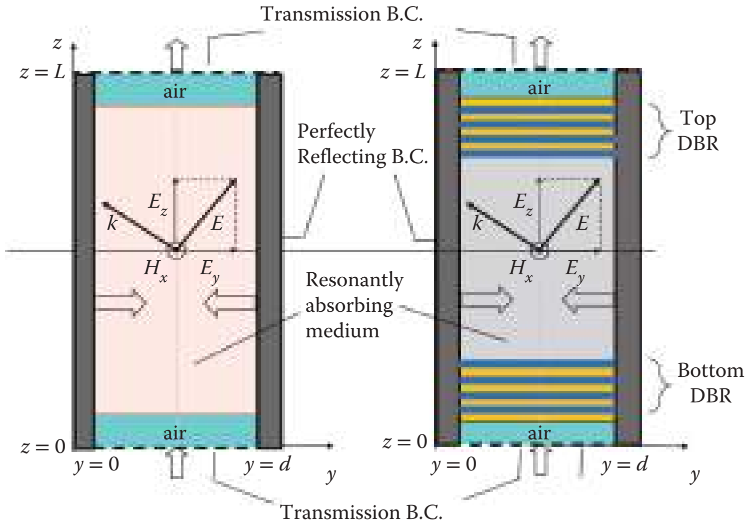


FIGURE 49.11 Planar parallel-mirror waveguide (left panel) and a semiconductor microcavity (right panel) filled with a resonantly absorbing three-level medium. Only TM-guided modes are shown. Perfectly transmitting boundary conditions are imposed on the upper/lower boundaries (upward-pointing arrows) and perfectly reflecting side boundaries are assumed (left/right arrows). The pulse initially propagates through a free-space region (air), then penetrates further, interacting with the degenerate three-level medium, and finally exits through another free-space region. B.C., boundary conditions; DBR, distributed Bragg reflector.

from this system, since there is no coupling to "level 3" in this case. In the FDTD numerical implementation, we use as a source field the TEM (i.e., TM_0) guided mode of a parallel-mirror slab waveguide. The electric field is a plane wave linearly polarized in the transverse x-direction perpendicular to the waveguide plane

$$E_y\left(z=0,y,t\right)=\begin{cases} E_0 \operatorname{sech}(10\Gamma)\sin\left(\omega_0 t\right) \\ E_0 \exp\left[\left(\frac{t-t_0}{t_\mathrm{d}}\right)^2\right]\sin\left(\omega_0 t\right) \end{cases} \quad E_z\left(z=0,y,t\right)=0\,, \tag{49.35}$$

where $\Gamma=\frac{t-\frac{T_\mathrm{p}}{2}}{\frac{T_\mathrm{p}}{2}}$. Here we have allowed for two electric-field envelopes: hyperbolic secant and Gaussian. We should therefore expect that the PAT [18] still holds. The pulse area at a given time is given by

$$\theta(z,t)=\frac{\wp}{\hbar}\int_{-\infty}^{t} A\left(z,t'\right)dt'\,, \tag{49.36}$$

where the electric-field envelope is as follows:

$$A(z,t)=E_0 \operatorname{sech}\left(\frac{t-\frac{z}{v}}{\tau}\right)=E_0 \operatorname{sech}(10\Gamma)\,. \tag{49.37}$$

The area under the entire pulse is given by

$$\theta_{\text{pulse}}(z) = \frac{\wp}{\hbar} \int_{-\infty}^{\infty} A\left(z, t'\right) \mathrm{d}t' . \tag{49.38}$$

According to PAT, pulses with areas less than π should be completely absorbed within several absorption lengths; however, pulses with an area of even multiples of π should continue to propagate undistorted as a solitary wave—known as SIT soliton—through the resonantly absorbing medium. We set all resonant dipoles initially to the ground state, i.e., $S_{7e} = -1, S_{8e} = -\frac{1}{\sqrt{3}}$. We then test the predictions of PAT by calculating the pulse amplitude from the PAT, which is given by the following relation for a pulse with a hyperbolic secant-shaped envelope:

$$E_0 = \frac{2\pi\hbar f_0}{\wp \arctan\left(\sinh\left(u\right)\right)\Big|_{-10}^{+10}} \frac{\theta_{\text{pulse}}(z=0)}{2\pi}, \tag{49.39}$$

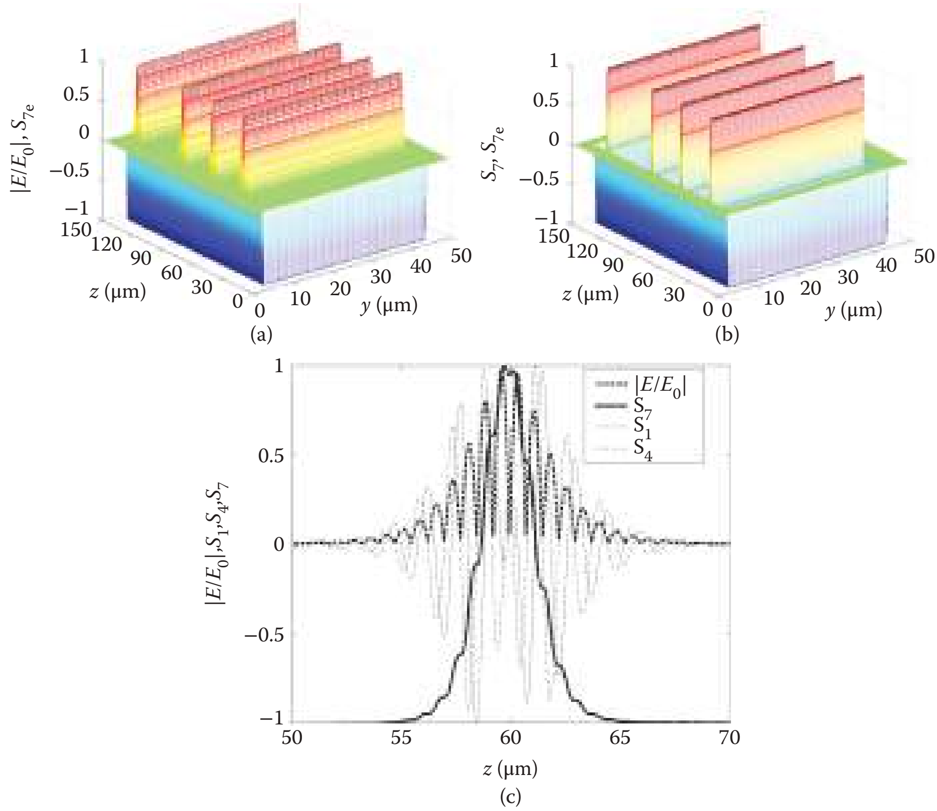
where $f_0 = 1/\tau_{\text{p}}$ and $\tau_{\text{p}} = T_{\text{p}}/20$, with $T_{\text{p}} = 100$ fs being the pulse duration chosen for this simulation. If we set the initial pulse area $\theta(z = 0)$ to 2π, we get a pulse amplitude corresponding to a soliton pulse of $E_0 = 4.2186 \times 10^9\ \text{V} \cdot \text{m}^{-1}$. The simulated waveguide structure is 150 μm long and 50 μm wide. 3D plots of the soliton field modulus (a) and population relaxation term S_7 (b) distributions in the waveguide *y-z* plane and a slice through the middle of the waveguide's width are displayed in Figure 49.12.

The stable SIT soliton propagation is shown in Figure 49.13a and b: the initial 2π pulse propagates in the resonantly absorbing medium while maintaining its hyperbolic secant shape without any loss. The atoms are driven into the excited state by absorbing ultrashort pulse energy, and by reradiating this energy to the field they return to the ground state. The corresponding population inversion exhibits complete transition from the ground level to the excited state and back to the initial state within one Rabi period. By contrast, for a 0.9π pulse with pulse amplitude $E_0 = 1.18957 \times 10^9\ \text{V} \cdot \text{m}^{-1}$, the initially symmetric pulse is distorted and reshaped during its propagation (Figure 49.13c). The initial equilibrium between the energy absorbed by the leading pulse edge and the energy reemitted by the trailing edge is violated and the pulse becomes increasingly asymmetric. In addition, the system is only partially inverted (Figure 49.13d). The reemitted energy of the continuously growing trailing edge is transmitted through the lower ($z = 0$) boundary. After a sufficiently long time, this pulse should eventually be completely absorbed, in agreement with the PAT.

Let us consider now the more interesting case of a TM_1-guided mode of a parallel-mirror slab waveguide, which cannot be reduced to the two-level system case. We solve self-consistently Maxwell's equations for a TM wave, Equation 49.22, with the pseudospin equations, (Equation 49.4), for a three-level system. As a source field we use the TM_1 slab waveguide mode

$$\begin{aligned}
&E_y\left(z=0,y,t\right) = E_o c_1 \cos\left(\frac{\pi y}{w}\right) \operatorname{sech}\left(10\Gamma\right) \sin\left(\omega_o t\right) = \tilde{E}_y\left(y,t\right) \sin\left(\omega_o t\right) \\
&E_z\left(z=0,y,t\right) = -E_o c_2 \sin\left(\frac{\pi y}{w}\right) \operatorname{sech}\left(10\Gamma\right) \cos\left(\omega_o t\right) = \tilde{E}_z\left(y,t\right) \sin\left(\omega_o t\right) \\
&c_1 = \frac{\sqrt{\frac{\omega_o^2 n^2}{c_o^2} - \left(\frac{\pi}{w}\right)^2}}{\omega_o \varepsilon_o n^2}; \qquad c_2 = \frac{\pi/w}{\omega_o \varepsilon_o n^2},
\end{aligned} \tag{49.40}$$

where *w* is the waveguide width. In order to obtain a stable soliton solution, we generalize the PAT to multiple spatial dimensions, defining the pulse area as the area below the modulus of the field envelope,



(a) (b) (c)

FIGURE 49.12 (a) Superimposed snapshots of the modulus of the electric field for a TEM guided mode propagating in a planar parallel-mirror waveguide at times $t = 150, 250, 350,$ and 500 fs after the arrival of the pulse at the left boundary of the waveguide. The initial population profile, S_{7e}, is also shown as an indication of the 2D boundaries of the active medium; the resonant dipoles are initially in the ground state with $S_{7e} = -1$, absorbing the incoming field radiation. (b) Corresponding population inversion term S_7 distribution across the waveguide plane, showing Rabi flops. (c) Slice straight across the middle of the waveguide along the propagation direction, showing the modulus of the electric field, dispersive, and absorptive polarization components, S_1 and S_4, respectively, and the population, S_7, performing a full Rabi flop along the waveguide axis. $\Delta z = 30\,\text{nm}$, $\Delta y = 500\,\text{nm}$, $\Delta t = 9.989 \times 10^{-2}$ fs, the pulse wavelength is $\lambda = 1.5\,\mu\text{m}$ and the duration $T_p = 100\,\text{fs}$. (Adapted with permission from G. Slavcheva et al., *Physical Review A*, 66, 063418, 2002. Copyright 2002 by the American Physical Society.)

which depends on the point in the waveguide plane we choose to consider:

$$\theta\left(y,z\right) = \frac{\wp}{\hbar}\int_{-\infty}^{\infty}\left|\tilde{E}\left(y,z,t'\right)\right|\mathrm{d}t' = \frac{\wp}{\hbar}\int_{-\infty}^{\infty}\sqrt{\tilde{E}_y^2 + \tilde{E}_z^2}dt'. \tag{49.41}$$

Assuming that the initial pulse is injected into the active medium at a point $y = y_0$ at the boundary ($z = 0$), stable pulses in the absorbing medium are obtained when

$$\theta\left(y = y_0, z = 0\right) = \frac{\wp}{\hbar}\int_{-\infty}^{\infty}\sqrt{\tilde{E}_y^2\left(y_0,0,t'\right) + \tilde{E}_z^2\left(y_0,0,t'\right)}\mathrm{d}t' = 2n\pi, \tag{49.42}$$

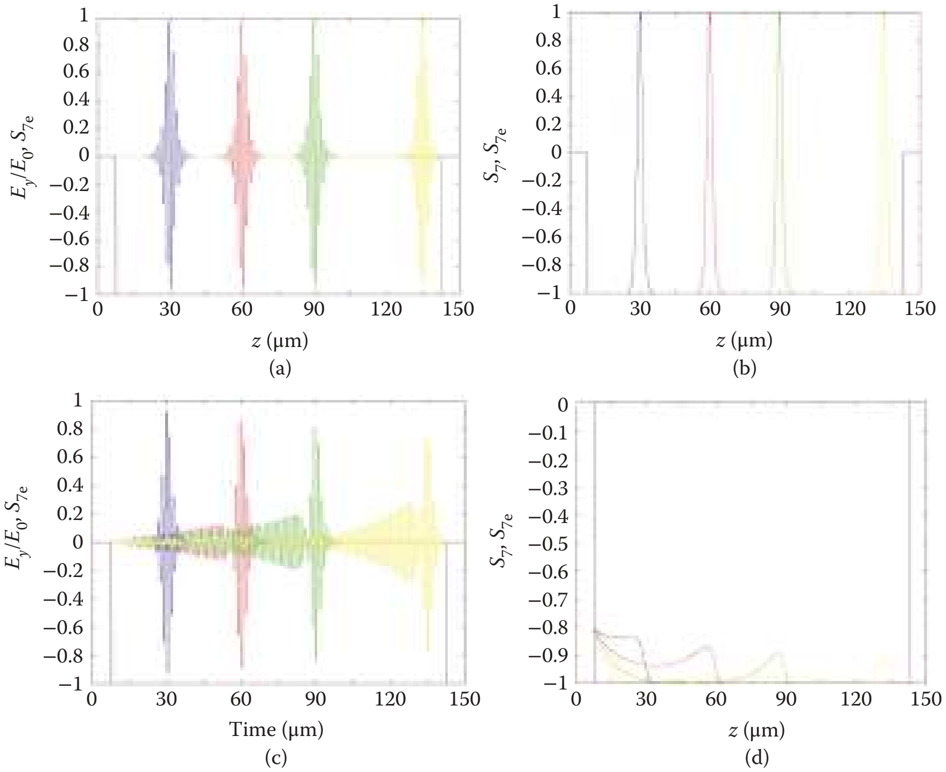


FIGURE 49.13 (a) Time evolution of a 2π initial pulse ($E_0 = 4.2186 \times 10^9$ V · m^{-1}) at $t = 150, 250, 350$, and 500 fs after impinging on the left boundary of a planar parallel-mirror waveguide. The soliton propagates undistorted through the resonant medium, the boundaries of which are shown by the initial population profile, $S_{7e} = -1$, i.e., all dipoles are initially in the ground state. (b) The corresponding population inversion term, S_7, exhibits full Rabi flops, whereby the absorbed energy is fully balanced by the emitted energy over one Rabi period. (c) Time evolution of a pulse with initial pulse area 0.9π (< π) ($E_0 = 1.18957 \times 10^9$ V · m^{-1}) at the same simulation times. In this case, the pulse becomes distorted due to imbalance of the absorbed and reemitted energies, and will eventually decay to zero in agreement with the PAT. (d) The resonant system is only partially inverted: there is no population swap between the ground and excited levels. (Adapted with permission from G. Slavcheva et al., *Physical Review A*, 66, 063418, 2002. Copyright 2002 by the American Physical Society.)

with n being an integer. A 2π pulse has an initial pulse amplitude $E_0 = 1.123 \times 10^7$ V · m^{-1}. The time evolution of the TM$_1$ soliton is shown in Figure 49.14.

49.3.5.2 2D Semiconductor Microcavities

Using the 2D model, we numerically investigate conditions of onset of SIT cavity soliton formation in a semiconductor cavity driven by a coherent ultrashort pulse launched from the lower boundary (see Figure 49.11, right panel) [53]. We have designed a number of 2D microcavity structures in order to test the influence of the cavity length on the localization phenomena and pattern formation. A typical 5λ-cavity (where the optical path length between the mirrors of the cavity is equal to an exact multiple of five wavelengths) designed at λ = 1.3 μm, filled with resonantly absorbing medium sandwiched between a bottom DBR (31.5 pairs of GaAs/AlAs) and a top DBR (24 pairs of $Al_{0.1}Ga_{0.9}As/Al_{0.8}Ga_{0.2}As$) is shown in Figure 49.15a and b. We apply a driving pulse excitation with duration 100 fs at the left boundary

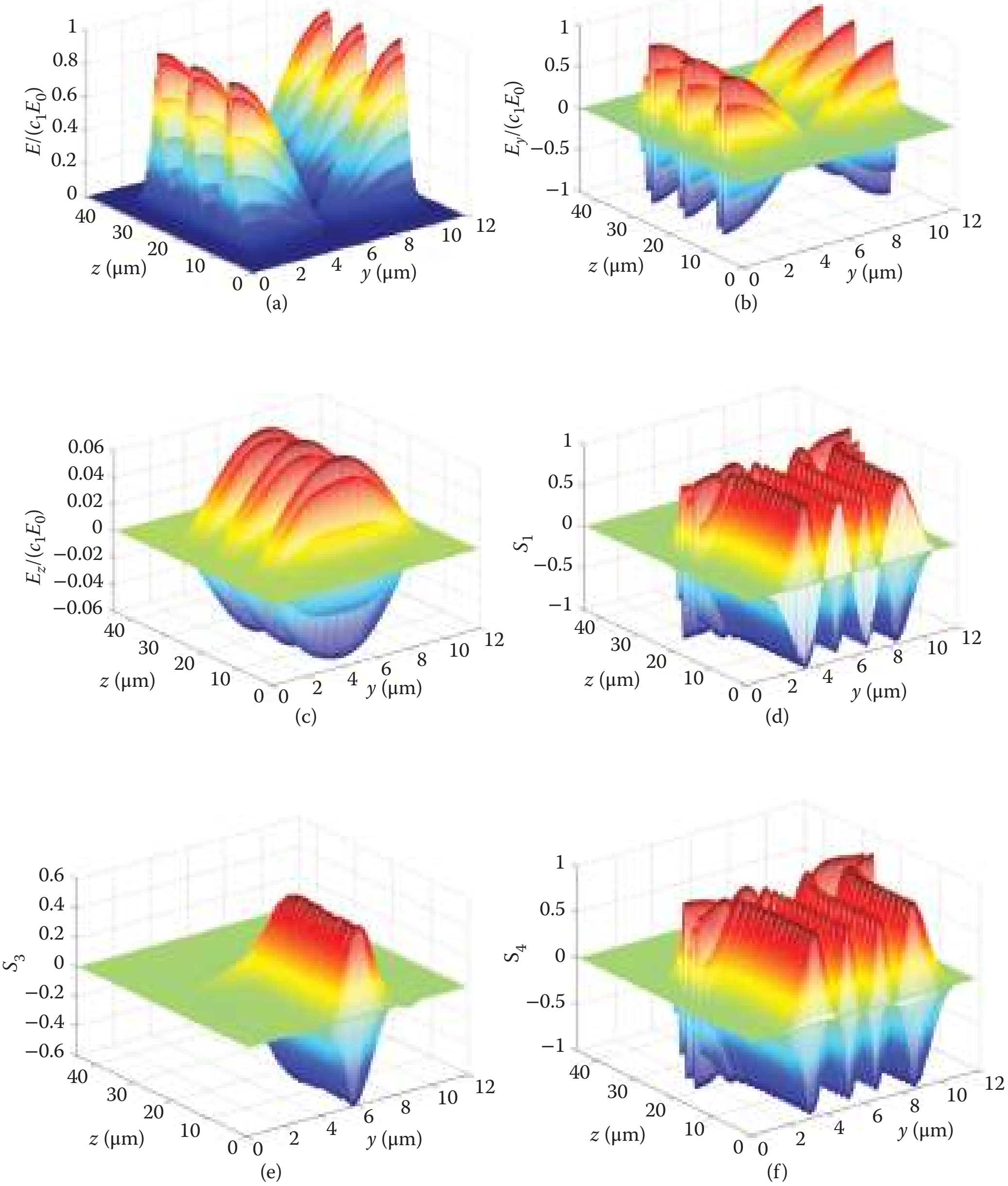


FIGURE 49.14 Time evolution of the (a) modulus, (b) E_y, and (c) E_z electric-field vector component of the TM_1 soliton at times $t = 90, 125$, and 155 fs after its arrival at the left boundary of a planar parallel-mirror waveguide. The 2D pulse travels undistorted in the resonant medium with refractive index, $n = 1$. The waveguide is 50 μm long and 9.9185 μm wide. Discretization grid: $\Delta z = 1.5$ nm; $\Delta y = 119.5$ nm; $\Delta t = 5 \times 10^{-3}$ fs. (d–g) Polarization components of significance (i.e., those corresponding to transitions from one level to another) in a degenerate three-level system. S_1 and S_4: in-phase dispersive components; S_3 and S_6: in-quadrature absorptive components. (*Continued*)

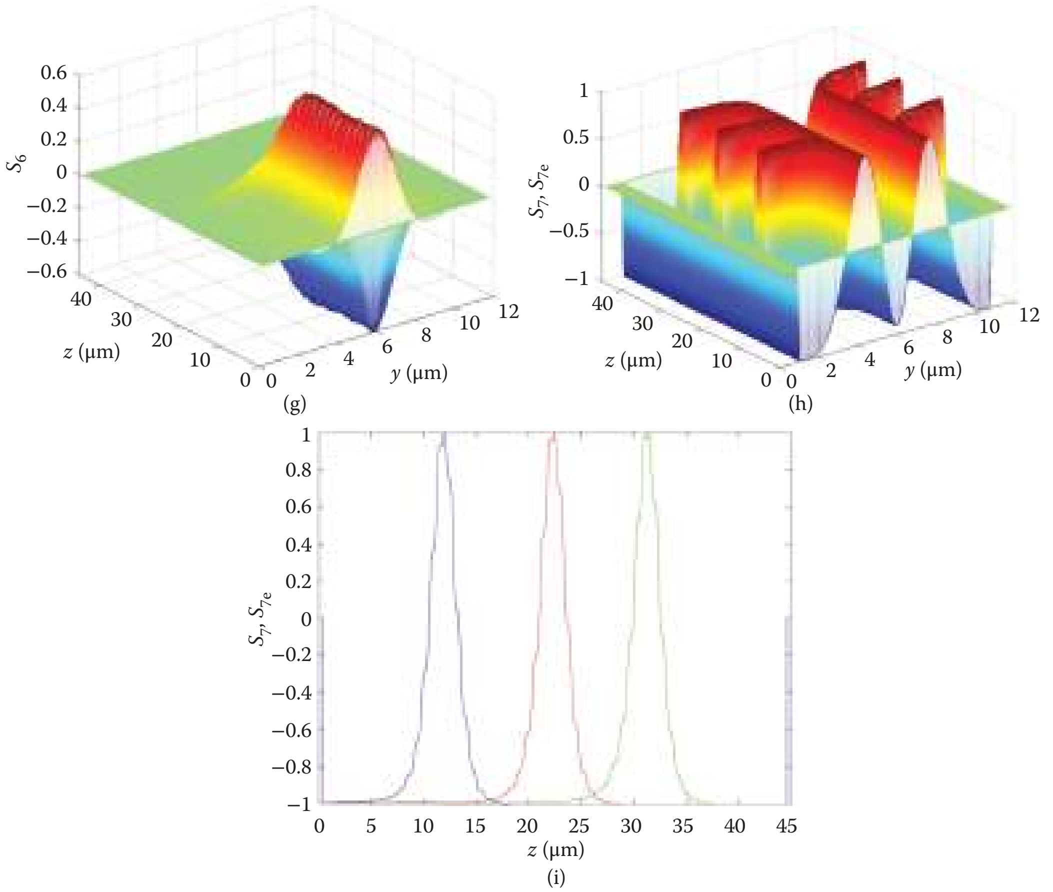


FIGURE 49.14 (*Continued*) (d–g) Polarization components of significance (i.e., those corresponding to transitions from one level to another) in a degenerate three-level system. S_1 and S_4: in-phase dispersive components; S_3 and S_6: in-quadrature absorptive components. (h) Population component S_7 showing population Rabi flops along the waveguide and in a transverse direction (two complete Rabi flops are visible). (i) lengthways cross section of (h) showing complete Rabi flops in the propagation direction, z.

($z = 0$) of the microcavity (Figure 49.15a), which is 12 μm long and $w = 9.92$ μm wide. The dipole number density is $N_a = 10^{24}\ \mathrm{m}^{-3}$ and the dipole matrix element, $\wp = 1 \times 10^{-29}\ \mathrm{C \cdot m}$. The relaxation times are assumed uniform and equal to 100 ps, satisfying the SIT criterion $T_\mathrm{p} \ll T_i,\ i = 1, 2, \ldots, 8$. The initial pulse amplitude $E_0 = 4.8643 \times 10^9\ \mathrm{V \cdot m^{-1}}$ is calculated for a pulse area of 2π, assuming that the PAT still holds. As pointed out in [54–56], this is not a necessary condition for SIT in multilayered media. Even a relatively weak pulse would lead to SIT intracavity patterns after sufficiently long simulation times, due to field enhancement by constructive interference from multiple reflections in the DBR mirrors. The population inversion gradually builds up with time and begins to perform Rabi flopping. In Figure 49.15c through f, the electric-field modulus and the corresponding population inversion profile are plotted for the cavity region. The clearly visible instantaneous roll patterns for the electric field and population inversion, S_7, feature a quasi-stationary standing-wave (i.e., quiescent) SIT soliton. The cavity induces a quasi-stationary population "grating", and the number of the quasi-complete Rabi flops (10) of the population equals the number of half-wavelengths along the cavity length, $L_c = (10\lambda)/(2n)$. At the

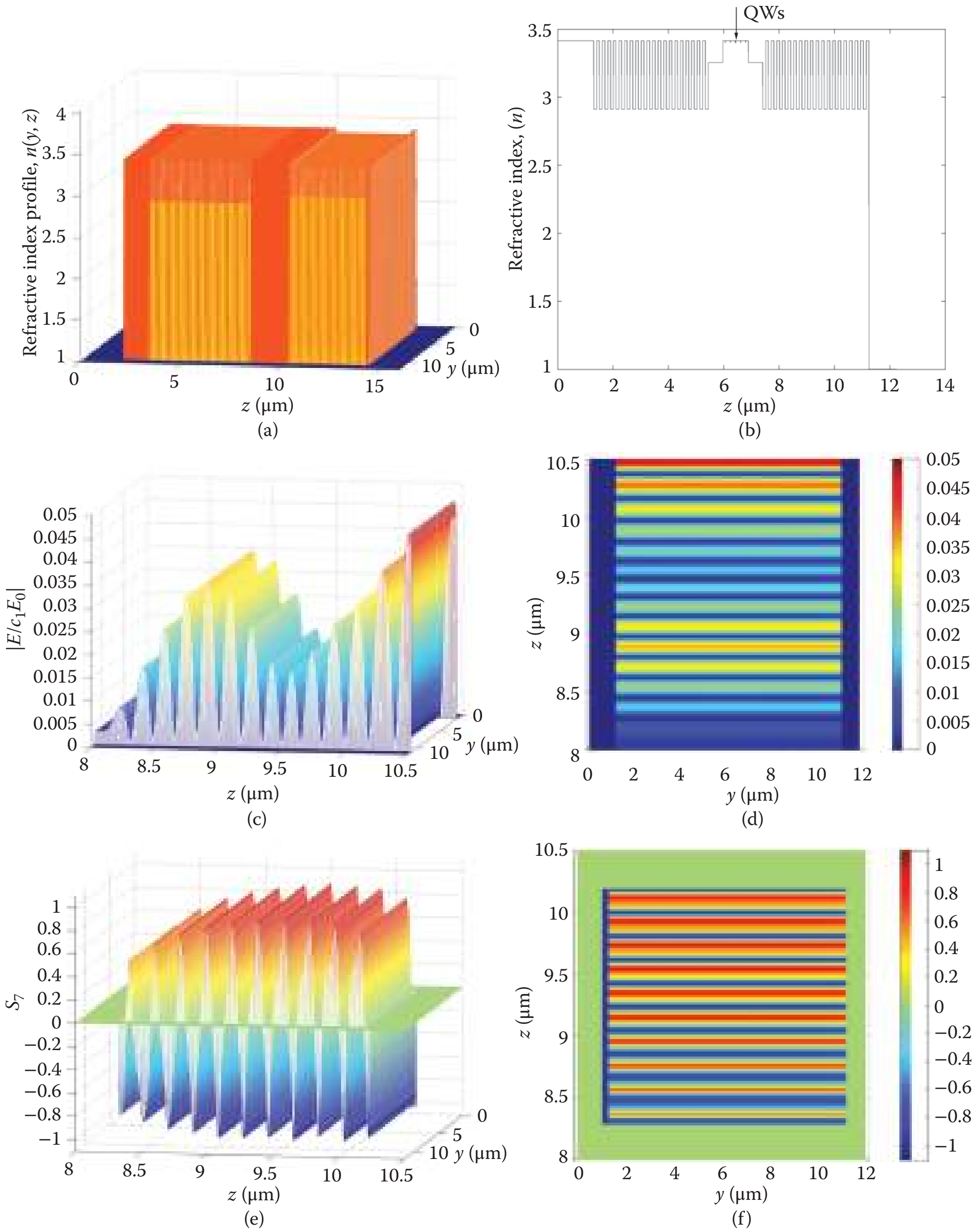
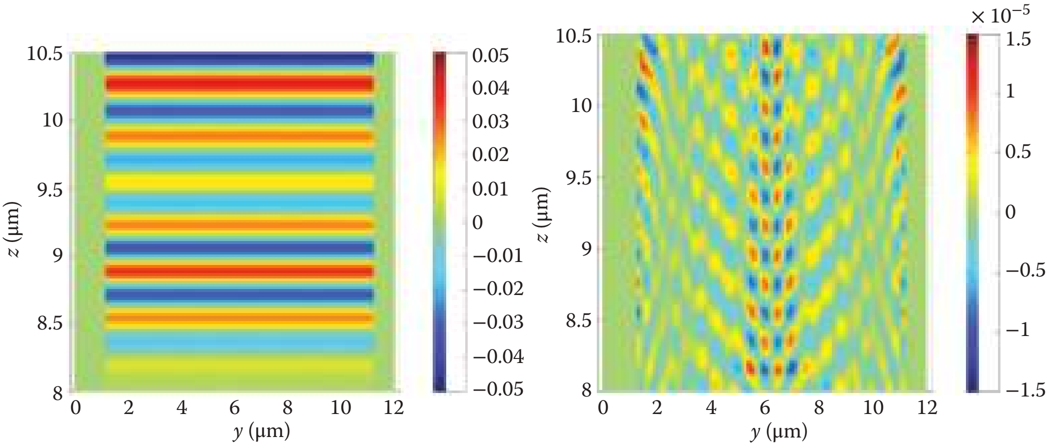


FIGURE 49.15 (a) 2D refractive-index profile across the semiconductor microcavity. (b) Longitudinal slice of (a) at a distance $y_0 = 6$ μm from the left interface, $y = 0$. (c) A snapshot at time $t = 3.6055 \times 10^{-13}$ s of the electric-field modulus distribution across the cavity. (d) Top view of (c) showing the roll patterns formed. (e) Snapshot of the corresponding population inversion profile across the cavity, showing 10 complete Rabi flops. (f) Top view showing the intracavity population "grating" (roll pattern). (G. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2003 IEEE.)

(*Continued*)



(g) (h)

FIGURE 49.15 (*Continued*) (g) Top view of the E_y-component roll pattern. (h) Top view of the 2D pattern of the E_z electric-field component. (G. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2003 IEEE.)

same time, a roll pattern for the E_y electric-field component, Figure 49.15g, and a more elaborate 2D one, Figure 49.15h, for the E_z field component are observed. Performing simulations with different driving pulse durations on a number of different microcavity designs, we demonstrate numerically the existence of a new type of 2D spatiotemporally localized standing-wave soliton ("light bullet"). The SIT cavity soliton results from the interplay of the resonant nonlinearity and the Bragg reflection interference and dispersion in the microcavity.

49.4 Optical Orientation in Quantum Wells and QDs

Circularly polarized optical excitations of semiconductor nanostructures excite particular interest due to their ability to transfer angular momentum and thus to drive the quantum system to a specific angular-momentum state. The conversion between the flying qubit encoded in photon polarization and the material qubit encoded, for example, in the carrier spin, is known as optical orientation, and this technique has been successfully applied to spin-polarized states in semiconductors [14]. This, in turn, opens up the possibility for faithful transmission of flying qubits between distant locations. If sufficiently short (and therefore high-intensity) optical pulses are used, so that they can interact with the system before it can be affected by its environment, the spin does not have time to decohere (i.e., lose its phase information) and could be coherently manipulated by such light pulses. In this high-intensity, coherent regime, optical Rabi oscillations take place; one-half of each oscillation consists of the population being driven from the ground to the excited state (or vice versa), hence performing a Rabi flop. A Rabi flop corresponds to a one-qubit rotation, whereby the spin is rotated through π; this entails a transition either from spin-down (state 0) to spin-up (state 1) or vice versa (Figure 49.1). The achievement of control and manipulation of the single charge-carrier spin dynamics and the collective many-body dynamics of multiple entangled spins is the key to the physical implementation of quantum computation.

49.4.1 Circularly Polarized Optical-Pulse Interactions with Two-Level Quantum Systems: Selective Spin Initialization and Readout

In bulk semiconductors and in semiconductor structures with reduced dimensionality, dipole optical transitions (including excitonic transitions) between energy levels are only allowable subject to satisfying the angular-momentum optical selection rules, $\Delta J_z = 0, \pm 1$, for linear and circular photon polarization respectively, where J_z is the total angular-momentum projection on the electromagnetic wave propagation direction coinciding with the QD growth axis. We shall consider optical transitions with $\Delta J_z = \pm 1$ that are excited by circularly (or elliptically) polarized light, and will therefore apply the pseudospin equations derived for circularly polarized light in Equation 49.8, coupled to Maxwell's vector equations for an electromagnetic wave propagating along the z-direction, circularly (elliptically) polarized in a plane perpendicular to z [57]:

$$\begin{aligned}
\frac{\partial H_x(z,t)}{\partial t} &= \frac{1}{\mu}\frac{\partial E_y(z,t)}{\partial z} \\
\frac{\partial H_y(z,t)}{\partial t} &= -\frac{1}{\mu}\frac{\partial E_x(z,t)}{\partial z} \\
\frac{\partial E_x(z,t)}{\partial t} &= -\frac{1}{\varepsilon}\frac{\partial H_y(z,t)}{\partial z} - \frac{1}{\varepsilon}\frac{\partial P_x(z,t)}{\partial t} \\
\frac{\partial E_y(z,t)}{\partial t} &= \frac{1}{\varepsilon}\frac{\partial H_x(z,t)}{\partial z} - \frac{1}{\varepsilon}\frac{\partial P_y(z,t)}{\partial t}.
\end{aligned} \tag{49.43}$$

We phenomenologically construct the Hamiltonian corresponding to initially absorbing energy-level schemes (i.e., $S_{30} = -1$), shown in Figure 49.16:

$$\hat{H} = \hat{H}_0 + \hat{H}_{\text{int}} = \hbar \begin{pmatrix} 0 & -\frac{1}{2}\left(\Omega_x - i\Omega_y\right) \\ -\frac{1}{2}\left(\Omega_x + i\Omega_y\right) & \omega_0 \end{pmatrix}. \tag{49.44}$$

The Hamiltonian for an initially amplifying two-level system (i.e., $S_{30} = 1$) is the transposed matrix. Here we have introduced a complex Rabi frequency, whose real and imaginary parts correspond to the $E_{x(y)}$ electric-field vector components, $\Omega_x = \frac{\wp}{\hbar}E_x$ and $\Omega_y = \frac{\wp}{\hbar}E_y$.

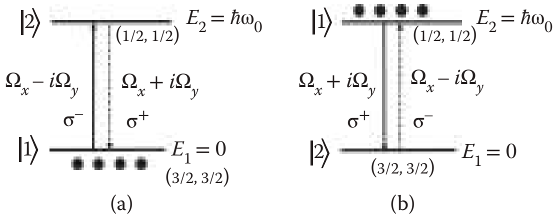


FIGURE 49.16 Energy-level diagram of (a) initially absorbing (i.e., entire charge-carrier population in the ground state) and (b) amplifying (i.e., entire charge-carrier population in the excited state) two-level system, with energy-level separation $\hbar\omega_0$ and allowed dipole optical transitions subject to the condition $\Delta J_z = \pm 1$. States are additionally labeled (J, J_z), where J is the total angular momentum and J_z is its projection along z. (a) A left circularly polarized optical field, σ^-, drives the population into the upper excited level, $|2\rangle$; the population is returned to the ground level by an optical pulse with opposite helicity, σ^+. (b) Having been initially prepared in the excited state, $|1\rangle$, the population in the system is driven by a right circularly polarized pulse, σ^+, to the ground state $|2\rangle$ and returned to the initial excited state by a left circularly polarized pulse, σ^-.

Employing the general methodology in Section 49.2.3, we can easily show that

$$P_x = -\wp N_a S_1, \qquad (49.45)$$
$$P_y = -\wp N_a S_2 \,.$$

We solve numerically by the FDTD method the system of equations, Equations 49.43 and 49.8, coupled through Equation 49.45. The source optical fields for left and right circularly polarized light are each given by two orthogonal, linearly polarized pulses, one time-delayed relative to the other by $\pi/2$, with carrier frequency ω_0 tuned in resonance with the two-level system:

$$\sigma^- \begin{cases} E_x(z=0,t) = \tilde{E}_x(t)\cos(\omega_o t) \\ E_y(z=0,t) = -\tilde{E}_y(t)\sin(\omega_o t) \end{cases} ; \; \sigma^+ \begin{cases} E_x(z=0,t) = \tilde{E}_x(t)\cos(\omega_o t) \\ E_y(z=0,t) = \tilde{E}_y(t)\sin(\omega_o t), \end{cases} \qquad (49.46)$$

where the pulse shape is determined by the envelopes:

$$\tilde{E}_{x,y}(t) = \begin{cases} E_0 \mathrm{sech}\,(10\Gamma) \\ E_0 \exp\left[-\left(t-t_0\right)^2/t_d^2\right], \end{cases} \qquad (49.47)$$

which we choose to be either hyperbolic secant or Gaussian. We consider a 15 μm-long simulation structure filled with either a resonantly absorbing or amplifying ($S_{30} = -1$ or $S_{30} = +1$) two-level medium, assumed to be GaAs (refractive index, $n = 3.3827$). We use a pulse wavelength, $\lambda_0 = 1.5$ μm and a pulse duration, $T_p = 30$ fs; the dipole coupling $\wp = 1 \times 10^{-29}$ C · m, and two-level system ensemble density $N_a = 1 \times 10^{24}\,\mathrm{m}^{-3}$ are taken from [47]. We restrict ourselves to the linear regime (initial electric-field amplitude, $E_0 = 1\,\mathrm{V \cdot m^{-1}}$). The material relaxation times are $T_1 = 100$ ps and $T_2 = 70$ fs. We launch a left circularly polarized pulse, σ^-, from the left boundary of the simulation domain and sample the electric field at two locations, z_1 and z_2, along the active medium. The complex propagation factor (see Equation 49.29) over one dielectric wavelength $l = z_2 - z_1 = \lambda/n$ is calculated for the E_x and E_y components, and the results for the gain/absorption coefficient spectra are compared with the steady-state density-matrix theory spectra (see Equation 49.30). A snapshot of the electric-field components, E_x and E_y, at the input and output boundaries are shown in the first column of Figure 49.17. The corresponding Fourier spectra and gain/absorption coefficients are shown in the second and third columns, respectively, of Figure 49.17 for an initially absorbing (Figure 49.17a through f) and initially amplifying (gain) medium (Figure 49.17g through l). The optical wave interacts with the two-level system, driving the population into the upper state (or returning the population to the ground state) by absorbing (or emitting) pulse energy. Therefore, we clearly observe gain/absorption in the Fourier spectra. The gain/absorption coefficients coincide with those for linear polarization. In addition, we have shown that when pumping an absorbing system with a right circularly polarized pulse, the pulse fails to excite the system, and similarly, when pumping an initially inverted system (the whole population residing on the upper level) by a left circularly polarized pulse, the system remains unaffected. Therefore our model correctly describes the excitation using circularly polarized light of dipole optical transitions subject to the optical selection rules, $\Delta J = \pm 1$, in a two-level system. We have thus numerically demonstrated selective excitation of specific spin states by circularly polarized pulses with predefined helicity.

49.4.1.1 SIT and Polarized SIT Soliton Formation

The results in Section 49.4.1 were obtained in the linear (small-signal) regime. Let us now consider the nonlinear coherent regime of ultrashort pulse propagation through resonant media, whose atomic lifetimes greatly exceed the pulse duration (i.e., such that T_1, $T_2 \gg T_p$). The simulation domain is 150 μm long and a two-level resonantly absorbing medium with refractive index $n = 1$ is embedded between two symmetric free-space regions, each with length 7.5 μm. We inject a resonant, left circularly polarized pulse,

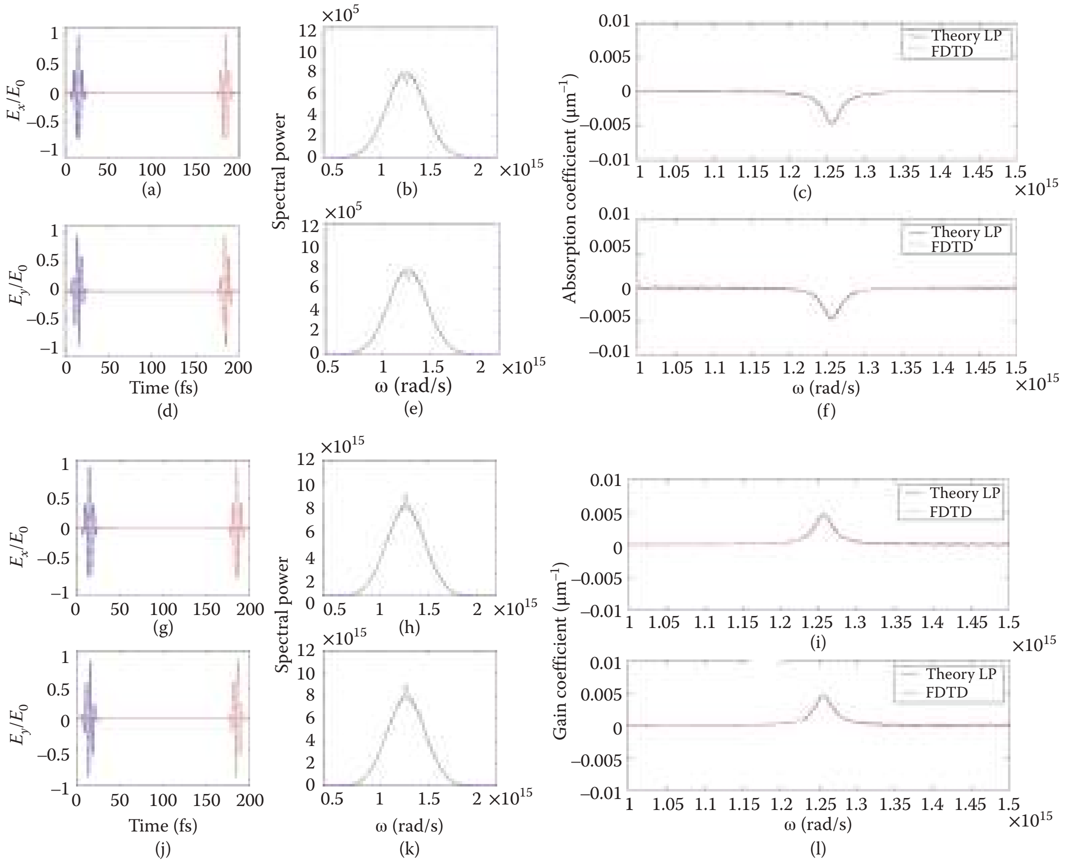


FIGURE 49.17 Time evolution at the input (solid line) and output (dashed line) structure boundaries of the electric-field components, E_x (a) and E_y (d), of a circularly polarized optical pulse with unit amplitude and pulse duration 30 fs for an initially absorbing two-level medium. (b), (e) Fourier spectra corresponding to the E_x and E_y field components, respectively, showing absorption of the pulse (dashed curve dip). (c), (f) Absorption coefficients for E_x and E_y, respectively; in this case, the dashed curve indicates the numerically calculated coefficients for linear polarization, while the solid curve signifies the analytically computed ones (Equation 49.30). (g–l) Respective plots for an initially amplifying (gain) medium; notations are the same as in (a–f). The gain is clearly visible in the Fourier spectra in (h) and (k) (see overshoot of dashed curve). (Adapted with permission from G. Slavcheva and O. Hess. *Physical Review A*, 72, 053804, 2005. Copyright 2005 by the American Physical Society.) (FDTD, finite-difference time-domain.)

σ^-, with hyperbolic secant envelope, since this is the well-known stable soliton solution of the Maxwell-Bloch system (see, e.g., [58]). We choose the electric-field component amplitudes so that the pulse area, according to the PAT, is 2π. This condition implies that $E_0 = 4.2186\times 10^9\ \text{V}\cdot\text{m}^{-1}$. The time evolution of a 2π pulse is shown in Figure 49.18a. Both circularly polarized pulse components travel undistorted through the resonantly absorbing two-level medium as a solitary wave, and the population inversion is driven through complete Rabi flops by the leading and trailing pulse edges. An expanded view of a snapshot of a circularly polarized SIT soliton at the simulation time $t = 0.2$ ps is shown in Figure 49.18b. By virtue of the PAT, stable soliton solutions in gain (initially inverted, $S_{30} = 1$) medium are obtained for pulse areas which are odd multiples of π. In order to confirm this prediction, we launch a right circularly polarized pulse with amplitude $E_0 = 2.1093 \times 10^9\ \text{V} \cdot \text{m}^{-1}$, which corresponds to a pulse area of π. As expected, a π-pulse completely de-excites the population back to the ground state (Figure 49.18c); the pulse amplitude increases,

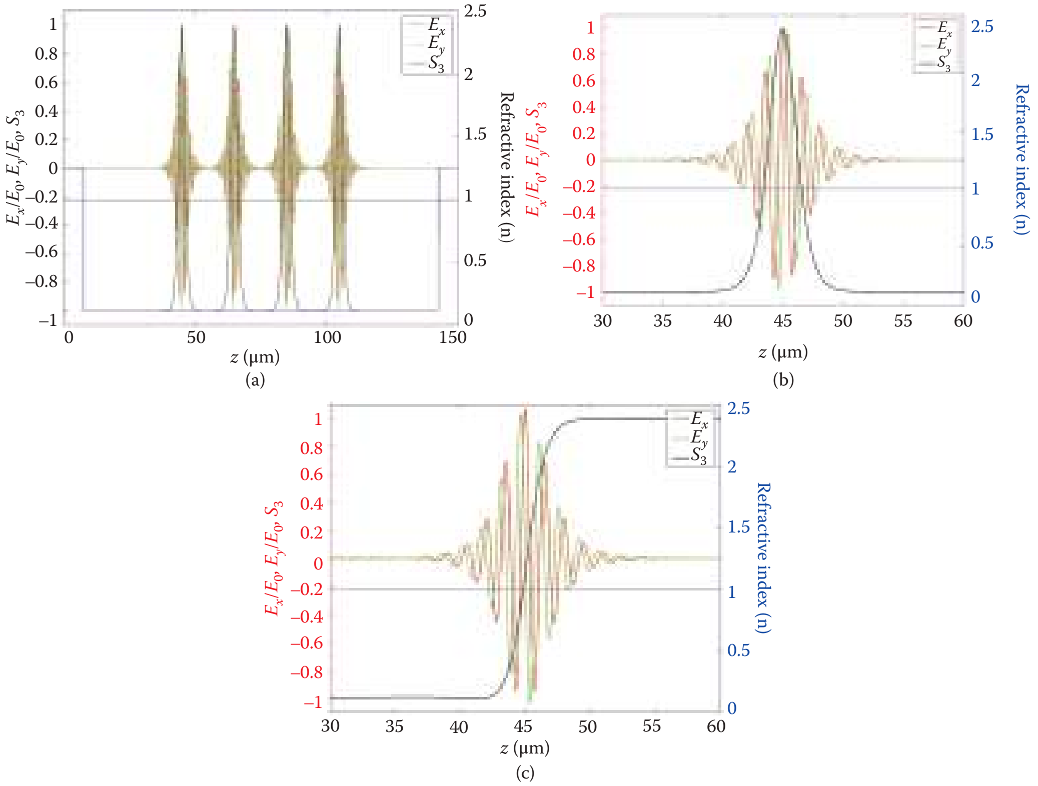


FIGURE 49.18 (a) Time evolution of a 2π, σ^- pulse propagating (from left to right) in a resonantly absorbing medium. The region where the initial charge carrier population profile is $S_{30} = -1$ demarcates the extent of the active medium on the diagram. The normalized electric-field components, E_x and E_y, and the instantaneous population profile, S_3, are plotted at the simulation times $t = 200, 267, 334$ and 400 fs from the moment the pulse is injected at the left boundary of the medium. The pulse travels without loss: it is a SIT soliton. (b) Expanded view of the σ^- SIT soliton in (a) at $t = 267$ fs showing the full Rabi flop of the population inversion. (c) A snapshot of a σ^+ SIT π pulse at $t = 267$ fs that completely inverts the population in the resonantly amplifying medium. (Adapted with permission from G. Slavcheva and O. Hess. *Physical Review A*, 72, 053804, 2005. Copyright 2005 by the American Physical Society.)

while the pulse duration decreases, thereby conserving the pulse area during the pulse propagation in the amplifying medium.

49.4.2 Circularly Polarized Optical Pulse Interactions with Degenerate Four-Level Systems

Let us consider now a more realistic discrete-level quantum system, such as excitonic transitions in bulk or low-dimensional semiconductor systems. For instance, we consider fundamental heavy-hole dipole optical transitions in QDs or quantum wells (QWs), whereby the band structure in the vicinity of the Γ-point ($\mathbf{k} = 0$) in the Brillouin zone (BZ) is isomorphic with a pair of two-level systems corresponding to σ^- and σ^+ heavy-hole excitonic transitions, as shown in Figure 49.19.

Consider a plane electromagnetic wave propagating along z direction, circularly polarized in a plane perpendicular to z, and tuned in resonance with the fundamental heavy-hole transition of the semiconductor system. The resonant nonlinearity is modeled by an ensemble of degenerate four-level systems of the type shown in Figure 49.19, with density N_a. The time-evolution of a four-level system under a dipole-coupling

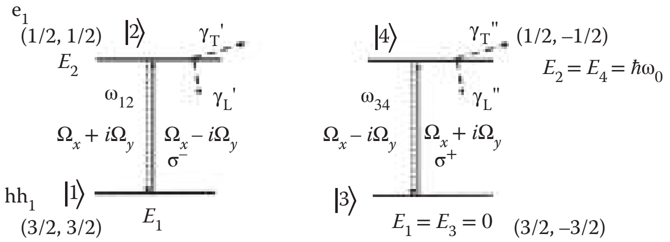


FIGURE 49.19 Energy-level diagram of σ^- and σ^+ heavy-hole transitions in a quantum well near the Γ-point ($\mathbf{k} = 0$) or between s-shell heavy-hole and electron levels in a QD. The energy separation between the upper and lower levels is $\hbar\omega_0$, and the allowed dipole transitions satisfy $\Delta J_z = \pm 1$. Ω_x and Ω_y are, respectively, the real and imaginary part of the Rabi frequencies associated with the coherent transitions. γ_L and γ_T are, respectively, the longitudinal relaxation and transverse dephasing rates of the excited states. e_1 is the bottommost electron energy level in the conduction band, and hh_1 is the topmost heavy-hole energy level in the valence band; together, these energy levels make up the fundamental bandgap.

external perturbation is described in terms of the SU(4) group generators by an equation of motion for a 15-dimensional real state pseudospin vector (see Equation 49.12 for $j, k, l = 1, 2, ..., 15$), where the dissipation in the system is accounted for by introducing nonuniform relaxation times, T_j. Without making any assumptions about the initial population redistribution between the levels, the system Hamiltonian is given by [59]

$$\hat{H} = \hbar \begin{pmatrix} 0 & -\frac{1}{2}\left(\Omega_x - i\Omega_y\right) & 0 & 0 \\ -\frac{1}{2}\left(\Omega_x + i\Omega_y\right) & \omega_0 & 0 & 0 \\ 0 & 0 & 0 & -\frac{1}{2}\left(\Omega_x + i\Omega_y\right) \\ 0 & 0 & -\frac{1}{2}\left(\Omega_x - i\Omega_y\right) & \omega_0. \end{pmatrix}. \tag{49.48}$$

We investigate the nonlinear coherent propagation regime, where the SIT condition $T_\mathrm{p} \ll T_j$ is satisfied with $T_\mathrm{p} = 100\,\mathrm{fs}$ and $T_j = 100\,\mathrm{ps}$. A circularly polarized pulse, amplitude-modulated by a hyperbolic secant function and whose carrier-wave frequency is resonant with the optical transition (at $\lambda = 1.5\,\mu\mathrm{m}$), Equation 49.46 and 49.47, is applied at the left boundary $z = 0$ of a 150-μm-long simulation domain, with either resonantly absorbing or amplifying media embedded between two free-space regions of thickness 7.5 μm each. We choose the initial pulse to be a π-pulse, giving $E_0 = 1.093 \times 10^9\ \mathrm{V \cdot m^{-1}}$ according to the PAT. We wish to achieve selective excitation of dipole optical transitions with $\Delta J_z = -1$ or $\Delta J_z = +1$, with a view to manipulate a particular spin state simply by selecting the helicity of the impinging optical pulse. The two leftmost columns of Figure 49.20 show all four possible cases where the initial state population of each two-level system is either entirely in the ground or entirely in the excited state. The two rightmost columns of the same figure also show the outcome of pulse excitation of each respective system. The conclusion that can be drawn is that we can indeed coherently control the spin population of specific spin states in the four-level system by choosing the proper polarization of the injected pulse. We have numerically demonstrated SIT and SIT soliton propagation in a four-level system initially prepared in a state with uniformly distributed populations in either the ground or excited states of the two two-level systems [60]. The ultrashort pulse travels undistorted in the resonant, degenerate four-level medium, Figure 49.21a, and the population is locally driven by the electric field through full Rabi flops simultaneously in both two-level systems, Figure 49.21b (compare with the two-level system soliton, Figure 49.18b).

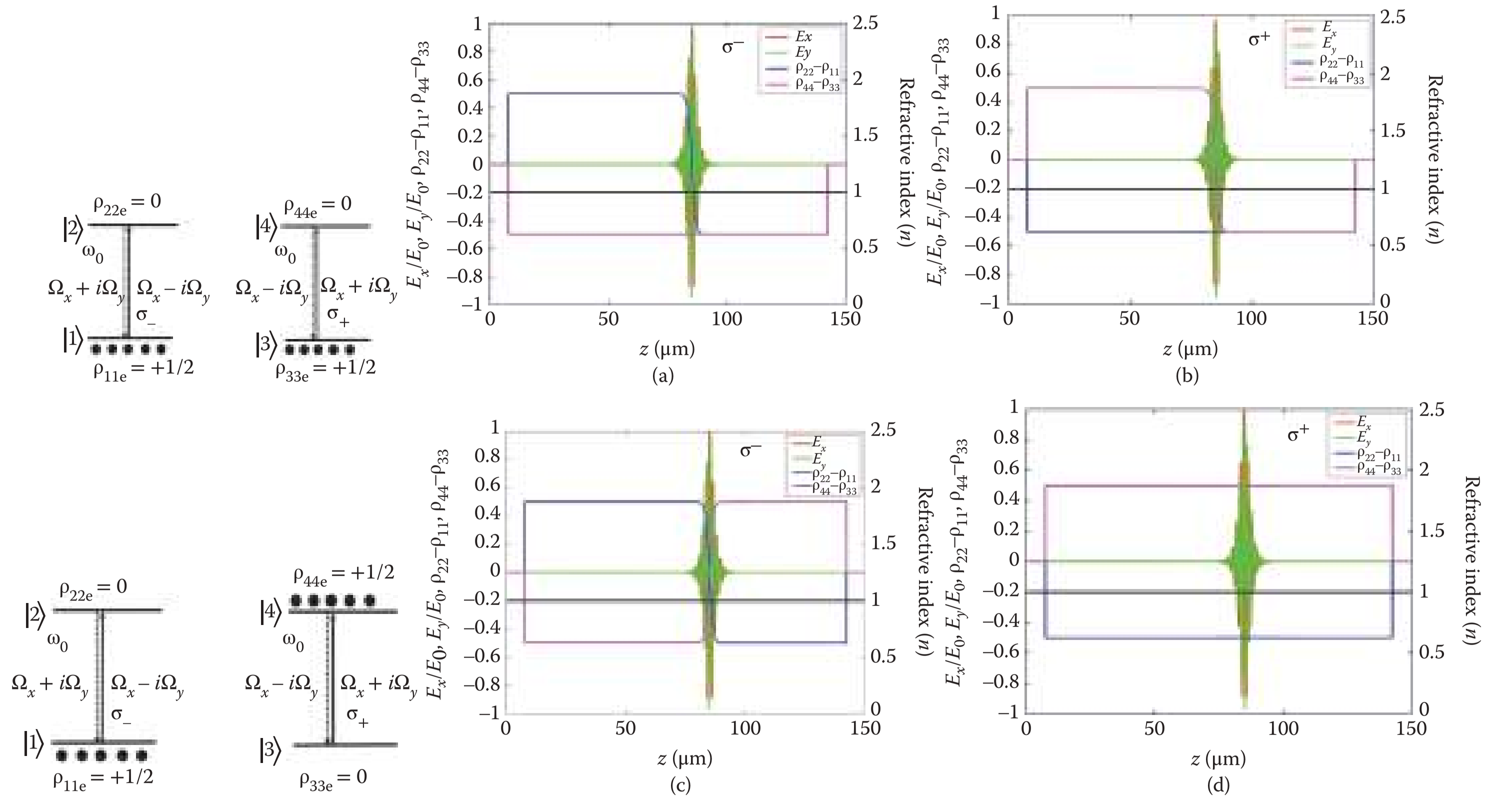


FIGURE 49.20 A snapshot of a circularly polarized pulse at time $t = 333$ fs after injection of the pulse at the medium's left boundary. E_x and E_y electric-field components and the population inversion $\rho_{22} - \rho_{11}$ in the first, and $\rho_{44} - \rho_{33}$ in the second two-level systems are plotted, tracing out the boundaries of the active medium with refractive index, $n = 1$. The leftmost two columns display the initial population distribution between the levels (equal split of the population between the two sets of two-level systems is assumed at the outset). (a) A σ^- π-pulse completely excites the first two-level system, but does not affect the second system. (b) A σ^+ π-pulse completely de-excites the initially inverted second two-level system, but does not affect the first one. (c) A σ^- π-pulse completely excites the first and de-excites the second system. (d) A σ^+ π-pulse does not affect either of the systems: the population in the first remains in the ground state and the population of the second remains in the excited state. (*Continued*)

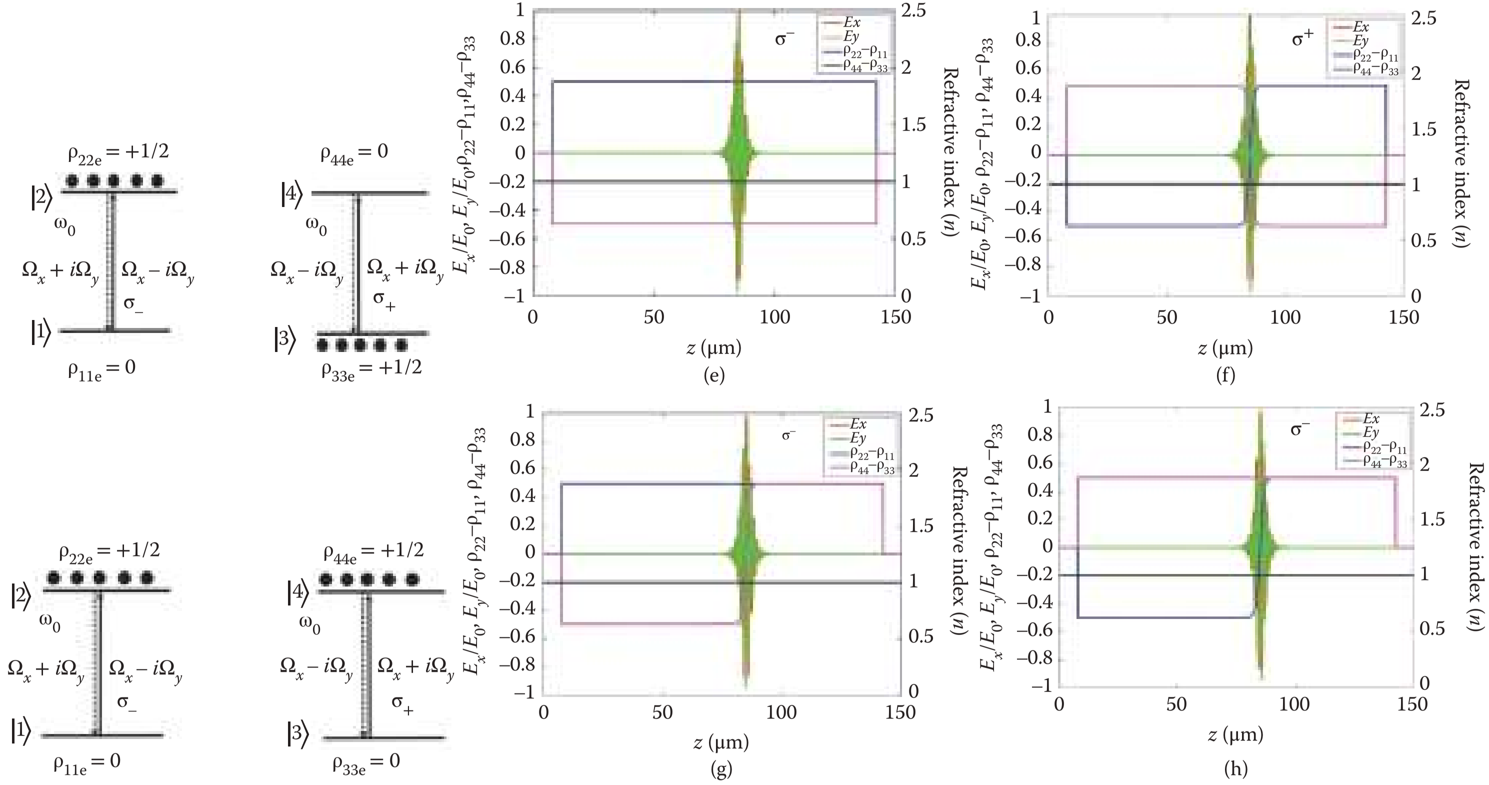
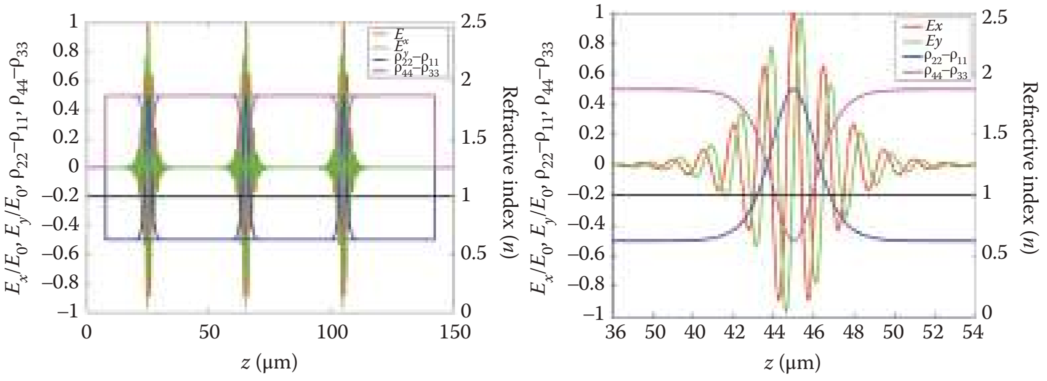


FIGURE 49.20 (*Continued*) (e) A σ^- π-pulse does not change the population in either of the systems. (f) A σ^+ π-pulse inverts simultaneously the population in both systems. (g) A σ^- π-pulse does not affect the first system, but de-excites the second. (h) A σ^+ π-pulse inverts the population in the first system, but does not affect the second system.



(a)

(b)

FIGURE 49.21 (a) Polarized SIT soliton propagation through a four-level medium resonant with the carrier wave of the injected pulse. The diagram shows a superimposition of the soliton at three different times: $t = 133$, 266 and 400 fs from the moment the pulse is injected at the left boundary of the medium. (b) A snapshot of the polarized soliton in the four-level medium, consisting of two pairs of two-level systems, at $t = 200$ fs. The medium is initially prepared in a state with the charge-carrier population residing in the lower-lying level, $|1\rangle$, of the first two-level system and in the upper level, $|4\rangle$, of the second two-level system (see Figure 49.19).

49.5 Modeling Coherent Spin Generation, Manipulation, and Readout in Charged QDs

In Section 49.4, we demonstrated coherent, selective spin excitation dependent on the helicity orientation of an impinging ultrashort optical pulse. In this section, we will show how spin states in low-dimensional semiconductor structures can be generated, manipulated, and read out using ultrashort circularly polarized pulses. Our goal will be to compute the time evolution of a discrete-level system and map the allowed dipole optical transitions that occur in a semiconductor nanostructure (e.g., a single QD or an ensemble of QDs) upon ultrashort, circularly polarized optical excitation. The pulse leaves behind a long-lived coherence, and we shall look for characteristic signatures in the dynamics that would allow us to determine the initial spin state and find useful regimes and schemes for high-fidelity optical manipulation of the spin confined in a QD. One promising approach is to address optically individual carrier spins in semiconductor QDs and manipulate them through optically excited states (charged excitons) by employing the techniques of coherent quantum control and optical orientation.

49.5.1 Single-Charged QD Embedded in a Waveguide

We will study n-doped QDs with an overall negative charge of a single electron. Injection of a single electron into a QD can be achieved either by modulation doping of the barrier region, adjusting the impurity doping within a delta-doped layer to transfer on average one electron per dot to the lowest energy states, or by electrical injection. A large class of self-assembled QDs posses quasi-cylindrical geometry about the quantization axis, z, and the electron and hole single-particle states in the conduction and valence band, respectively, can be approximated by harmonic oscillator potentials [61] (Figure 49.22). The resonant, circularly polarized optical excitation of a charged QD leads to the formation of a three-particle complex—a trion—consisting of two electrons occupying the same lowest conduction-band electron level in a spin singlet state, and a hole occupying the lowest valence-band hole level (Figure 49.22c and d). There is strong evidence from both experiment [62] and theory [63,64] that an intense resonant excitation of the trion transition suppresses the electron spin relaxation due to hyperfine interaction of the electron spins with

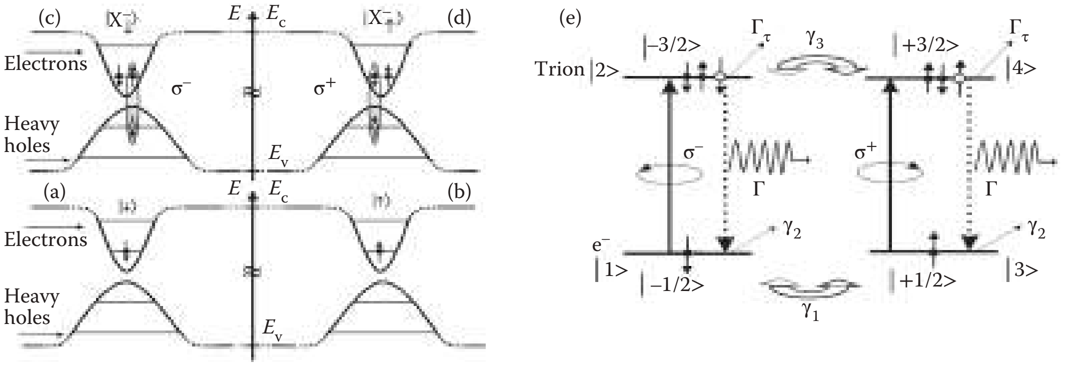


FIGURE 49.22 (a), (b) Schematic of the initial electron states (spin down and spin up) in a singly negatively charged QD and (c), (d) the ground singlet negatively charged exciton, X^-, which forms upon left circularly polarized, σ^-, and right circularly polarized, σ^+, optical excitation of the heavy-hole excitonic transitions. (e) Energy-level diagram of a negatively charged trion (a three-particle complex of two electrons and a hole), equivalent to (a–d); the levels are labeled with the J_z quantum number, and the allowed dipole transitions correspond to $\Delta J_z = \pm 1$ with an energy gap of $\hbar\omega_0$. [Solid upward arrows, coherent optical transitions excited by σ^- or by σ^+; dashed arrows, spontaneous emission rate, Γ; double-headed arrows, transitions due to electron- and hole-spin relaxation, with rates γ_1 and γ_3, respectively; γ_2 and Γ_τ, spin decoherence rates for electrons and holes, respectively.]

the lattice ions' nuclear spins. The latter limits the ability to measure accurately the electron spin orientation at low temperatures. It is obvious that the longer the spin lifetime, the better the chances for optical manipulation and readout of that spin state. We shall investigate both low- and high-intensity nonlinear excitation regimes.

In our treatment of a single quantum object, we shall assume that the ergodic hypothesis holds—i.e., the equivalence of time averages of an observable (in this case, the single-dot polarized, time-resolved photoluminescence TR [PL]) and the ensemble averages at one time over a large number of identical replicas of the dot system. Under such conditions, quantum mechanics allows the prediction of the single-system properties on the basis of macroscopic ensemble averages. Therefore, we shall assume that the time dependence of the optically induced coherent spin generation and subsequent relaxation in a single QD, averaged over a large number of successive measurements, is equivalent to the corresponding spin dynamics of an ensemble of identically prepared QDs. The ensemble of degenerate four-level systems describing the QD resonant nonlinearities is resonantly coupled to an optical wave propagating along the QD growth axis, z, and circularly polarized in a plane perpendicular to z. We use our general methodology to tackle this problem [44] and solve the 1D Maxwell's curl equations (Equation 49.43) coupled to the master pseudospin equations for a four-level system (Equation 49.12) through the induced macroscopic polarization given by

$$\begin{aligned} P_x &= -\wp N_a S_1 \\ P_y &= -\wp N_a S_7. \end{aligned} \tag{49.49}$$

The system under investigation is a GaAs/AlGaAs self-assembled, molecular beam epitaxy (MBE)-grown QD with a height of 5 nm and refractive index, $n_{\mathrm{GaAs}} = 3.63$, sandwiched between two $Al_{0.3}Ga_{0.7}As$ barrier regions, each with a width of 50 nm and refractive index, $n_{\mathrm{AlGaAs}} = 3.46$, at the trion fundamental transition resonance wavelength of 757 nm. The circularly polarized pulse, whose carrier wave is resonant with the trion transition, is described by a hyperbolic secant envelope (see Equation 49.46 and first line of Equation 49.47) and has a duration of 1.3 ps. The simulations are run repeatedly with a different initial pulse amplitude, E_0, which is varied from a low value, $E_0 = 550\,\mathrm{V \cdot m^{-1}}$, through $E_0 = 3 \times 10^6\,\mathrm{V \cdot m^{-1}}$, representing a π-pulse completely inverting the spin population in the system, to $E_0 = 4 \times 10^7\,\mathrm{V \cdot m^{-1}}$, corresponding to a pulse area of 12π and inducing six full Rabi flops. The fundamental trion transition dipole moment is estimated to be $\wp = 4.8 \times 10^{-28}\,\mathrm{C \cdot m}$, and the 3D number density of the resonant four-level dipoles, $N_a = 2.5 \times 10^{24}\,\mathrm{m^{-3}}$, is calculated from the QD surface density to give on average one dot within the microscopic single-dot volume. The trion recombination (spontaneous emission) rate, $\Gamma = 400\,\mathrm{ps^{-1}}$, is taken from [62]; the electron spin-flip rate due to hyperfine interaction with the lattice ions' spins, $\gamma_1 = 0.5\,\mathrm{ns^{-1}}$, has been calculated in [65,66]; the hole spin-flip relaxation is widely assumed to be due to phonon-assisted processes, and an estimate of $\gamma_3 \sim 170\,\mathrm{ps^{-1}}$ [67,68] has been given. The trion-bound electron and hole spin phonon-assisted decoherences are taken as $\gamma_2 = 450\,\mathrm{ps^{-1}}$ and $\Gamma_\tau = 340\,\mathrm{ps^{-1}}$ [69]. We shall assume that the initial spin population resides either in state $|1\rangle$ with spin-down, or in state $|3\rangle$ with spin up. We initially excite the $|1\rangle \rightarrow |2\rangle$ transition by a σ^- pulse. The simulation results for the time evolution of the pulse's electric-field vector components and the populations of all four levels for three initial electric-field amplitudes, E_0, are summarized in Figure 49.23. Excitation of the $|3\rangle \rightarrow |4\rangle$ optical transition is, of course, identical to exciting the $|1\rangle \rightarrow |2\rangle$ transition. The simulated detection of the time-resolved signals shows that there is a sufficiently long time interval—on the order of 400 ps—within which it is possible to determine the spin of the initial states. Indeed, there are two reliable ways for initial spin-state identification: through examining the PL trace profile and through the presence or absence of a second echo pulse after a $\sigma^{+/-}$ π-pulse excitation. The observation of a nonmonotonic PL trace, rather than an exponential decay, is always associated with a spin-up state; in addition, the appearance of a second echo pulse post σ^- π-pulse excitation alludes to a spin up, whereas a repeated echo after a σ^+ π-pulse excitation is indicative of a spin down. The simulations show the onset of the high-intensity Rabi oscillations regime, which suppresses the spin-relaxation processes.

49.5.2 Coherent Spin Manipulation through Hot-Charged Exciton States in QDs

Optical manipulation of the spin of a QD-confined single electron or hole through the resonantly driven trion ground singlet transition (described in the previous section) is considered to be one of the most promising schemes for implementation of spin-based quantum computing, due to the extended spin lifetimes which are limited only by the hyperfine interaction. However, the requirement of a resonantly driven ground trion transition in an inhomogeneously broadened ensemble of charged QDs represents an obstacle for the scalability of architectures based on this quantum system. This problem can be overcome by optical excitation into the excited trion states, taking advantage of a recently discovered effect, herein referred to as "spin filtering". The effect is observed in p-doped QDs under nonresonant, circularly polarized optical excitation, and constitutes an enhanced, photo-induced circular dichroism in the excited trion-state emission, compared to the dichroism exhibited by the ground singlet trion state [70]. The degree of spin polarization is nearly doubled when a high-energy level in the QD is excited resonantly. This is due to an increased spin "injection" efficiency—i.e., the resonant excitation prevents relaxation processes from occurring and the desired spin is maintained. In addition, nonmonotonic dependence of the degree of spin polarization on the optical pulse power is observed, allowing to maximize it by optimizing the pulse characteristics. This spin-filtering effect is promising for realization of high-fidelity schemes for all-optical spin manipulation, since the increased polarization contrast for the two pulse helicities enables highly efficient selective excitation and readout of the spin-up and spin-down populations.

We consider quasi-resonant σ^- or σ^+ circularly polarized excitation of a p-doped QD ensemble into the p-shell in the presence of a resident s-shell hole, and subsequent cascade relaxation to the bright trion ground singlet state, whose decay is detected as PL. The optical excitation creates X^{+*} states, consisting of one electron–hole pair in the s-shell and a resident hole in the p-shell ($1e^1 1h^1 2h^1$), grouped in four degenerate doublets [71]. The exchange interactions can be revealed in the excited exciton spectra, and the energy-level diagram and the spin configurations of the X^{+*} trion are given in Figure 49.24.

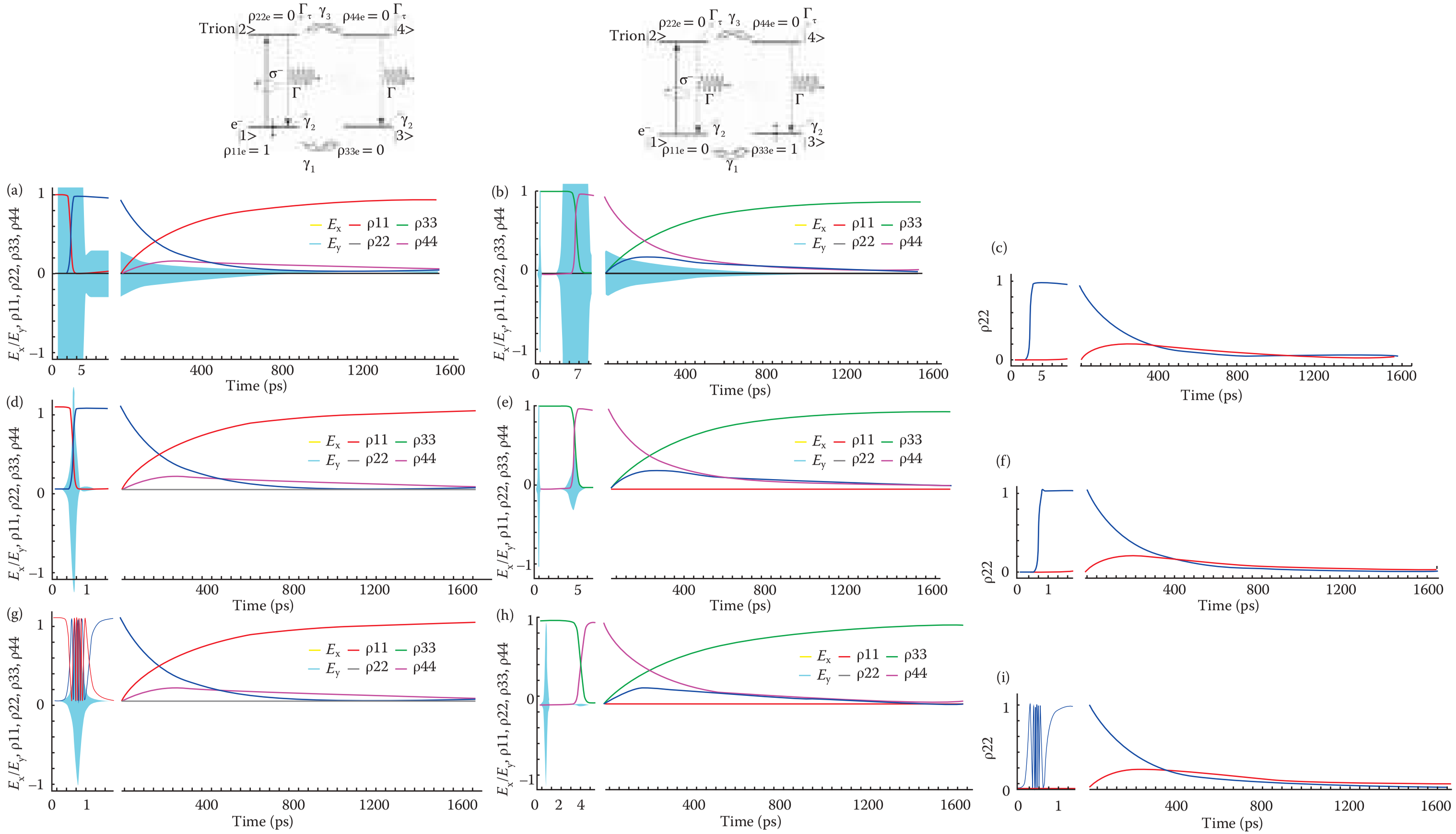


FIGURE 49.23 Top row: initial spin-down (left) and spin-up (right) state. The system is excited by a σ^- polarized pulse. (a–c) Excitation with $E_0 = 550\ \mathrm{V \cdot m^{-1}}$. Time evolution of (a) initial spin-down state $|1\rangle$; (b) initial spin-up state in $|3\rangle$; ρ_{22} (blue curve) represents the trion $|-3/2\rangle$ state spin population proportional to σ^- polarized PL signal detected experimentally. (c, f, i) TRPL signal for ⇓ (blue curve) and ⇑ (red curve) exhibiting different dynamics in the short term: ρ_{22} populations from the graphs to their respective left are plotted; (d–f) Excitation with a π-pulse with $E_0 = 3\times10^6\ \mathrm{V \cdot m^{-1}}$. Time-resolved dynamics for (d) spin-down initial state; (e) spin-up initial state. (g–i) Excitation with a 12π-pulse with $E_0 = 4\times10^7\ \mathrm{V \cdot m^{-1}}$. Time-resolved dynamics for (g) spin-down and (h) spin-up initial states. Excitation of initial spin-down state results in six full spin population Rabi flops. (Adapted with permission from G. Slavcheva, *Physical Review B*, 77, 115347, 2008. Copyright 2008 by the American Physical Society.)

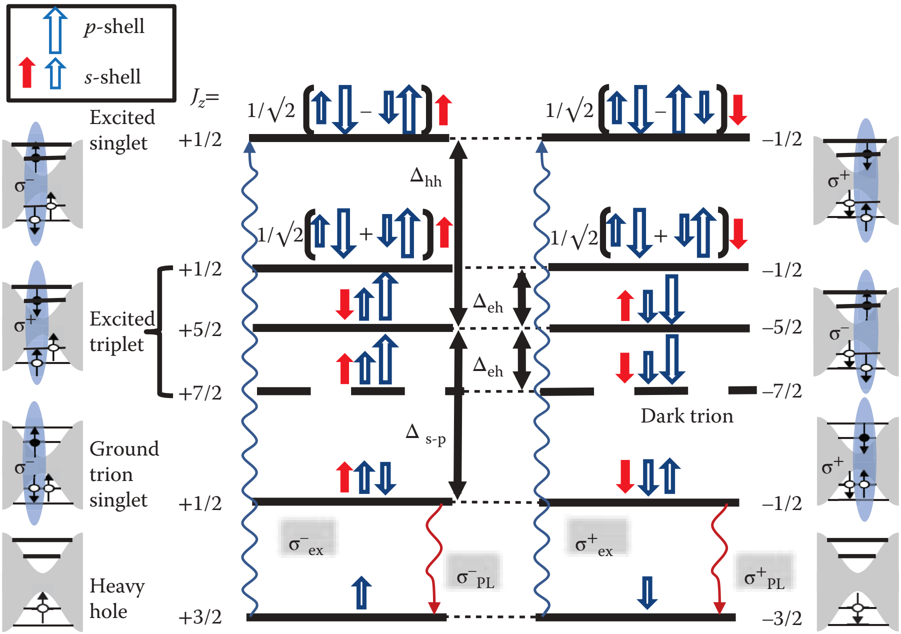


FIGURE 49.24 Energy levels and spin configurations of the hot X^{+*} trion states under σ^- (left six-level system) and σ^+ (right six-level system) optical excitation. The excited singlet is energetically split from the triplet by Δ_{hh}; the excited triplet states are separated from each other by energy Δ_{eh}. The levels are labeled with the total spin projection, J_z, of the electrons and holes. [Small solid arrows, s-shell electron-spin projection; small open arrows, s-shell hole-spin projection; large open arrows, p-shell hole-spin projection; upward wavy arrows, either σ^- or σ^+ excitation; downward wavy arrows, photoluminescence (PL); shaded ellipses, exciton coupling in the radiative states; Δ_{s-p}, energy separation between p-shell and s-shell trion states.]

Spin decoherence mechanisms that need to be considered include the electron–hole anisotropic exchange interaction (AEI), which remains relevant for excited trion states, and the hyperfine interaction between nuclei and either electrons [72] or holes through dipole–dipole interaction [73]. The two degenerate six-level systems for σ^- and σ^+ excitation are coupled through transverse spin decoherence mechanisms, as shown in Figure 49.25.

The system Hamiltonian of a circularly polarized pulse resonantly coupled to an ensemble of six-level resonant absorbers with density N_a is given by

$$\hat{H}^{\mp} = \hbar \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & -\frac{1}{2}\left(\Omega_x \mp i\Omega_y\right) \\ 0 & \omega_0 - \Delta_{hh} - \Delta_{sp} & 0 & 0 & 0 & 0 \\ 0 & 0 & \omega_0 - \Delta_{hh} - \Delta_{eh} & 0 & 0 & 0 \\ 0 & 0 & 0 & \omega_0 - \Delta_{hh} & 0 & 0 \\ 0 & 0 & 0 & 0 & \omega_0 - \Delta_{hh} + \Delta_{eh} & 0 \\ -\frac{1}{2}\left(\Omega_x \pm i\Omega_y\right) & 0 & 0 & 0 & 0 & \omega_0 \end{pmatrix}, \quad (49.50)$$

where −/+ correspond to σ^-/σ^+ polarization, $\Omega_x = \wp\frac{E_x}{\hbar}$ and $\Omega_y = \wp\frac{E_y}{\hbar}$ are the time-dependent Rabi frequencies associated with the E_x and E_y electric-field components, and ω_0 and $\wp$ are the resonant transition frequency and the optical dipole matrix element of the ground to excited singlet state transition ($|1\rangle \rightarrow |6\rangle$).

In the presence of relaxation processes (Figure 49.25), under time-dependent external perturbation, e.g., a laser pulse, the time evolution of the six-level quantum system is governed by a master equation,

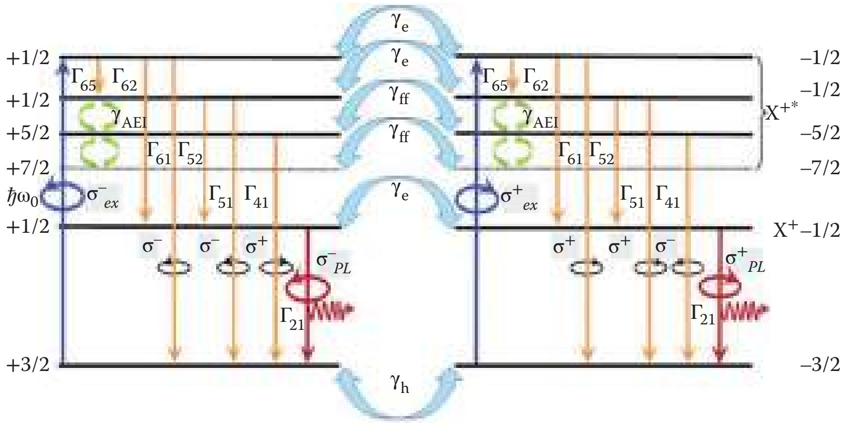


FIGURE 49.25 Discrete-level model of the X^{+*} states in a p-doped QD. The two degenerate six-level systems are coupled via spin decoherence mechanisms represented by double-headed arrows. [γ_h, hole-spin decoherence; γ_e, electron-spin decoherence; γ_{ff}, spin decoherence due to spin flip-flop processes, i.e., anisotropic electron-hole exchange interactions (AEI); ↺ or ↻, circularly polarized transitions; upward arrows, σ_{ex}^- and σ_{ex}^+ resonant pumping; downward arrows, radiative or nonradiative transitions; downward arrows denoted Γ_{21}, detected polarized PL; vertical small curved arrows, spin flip-flop coupling due to AEI.]

Equation 49.12, for the $N^2 - 1 = 35$-dimensional real state pseudospin vector, S_j, coupled to Maxwell's curl equations, Equation 49.43 [74], through the macroscopic polarization induced by the electromagnetic pulse:

$$\begin{aligned} P_x &= -\wp N_a S_5 \\ P_y &= \mp \wp N_a S_{20} \end{aligned} \qquad (49.51)$$

where −/+ again correspond to σ^- or σ^+ excitation, respectively.

The QD samples modeled are chosen to be identical to the ones used in TRPL experiments, namely MBE grown on a semi-insulating GaAs (001) substrate. The QD layers are sandwiched between two GaAs barriers and the dot areal density is approximately 2×10^{10} cm^{-2} with an average uncapped height of 4 nm, equivalent to a volume dot density, $N_a = 5 \times 10^{22}$ m^{-3}. The simulation domain consists of an InAs QD layer with nominal thickness given by the height of the typical QD (i.e., 4 nm) embedded between two GaAs barrier regions, each with thickness 50 nm (see Figure 49.26).

The circularly polarized Gaussian pulse at the left boundary of the simulation domain, $z = 0$, is modeled using two x and y linearly polarized waves with resonant carrier frequency, ω_0, phase-shifted by $\pi/2$ according to Equation 49.46 and bottom row of Equation 49.47. The system of equations is discretized in space and time, with spatial and temporal steps, $\Delta z = 1$ Å and $\Delta t = 3.33 \times 10^{-4}$ fs, respectively, and solved numerically directly in the time domain using the FDTD technique.

For comparison with the theory, PL experiments with short (i.e., 50 ps) σ^- or σ^+ optical excitation pulses with pumping wavelength $\lambda_{\text{res}} = 1065$ nm inducing a resonant transition to the excited dot states were carried out on QD ensembles nominally doped with one hole. The polarized PL was detected at the ground singlet X^+ transition, exhibiting a peak at $\lambda_{\text{det}} = 1148$ nm. In our simulations, an optical pulse of duration 50 ps is injected at $z = 0$. The pulse center frequency, ω_0, is tuned in resonance with the energy splitting between the ground heavy-hole level $|1\rangle$ and the excited singlet trion state $|6\rangle$: $\Delta E_{1\to6} = \Delta_{hh} + E_{X^+} + \Delta_{\text{s-p}}$, where $\Delta_{\text{s-p}}$ is the energy separation between the s- and p-shell trion states. The ground trion singlet energy, $E_{X^+} = 1.0815$ eV, is determined from the resonant PL spectra at λ_{det}. $\Delta_{\text{s-p}} \approx 73$ meV is inferred from the PL spectra [70], and $\Delta_{hh} = 12$ meV is taken in agreement with [71]. In order to calculate the Hamiltonian in Equation 49.50, $\Delta_{eh} \approx 0.5$ meV is used, as obtained in [71,75].

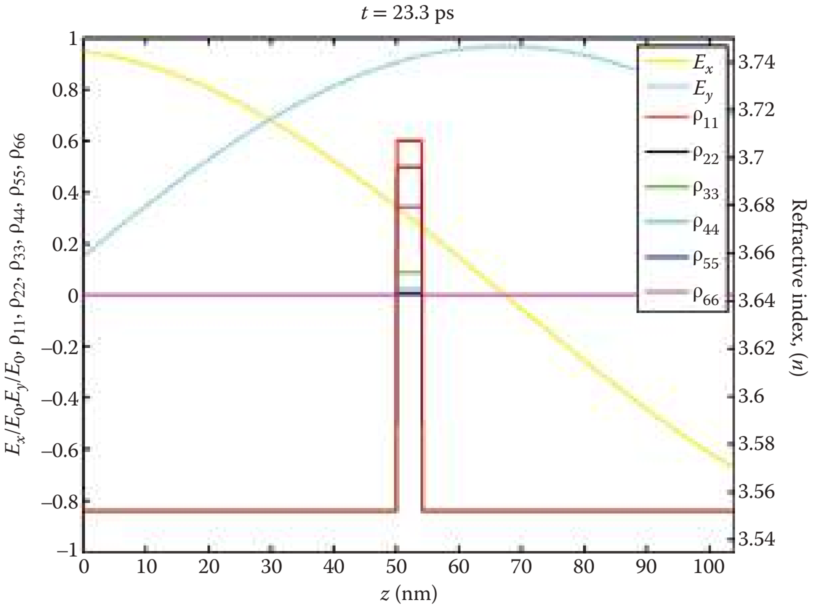


FIGURE 49.26 A snapshot of the spatial distribution of the circularly polarized electric-field components of an excitation pulse interacting with a six-level quantum system. The timing of the snapshot is $t = 23.3$ ps after the source pulse impinges on the left boundary ($z = 0$) of the medium incorporating the quantum system. The charge-carrier spin populations (diagonal density-matrix elements) of all six levels are shown; these indicate the fractional occupancy of each level compared to the total number of charge carriers spread across the different levels. The refractive-index profile of the simulation domain (bottommost curve, right-hand side scale) is produced by an InAs QD layer of thickness 4 nm embedded between two 50-nm-thick GaAs barrier regions.

The pulse area is chosen to be π, so that the pulse completely excites the ground-state spin population to the excited singlet level $|6\rangle$; this corresponds to choosing $E_0 = 2.69 \times 10^5\ \mathrm{V \cdot m^{-1}}$, assuming a dipole matrix element $\wp = 9.83 \times 10^{-29}\ \mathrm{C \cdot m}$. The dipole matrix element's value is comparable to the dipole moment of the X^+ ground singlet transition, and is calculated using the Fermi Golden Rule [50] from the energy separation between the levels and the spontaneous emission rate $\Gamma_{61} \sim 1.2\ \mathrm{ns^{-1}}$, which is close to the experimentally observed value.

The longitudinal relaxation times are taken as follows: the radiative spontaneous decay, $\tau_{21} \approx 1.27$ ns, is experimentally measured from a sample with one hole; spontaneous emission times, $\tau_{41} = 1.35$ ns and $\tau_{51} = \tau_{61} = 1.2$ ns are estimated from the energy-level separation using the Wigner–Weisskopf formula [50]. The nonradiative decay times are taken from theory [76], the transverse spin decoherence rates are obtained from experiment: $\tau_e = 500$ ps [72] and $\tau_h = 14$ ns [73], and a theoretical estimate for $\tau_{ff} = 125$ ps is made on the basis of the Heisenberg uncertainty relations. We should note that there are only two adjustable parameters in our model—namely the largely unknown nonradiative spin relaxation times, τ_{52} and τ_{65}.

The theoretically computed TRPL traces that are displayed in Figure 49.27 clearly exhibit photo-induced circular dichroism in the polarized TRPL emission; this is manifested as a different evolution of PL in time, depending on the pulse helicity used. In contrast to the widely used rate equations model which assumes mono-exponential decays, the dynamics of the quantum system described by the present model is nonlinear and, as a result, the simulated PL decay is not mono-exponential.

The experimentally detected TRPL trace shape strongly depends on detector characteristics: the faster the detector's response and the higher its sensitivity, the better the agreement between theory and

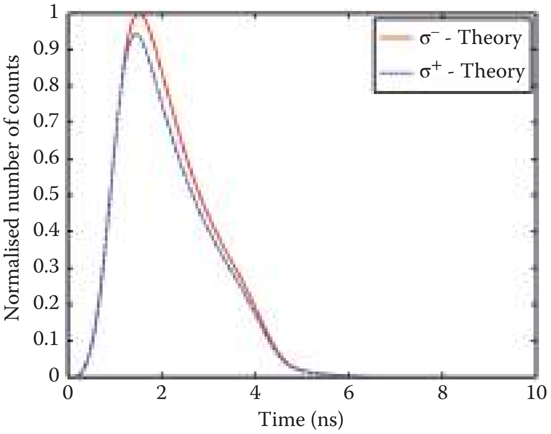
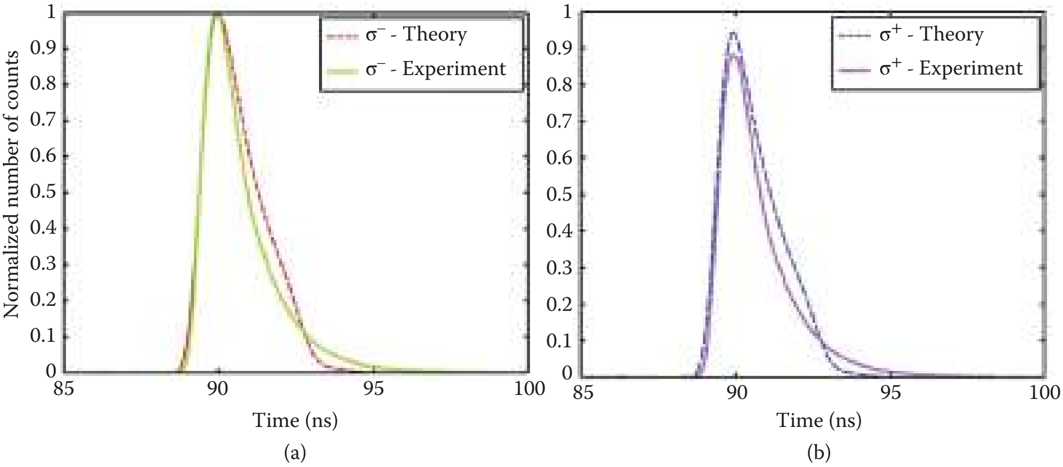


FIGURE 49.27 Theoretically computed polarized TRPL traces for σ^- and σ^+ excitation of a six-level QD. The presence of circular dichroism is clearly visible.



(a)

(b)

FIGURE 49.28 Comparison between the simulated TRPL trace (after convolution with the detector's response function) and experimentally detected TRPL trace for (a) σ^- excitation and (b) σ^+ excitation of a six-level QD.

experiment. For proper comparison with experiment, we convolve our theoretically calculated polarized traces with the detector's response function and plot them along with the experimental traces in Figure 49.28.

There is excellent agreement between the theoretical and the experimental results. The rising slope and initial decay after the peak in the experimental trace are almost perfectly reproduced by our simulations. The small discrepancy at longer times can be attributed to several factors. Firstly, it is not possible during the course of an experiment to determine precisely which hot trion state is being resonantly excited, so the assumed λ_{res} is only approximate. Secondly, Δ_{s-p} is inferred from the resonant PL spectra and is only a rough estimate, since the excited state does not correspond to a sharp spectral line, but rather a broadened line of finite width; in addition, the inhomogeneous broadening has not been taken into account. The lower peak height for σ^+ excitation can be attributed to the estimated value for the spin flip-flop coupling rate and requires further investigation.

Our theoretical model explains quantitatively the origin of the spin-filtering effect. The experimentally observed enhanced time-resolved circular dichroism in the excited trion state emission is shown to emerge from the dynamical spin flip-flop coupling between the two degenerate sets of excited charged excitonic states. The theory is in very good agreement with experiment, thereby allowing to obtain an estimate for largely unknown intra- and intershell spin-relaxation timescales. The approach also allows to predict optimum pulse parameters, such as power and duration, for efficient control of spin dynamics.

49.6 Quantum Stochastic Formalism for Modeling Cavity-Emitter Systems

Future advances in information systems are expected to exploit the quantum nature of their constituent parts—e.g., in the development of quantum simulators and design of architectures for quantum cryptography. The implementation of these technologies necessitates a fuller understanding of the underlying quantum systems, which in turn will require several fundamental aspects of quantum theory to be tested.

One example of a quantum-optical phenomenon is quantum noise. In contrast to the more familiar classical noise, which can be regarded as spontaneous random fluctuations from a steady state, quantum noise arises from the Heisenberg uncertainty relations, which are fundamental to quantum theory [77]. The concept of quantum noise stems from the statistical interpretation of quantum mechanics and our inability to access any particular one out of the infinite number of degrees of freedom of the electromagnetic field propagating in free space. Quantum noise is closely related to spontaneous emission, irreversible decay and the origin of the spectral line width described by the theory of Weisskopf and Wigner [78], who are considered to have laid the foundations of the quantum noise theory.

At large photon densities, quantum fluctuations are necessary to explain laser linewidth and the threshold properties. By contrast, at low photon densities, light is generated by spontaneous emission described in terms of single-photon emission events within the single-particle Dirac [79] picture. Control over such noise effects will be crucial in the design of devices for quantum information processing.

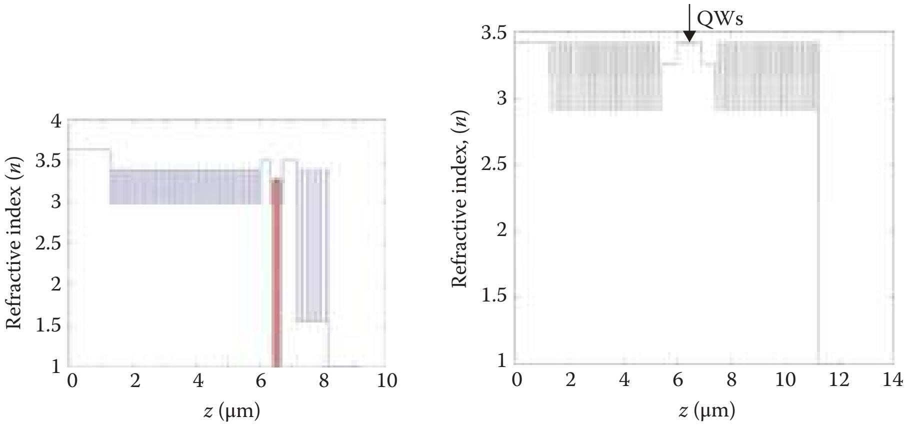
49.6.1 Langevin Formalism for Modeling Quantum-Optical Effects

Considering a two-level system with a few degrees of freedom immersed in a radiation-field heat bath with an infinite number of degrees of freedom, we exploit the quantum-classical correspondence in the presence of noise and employ the Langevin formalism. The essence of the Langevin approach lies in the realization that the action that the many unknown "bath" (i.e., external) variables have on the system is to modify its deterministic equation of motion, and that this can be accommodated by the inclusion of apparently random terms, known as "Langevin forces."

49.6.2 Spontaneous Emission of a QD Ensemble in a Semiconductor Microcavity and Onset of Lasing

Consider the semiconductor microcavity structures shown in Figure 49.29, designed at $\lambda = 850\,\text{nm}$ and $\lambda = 1.29\,\mu\text{m}$. The cavity of the first structure (bottom DBR: 35.5 pairs $\text{AlAs}/\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$; top DBR: five pairs $\text{AlO}/\text{Al}_{0.3}\text{Ga}_{0.7}\text{As}$) is filled with a resonant two-level medium ($\text{Al}_{0.5}\text{Ga}_{0.5}\text{As}$), whereas within the second cavity, there are six ($\text{Ga}_{0.63}\text{In}_{0.37}\text{N}_{0.012}\text{As}$) quantum wells filled with the resonant medium. To describe the interaction of an optical wave with these systems, we use the semiclassical 1D Maxwell–Bloch equations in the real vector pseudospin picture, Equation 49.26, for an optical wave linearly polarized along x propagating in the z-direction. In order to model the spontaneous emission of two-level atoms, we add a random electric-field fluctuation, δE_x, at each time step:

$$\left(E_x(z,t) + \delta E_x(z,t)\right) = -\frac{1}{\varepsilon}\frac{\partial H_y}{\partial z} - \frac{N_a \wp}{\varepsilon T_2} S_1 + \frac{N_a \wp \omega_0}{\varepsilon} S_2 . \tag{49.52}$$



(a) (b)

FIGURE 49.29 Refractive-index profile along a semiconductor microcavity (a) filled with gain medium (shaded area) with $n = 3.2736$ at $\lambda = 850$ nm; (b) with 6 GaInNAs-QWs, filled with gain medium of refractive index $n = 3.4$ at $\lambda = 1.29$ μm. (G. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2004 IEEE.)

We use a pseudorandom number generator based on the Box–Müller method for generating random deviates with a normal (Gaussian) distribution from uniformly distributed in the interval (0, 1) random numbers a and b. At each time step, j, white Gaussian noise with variance $\xi_E = \sigma^2 = 1 \times 10^{-3}\ \mathrm{V^2m^{-2}}$ is implemented according to

$$E_x(z)_j = E_x(z)_j + \sqrt{-2\xi_E \ln(a)}\cos(2\pi b). \tag{49.53}$$

White Gaussian noise has a mean zero and δ-correlated second moment of the random distribution:

$$\begin{aligned} \langle \delta E_x(z,t)\rangle &= 0 \\ \left\langle \delta E_x(z,t)\,\delta E_x\left(z,t'\right)\right\rangle &= \tilde{\xi}_E \delta\left(t-t'\right) = \xi_E R_{sp}\delta\left(\frac{t-t'}{\Delta t}\right), \end{aligned} \tag{49.54}$$

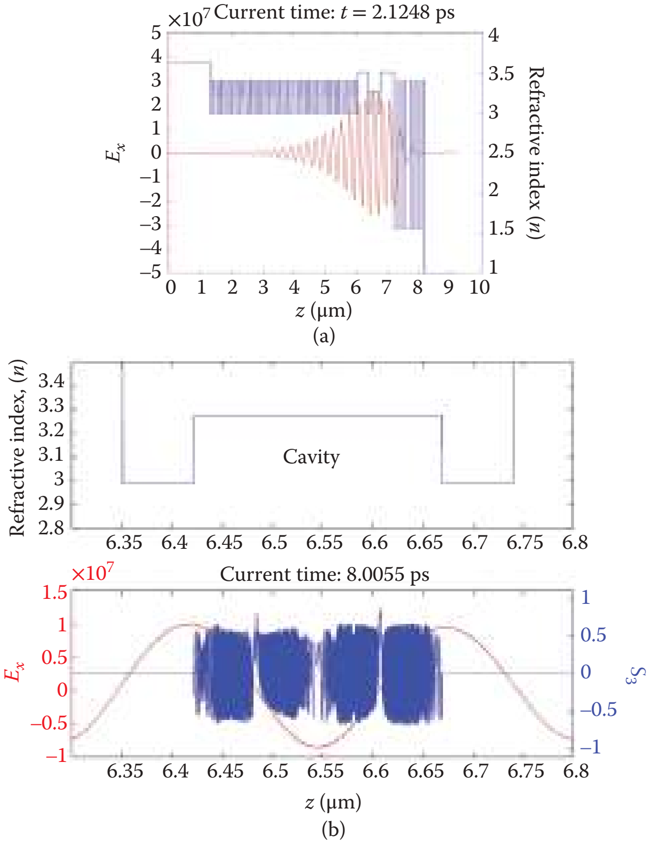
where an expression for the spontaneous emission rate has been derived:

$$R_{\mathrm{sp}} = \frac{\xi_E(\varepsilon_0\varepsilon)^2}{N_a^2 \wp^2 T_2^2}. \tag{49.55}$$

Consider the first microcavity: we assume that the resonant medium is initially inverted ($S_{30} = 1$), thus providing gain, and apply solely random white Gaussian noise at each point within the cavity according to Equation 49.53. The spatial distribution of the intracavity electric field as a function of time, and the time evolution of the electric field at the microcavity structure's output boundary are calculated. The parameter set used for this simulation is $\wp = 4.8 \times 10^{-28}\ \mathrm{C \cdot m}$, $n = 3.2736$, $N_a = 1 \times 10^{24}\ \mathrm{m^{-3}}$, initial field amplitude, $E_0 = 700\ \mathrm{V \cdot m^{-1}}$, and relaxation and dephasing times, $T_1 = 10$ ps and $T_2 = 70$ fs, respectively. The cavity field enhancement provided by the DBR mirrors leads to build-up of the electric field within the cavity borne out of the quantum noise. A snapshot of the intracavity electric-field build-up is shown in Figure 49.30, along with the corresponding Rabi flopping of the population dynamics. The amplitude that has arisen solely due to noise gradually builds up towards a standing-wave mode in the cavity.

After this initial amplitude build-up process, the lasing threshold is reached, and coherent oscillations appear at the output boundary. This is the onset of the lasing regime, shown for the unsaturated

$$\Omega_R^2 T_1 T_2 \approx 1.59 \tag{49.56}$$



(a)

(b)

FIGURE 49.30 (a) Stationary electric-field standing-wave profile arisen from quantum noise in the cavity. (b) (Top) Refractive-index profile within the cavity and in its vicinity. (Bottom) A snapshot of the spatial distribution of the intracavity E_x electric field (left axis, slowly-varying sinusoidal curve) and the corresponding population inversion (right axis, fast-oscillating curve) exhibiting fast Rabi oscillations at $t = 8$ ps after the instant when the quantum noise arises. These oscillations result in the fast modulation of the electric-field envelope observed at the output boundary of the structure, which is shown in Figure 49.31a and b.

and saturated

$$\Omega_R^2 T_1 T_2 \approx 16241 \gg 1 \tag{49.57}$$

cases in Figure 49.31 (a,b), respectively. The charge-carrier population performs multiple Rabi flops, oscillating about a steady population inversion value close to zero. The net gain and absorption are therefore similar, resulting in a gain saturation in time.

Let us now consider the structure designed at $\lambda = 1.29\,\mu$m. The population in all six quantum wells is initially inverted into the upper state, $S_{30} = 1$, representing a nonlinear gain medium, while the barrier regions within the cavity are assumed to be in the ground state, $S_{30} = -1$. Random noise is generated within the cavity with a parameter set corresponding to the unsaturated case above (49.56). In Figure 49.32, a snapshot of the spatial distribution of the electric-field amplitude and the steady-state population inversion are shown. The population inversion relaxes to this value after a sufficiently large number of time steps. The time evolution of the electric field at the output boundary is shown in Figure 49.33, showing continuous build-up of coherent self-sustained oscillations and lasing. The lasing threshold is clearly discernible, and the electric-field envelope exhibits rapid damped relaxation oscillations with a decay rate,

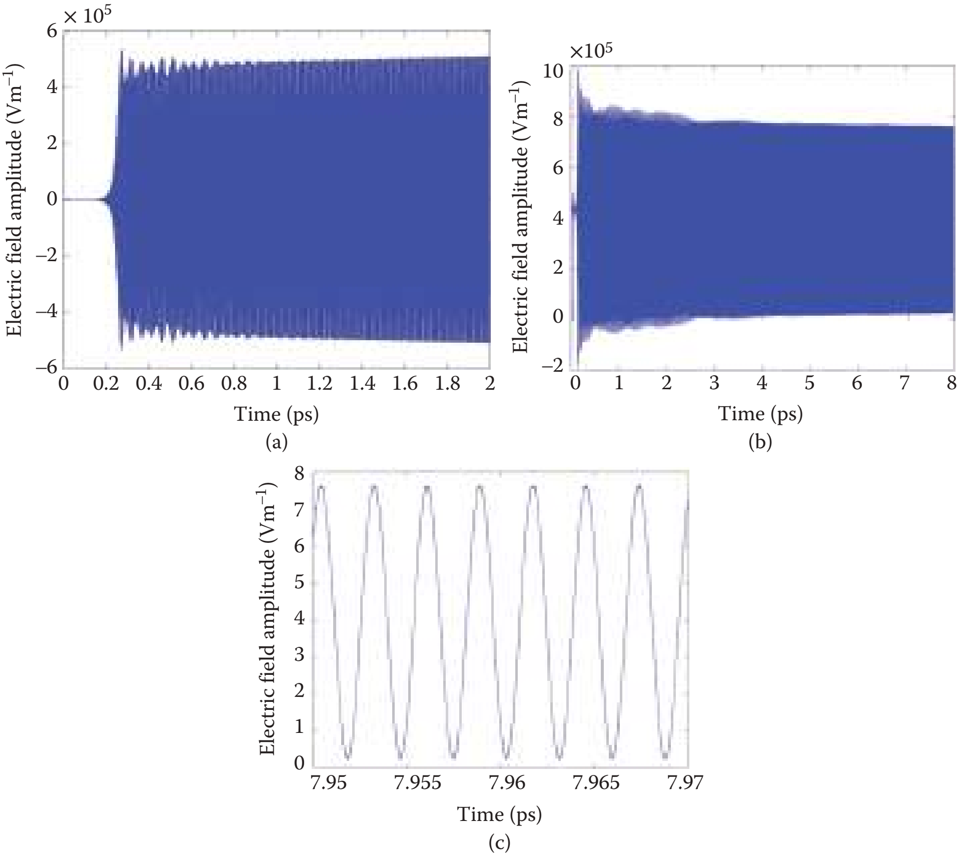


FIGURE 49.31 Time evolution of the electric field at the output boundary of a semiconductor microcavity, showing the noise-induced build-up of coherent oscillations and onset of lasing of the structure displayed in Figure 49.29a. Simulations were run for two different sets of relaxation and dephasing times (T_1 and T_2, respectively). (a) $T_1 = 10$ ps, $T_2 = 70$ fs; (b) $T_1 = 1$ ns, $T_2 = 10$ ps. (c) Expanded view of (b) in the steady-state region, showing single-mode oscillations and lasing at a frequency corresponding to $\lambda = 850$ nm. (G. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2004 IEEE.)

$\Gamma_R \approx 1.5 \times 10^{11}$ s^{-1}. These rapidly varying oscillations are a new feature which is not predicted by the usual rotating-wave (RWA) and slowly-varying envelope (SVEA) approximations. The oscillations eventually reach the steady-state gain saturation value, $S_{3sat} = 0.0025$, at simulation times exceeding 16 ps.

We perform a Fourier transform of the output electric-field trace in Figure 49.33 and obtain the laser spectral linewidth, shown in Figure 49.34. The obtained linewidth exhibits superfine structure: the main mode is centered at a frequency ω_0 corresponding to $\lambda = 1.289$ μm, and the wings exhibit satellite peaks at $\omega_0 + n\Omega_{rel}$. The frequency of the fast relaxation oscillations, $f_{rel} \approx 2.54$ THz, is determined from the figure.

We showed that an extension within the Langevin framework of our Maxwell-pseudospin model predicts the build-up of coherent, self-sustained oscillations and the onset of lasing, which is a sole consequence of noise being introduced into the cavity. This emphasizes the importance of quantum fluctuations for triggering the lasing regime of cavity operation. The simulations allow to investigate the lasing threshold behavior directly in the time domain and provide an estimate for the coherence time of the laser emission and the laser linewidth. We showed that several important parameters can be extracted from the simulations and that they can be optimized using numerical experiments on a variety of design geometries. Most importantly, the developed quantum stochastic approach is capable of modeling quantum effects, such as

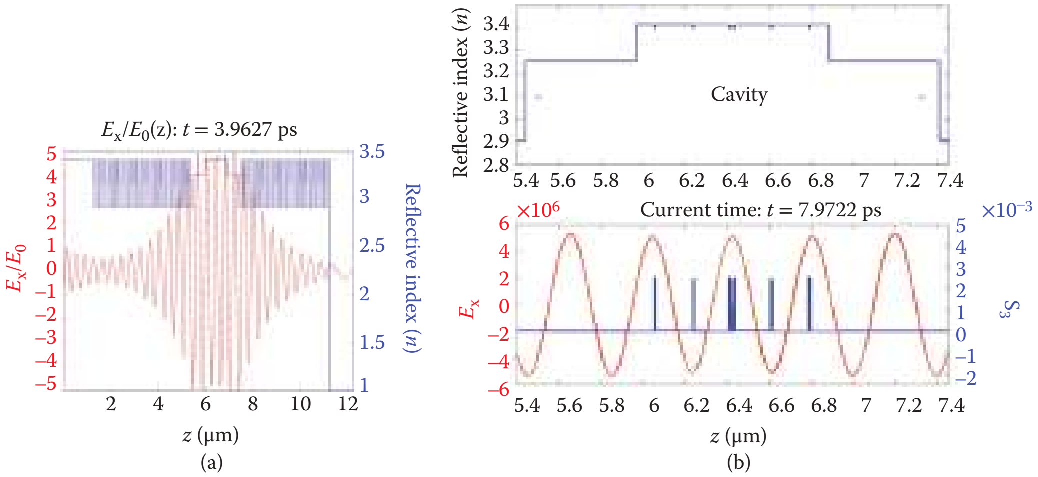


FIGURE 49.32 (a) Instantaneous profile of the electric-field standing wave borne out of quantum noise in the cavity. (b) (Top) Refractive-index profile implying the cavity boundaries and the location of the quantum wells. (Bottom) Spatial distribution of the stationary electric-field standing wave (left axis, sinusoidal curve) and the corresponding saturated population inversion, $S_{3\text{sat}} = 0.0025$, in the quantum wells within the cavity (right axis, spiky curve). (G. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2004 IEEE.)

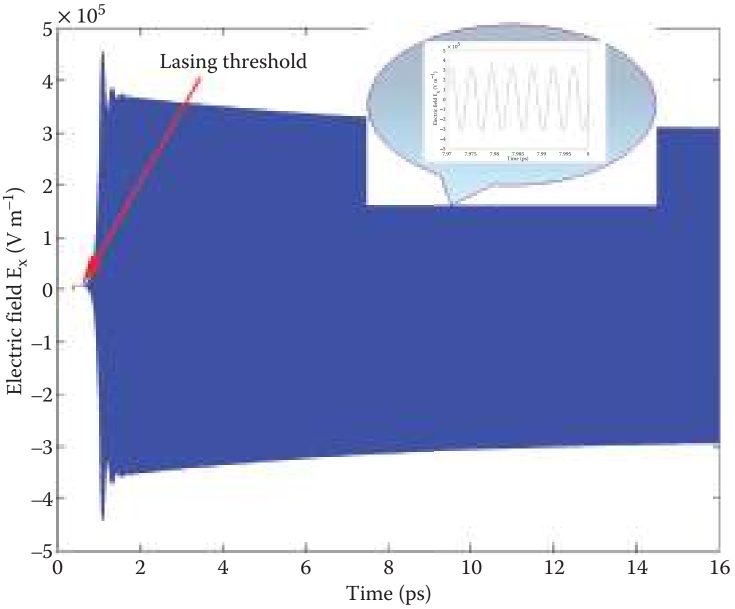


FIGURE 49.33 Build-up in time of coherent, self-sustained oscillations arisen from noise introduced into the cavity of the structure displayed in Figure 49.29b. The lasing threshold is apparent. The decay rate of the electric-field envelope is $\Gamma_R \approx 1.5 \times 10^{11}\ \text{s}^{-1}$. Inset: expanded view of the fast carrier-frequency oscillations with $\lambda = 1.289$ μm; the steady-state region occurs for $t > 16$ ps. (G. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2004 IEEE.)

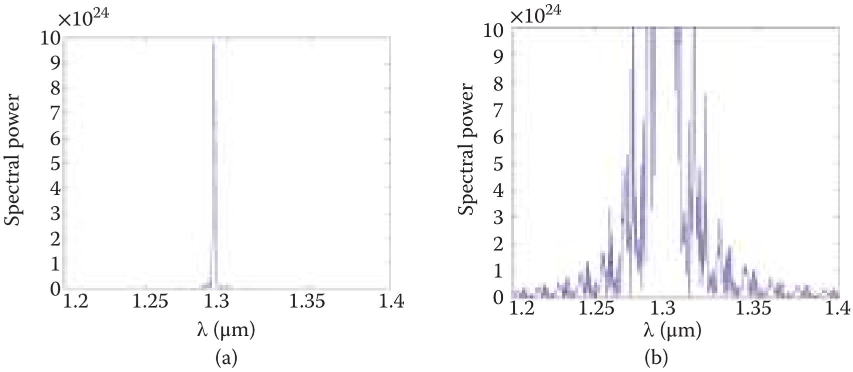


FIGURE 49.34 (a) Laser spectral line obtained by Fourier-transforming the output trace of the electric-field amplitude centered at the lasing frequency, corresponding to $\lambda = 1.289\,\mu\text{m}$. (b) Magnified view of the spectrum in (a) showing relaxation oscillation sidebands. (G. Slavcheva et al., *IEEE Journal of Selected Topics in Quantum Electronics* © 2004 IEEE.)

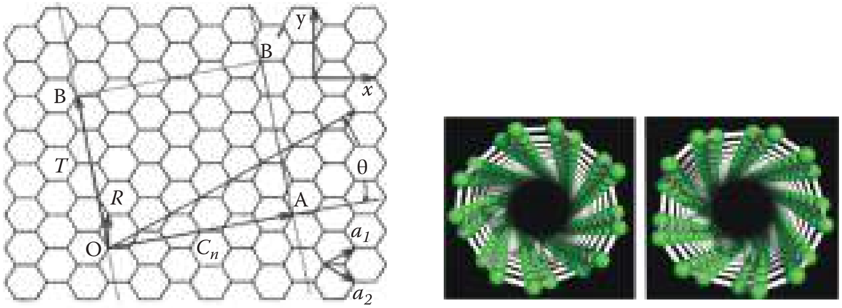
spontaneous emission, and thus can be used to model the system dynamics in the few-photon limit as well as demonstrate and predict quantum-optical effects.

49.7 Modeling Optical Rotation in Chiral Carbon Nanotubes with Master Maxwell-Pseudospin Equations

In this section, we shall employ our vector Maxwell-pseudospin model [44] to investigate how chirality affects the ultrafast nonlinear optical and magneto-optical response of a single chiral CNT [80]. We shall show that a simple discrete-level model of the optically active states near the bandgap edge of a chiral nanotube is sufficient to explain the circular dichroism and birefringence, as well as to quantify the optical rotation per unit length (also known as specific rotatory power), either in the absence or in the presence of an external magnetic field permeating the tube. The model predicts a giant natural gyrotropy for the specific chirality considered in this case, which is comparable with or exceeding that of artificially fabricated helical photonic structures. This is remarkable because the latter are considered to achieve the largest optical rotation of all known materials: orders of magnitude larger than that of crystal birefringent materials or of liquid crystals. Finally, we will give a quantitative estimate of the nonlinear coherent magneto-chiral optical effect in an axial magnetic field.

Chirality is one of the main symmetries of the CNT geometry that determines the optical properties of single-walled CNTs (SWCNTs). SWCNTs are uniquely determined by the chiral vector, $C_h = na_1 + ma_2 \equiv (n, m)$, with $0 < |m| < n$, or equivalently by a pair of integer numbers (n, m) in the planar graphene hexagonal lattice unit-vector basis (Figure 49.35). Two helical forms (enantiomers), e.g., AL (5, 4) and AR (4, 5), can exist. Topologically, the CNT can be viewed as a graphene sheet rolled up along the chiral vector into a cylinder. Chirality serves as a primary classification criterion. Achiral nanotubes, whose mirror image is superimposable, are subdivided into two classes: zigzag ($m = 0$) and armchair ($m = n$) nanotubes. The rest of the nanotubes, whose mirror reflection is not superimposable, are subsumed into the most general class of chiral nanotubes.

There is a fundamental relationship between chirality and optical activity. The interaction of polarized light with chiral materials gives rise to the phenomenon of optical rotation (or optical activity), whereby the polarization plane is rotated continuously during the propagation of the light through the nanotube. In the presence of an external magnetic field, magnetically induced optical activity—also known as Faraday effect—takes place. Both effects manifest themselves as a rotation of the transmitted light; however, the origin of the two effects is fundamentally different. While the natural optical activity is a result of



(a) (b) (c)

FIGURE 49.35 (a) Chiral vector, C_h (C_n on schematic) and chiral angle, $\theta = \angle\left(a_1, C_h\right)$, $0° < |\theta| < 30°$, where a_1 and a_2 are the elementary Bravais cell unit vectors. (b) Left-handed AL $(5, 4)$ (with $m > n$; m, n – integers, see text) and (c) right-handed AR $(4, 5)$ (with $m < n$) single-walled CNT molecular structure calculated by the tight-banding method [81]. Shown is a view along the tube axis looking in the negative z-direction. Chirality is determined by the rotation of the so-called armchair chains of carbon atoms either clockwise or counterclockwise when looking in the negative z-direction.

the nonlocal optical response of a medium lacking mirror symmetry, the magnetic optical activity results from time-reversal symmetry breaking by the magnetic field. The two phenomena are linked through the magneto-chiral optical effect which takes place when both symmetries are broken simultaneously [82].

The electronic band structure of an SWCNT is described by the quantization of the wavevector along the tube circumference perpendicular to the tube axis. This results in a discrete spectrum of allowed k-vector states forming subbands in the valence and conduction bands labeled by the quasi-angular momentum quantum number, μ [83]: $k_\perp = \frac{2\pi\mu}{L}$, $\mu = 0, \pm 1, \pm 2, \ldots$, where $L = \left|C_n\right|$ is the tube circumference (Figure 49.36a). We should note that due to symmetry in the graphene hexagonal unit cell, there are two equivalent points, K (with $\mu > 0$) and K′ (with $\mu < 0$) in the BZ (Figure 49.36c). The energy dispersion of a CNT can be visualized as intersections of the graphene conical dispersion at the quantized transverse wavevector, $k_\perp$ (Figure 49.36b).

The most widely studied excitation geometry in experiments on individual CNT is the one of linearly polarized excitation perpendicular to the tube axis, shown in Figure 49.37a. In this excitation geometry, the longitudinal E_z electric-field component is dominant, because the transverse E_x component is suppressed by depolarization effects, and linearly polarized intersubband optical transitions from the valence μ subband to the conduction μ subband with $\Delta m = 0$ can be excited (here m signifies the eigenvalue of the quasi-angular momentum operator projection, $\hat{J}_z$). We shall be interested, however, in the unconventional circularly polarized excitation geometry shown in Figure 49.37b, whereby a circularly x–y plane-polarized laser pulse propagates along the tube axis, exciting only one of the two allowed transitions, $\mu \to \mu \pm 1$, where "$+/-$" corresponds to right and left circularly polarized light, respectively (Figure 49.38b). The molecular structure and the allowed dipole optical intersubband transitions of a $(20, 10)$ chiral CNT are shown in Figure 49.38.

We shall consider aggregates or bundles of aligned SWCNTs grown by chemical-vapor deposition in an electric field [84], or aligned in a polymer matrix [85]. We model the single carbon-nanotube bandgap-edge structure at the K point of the BZ (where the fundamental bandgap opens) by a large ensemble of identical four-level systems, corresponding to the allowed transitions of the two valence subbands closest to the Fermi level for the AL and AR nanotube enantiomers shown in Figure 49.39. We note that the energy-level diagrams for the left- and right-handed helical forms in Figure 49.39 are nonsuperimposable: absorption of σ^+ light excites the $\mu \to \mu + 1$ transition in AL-handed SWCNT, while absorption of light with the same helicity excites the $\mu \to \mu - 1$ transition in AR-handed SWCNT. The difference between the optical selection rules for left and right circularly polarized light gives rise to the optical activity [86].

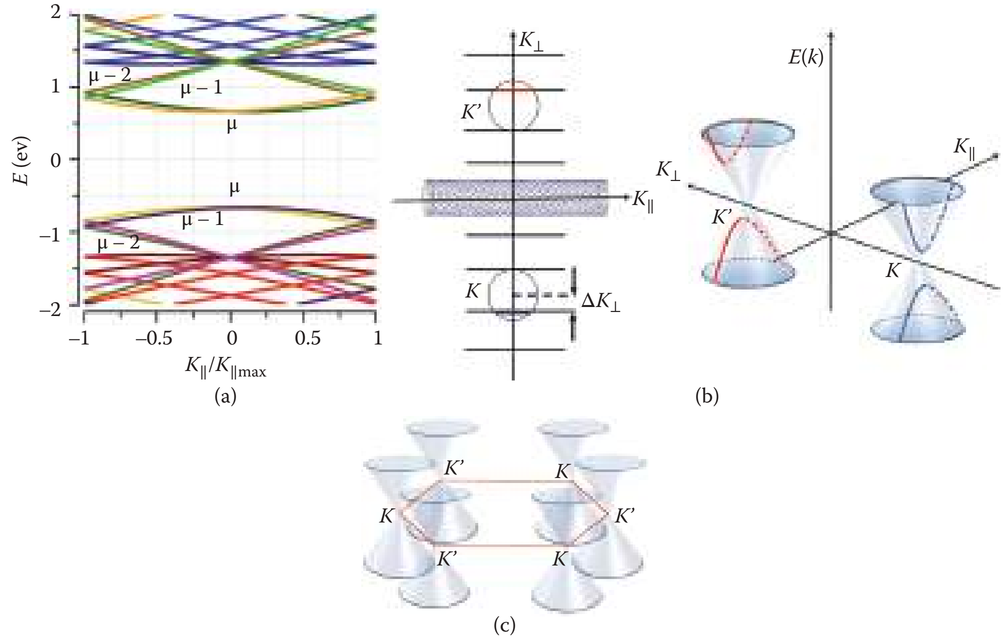


FIGURE 49.36 (a) Energy dispersion as a function of the longitudinal wavevector, $k_{||}$, normalized with respect to the wavevector at the boundary of the first BZ of a (5, 4) SWCNT. $k_{||}$ points along the tube axis. The tube for this chirality is semiconducting—i.e., a bandgap is opened at the center (Γ point) of the BZ. (b) Parabolic cross sections of the graphene linear dispersion, corresponding to points K and K′ in (a) representing the nanotube energy dispersion curves. (c) Unfolded 2D first BZ of graphene, showing the equivalent points K and K′ in the band structure.

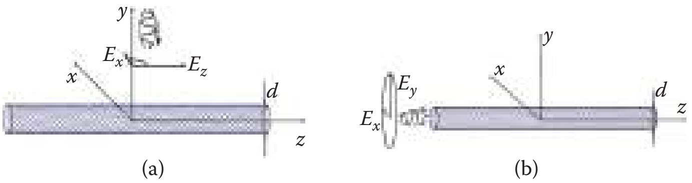
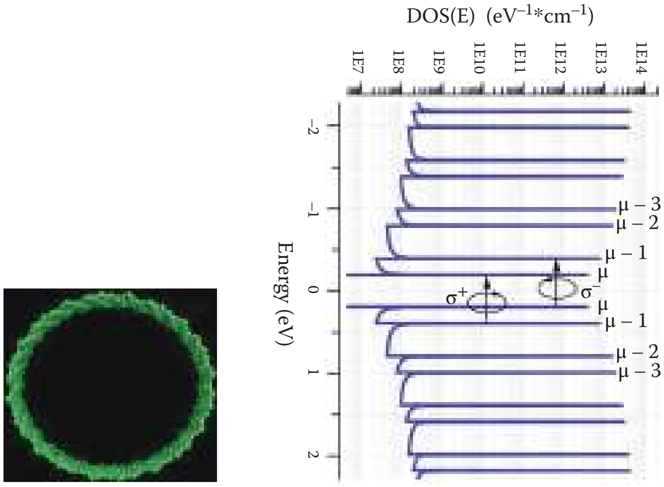


FIGURE 49.37 (a) Linearly polarized excitation perpendicular to the tube axis exciting intersubband optical transitions with $\Delta m = 0$. (b) Circularly polarized excitation along the tube axis exciting $\Delta m = \pm 1$ optical transitions.

49.7.1 Natural Optical Activity

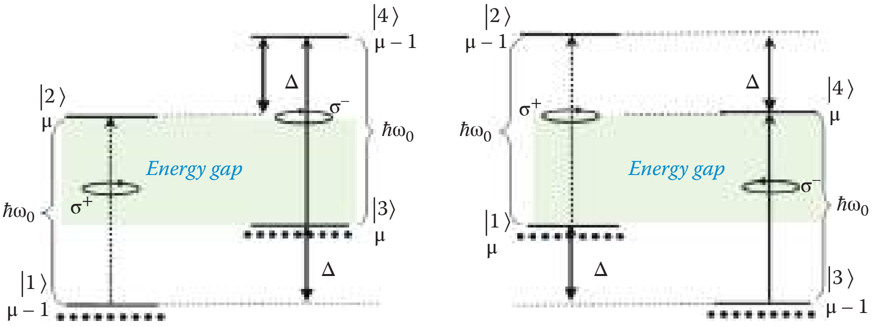
In this section, we compute the natural optical activity rotation angle (rotatory power per unit length) in the absence of an external magnetic field and demonstrate the possibility to manipulate it by engineering the nanotube chirality. We consider the specific case of a chiral, left-handed AL (5, 4) SWCNT, which is chosen for illustration of the general method valid for an arbitrary chirality. The nanotube length is 500 nm and all edge effects will be ignored in our 1D model. We performed tight-binding calculations of the electronic band structure (Figure 49.36a) and determined the fundamental bandgap, $E_g = E_{\mu,\mu} = 1.321$ eV, corresponding to a pumping wavelength of $\lambda = 939$ nm. The tube diameter is 0.611 nm, and the chiral angle, 26.33°. The resonant transition energy for circularly polarized excitations is $E_{\mu,\mu\pm 1} = 1.982$ eV, corresponding to a resonant wavelength, $\lambda_0 = 626.5$ nm.



(a)

(b)

FIGURE 49.38 (a) Molecular structure of an AL (20, 10) left-handed chiral CNT. (b) 1D electron density of states (DOS) plotted as a function of energy at the K point of the BZ ($\mu > 0$). The allowed dipole optical transitions for circularly polarized light are denoted by arrows for left- (σ^-) and right-handed (σ^+) pulse helicity.



(a)

(b)

FIGURE 49.39 Energy-level structure at the K (or K') point of the lowest subbands labeled by the subband index μ for an (a) AL (left-handed) and (b) AR (right-handed) SWCNT. The fundamental energy gap is shaded. The dipole optical transitions excited by σ^- and σ^+ circularly polarized light are designated by arrows. Only one of the two transitions is allowed for circularly polarized light, denoted by solid arrow (by contrast, the forbidden transition is indicated by a dashed arrow). Valence-band states below the bandgap are populated. (ω_0, resonant transition frequency; Δ, energy separation between the lowest subband and the second lowest subband near the bandgap.)

We consider a circularly polarized ultrashort pulse of duration 60 fs matching experiments [87], and a pulse area of π, giving an initial electric-field amplitude $E_0 = 6.098 \times 10^8\ \mathrm{V \cdot m^{-1}}$. We inject the pulse from the left boundary of the simulation domain, consisting of the tube embedded between two free-space regions of thickness 50 nm each. The pulse is resonant with the energy $E_{\mu,\mu\pm1}$ of the allowed dipole optical transition and is coupled to an ensemble of identical, HB four-level systems (Figure 49.39a for a left-handed AL SWCNT) used to describe the optical transitions involving the lowest-energy subbands lying nearest to the bandgap edge (Figure 49.40a).

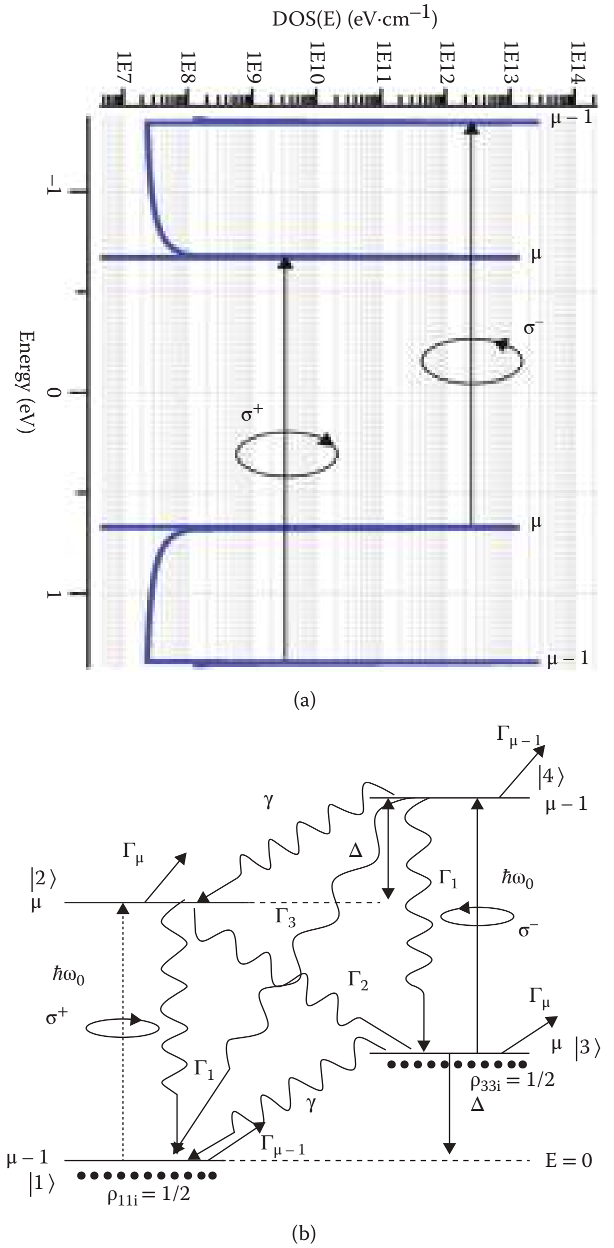


FIGURE 49.40 (a) Plot of the 1D density of states (DOS) showing the lowest-energy subbands near the bandgap edge involved in dipole optical transitions excited by circularly polarized light. Only one transition at a time can be excited by each helicity (σ^- or σ^+). (b) Energy-level diagram in the general case of a resonant optical excitation of $\mu - 1 \rightarrow \mu$ and $\mu \rightarrow \mu - 1$ interband transitions by a σ^+ and σ^- polarized pulse, respectively. The resonant transition energy is $\hbar\omega_0$. The initial population of the lowest valence states below the bandgap, ρ_{11i} and ρ_{33i}, is equally distributed between levels $|1\rangle$ and $|3\rangle$ (here, the subscript "i" denotes "initial"). (Δ, the energy separation between the first and the second lowest conduction (or valence) band states; wavy lines denote longitudinal relaxation processes between the levels, associated with population transfer; Γ_1, spontaneous emission (radiative decay) rate of $|2\rangle \rightarrow |1\rangle$ transitions, assumed to be equal to that of $|4\rangle \rightarrow |3\rangle$; Γ_2, spontaneous emission rate for the linearly polarized transition from the conduction μ subband to the valence μ subband; Γ_3, decay rate of the linearly polarized transition from the conduction $\mu - 1$ subband to the valence $\mu - 1$ subband; γ, intraband relaxation rate; Γ_μ and $\Gamma_{\mu-1}$, transverse relaxation (dephasing) rates.)

The corresponding energy-level diagram is shown in Figure 49.40b, with all coherent and spontaneous transitions indicated. Following our coherent vector Maxwell-pseudospin formalism, the system Hamiltonian for a four-level system that applies for either excitation by σ^- or σ^+ is given by

$$\hat{H} = \hbar \begin{pmatrix} 0 & -\frac{1}{2}\left(\Omega_x + i\Omega_y\right) & 0 & 0 \\ -\frac{1}{2}\left(\Omega_x - i\Omega_y\right) & \omega_0 & 0 & 0 \\ 0 & 0 & \Delta & -\frac{1}{2}\left(\Omega_x - i\Omega_y\right) \\ 0 & 0 & -\frac{1}{2}\left(\Omega_x + i\Omega_y\right) & \Delta + \omega_0 \end{pmatrix}, \qquad (49.58)$$

where $\Omega_x = \frac{\wp}{\hbar}E_x$ and $\Omega_y = \frac{\wp}{\hbar}E_y$ are the time-dependent Rabi frequencies associated with the E_x and E_y electric-field components, and $\wp$ is the optical dipole matrix element for $\mu \to \mu \pm 1$ transitions excited by circularly polarized light. This parameter is largely unknown. We have provided an estimate for it,

$$\wp \approx 3 \times 10^{-29}\ \text{C} \cdot \text{m}, \qquad (49.59)$$

based on an extension of the effective-mass method applied to chirality effects in carbon nanotubes (CNTs) [88], in agreement with the measured radiative lifetime of typical molecular transition in a single CNT [89], which is on the order of $\tau_{\text{spont}} \approx 10$ ns.

We consider a four-level resonant medium consisting of a large number of identical aligned CNTs with an average density $N_a = 6.8 \times 10^{24}\ \text{m}^{-3}$, calculated in such a way that the volume of the simulated nanotube with diameter 0.611 nm and length 500 nm contains just a single nanotube on average. A method called the effective medium approximation—used for describing the macroscopic properties of composite materials, such as the dielectric permittivity of a single CNT—has been successfully applied in FDTD modeling of the thermal radiative properties of vertical arrays of multi-walled CNTs [90]. We use two independent theoretical approaches to obtain an estimate for the effective dielectric constant of an isolated SWCNT: the aforementioned effective medium approximation [80] and tight-binding model calculations of the axial component of the imaginary part of the dielectric function tensor [91,92]. This gives a value for the refractive index, $n \approx 2.3$ along the nanotube axis. With dipole matrix element from (49.59), we use the expression for the spontaneous emission rate: $\tau_{\text{spont}} = \frac{3\pi\varepsilon_0\hbar c^3}{n\omega_0^3\wp^2}$, taking into account the energy separation of each transition, to obtain an estimate for the relaxation rates: $\Gamma_1 = 2.907\ \text{ns}^{-1}$, $\Gamma_2 = 9.812\ \text{ns}^{-1}$, and $\Gamma_3 = 1.227\ \text{ns}^{-1}$. We take the experimental value obtained in [87] for the intraband optical transitions, $\gamma = 130\ \text{fs}^{-1}$. Since the dephasing rates for the states involved are largely unknown, we treat them as phenomenological parameters adopting the following values: $\Gamma_\mu = 800\ \text{fs}^{-1}$ and $\Gamma_{\mu-1} = 1.6\ \text{ps}^{-1}$; however, our simulations show that the dynamics is largely insensitive to the choice of these dephasing rates.

We consider separately the cases of an ultrafast resonant optical excitation of the $|1\rangle \to |2\rangle$ transition with σ^+ helicity, and of the $|3\rangle \to |4\rangle$ transition with σ^- helicity (Figure 49.40b). Note that in the former case, level $|4\rangle$ does not participate in the relaxation dynamics and the system is effectively a three-level Λ-system, rather than a four-level system. The initial population is assumed to be equally distributed between the lower-lying levels. The ultrashort circularly polarized source pulse is injected into the medium, and the temporal dynamics of the electric-field vector components and population of all four levels are sampled at four different locations along the nanotube z-axis. The spatially resolved temporal dynamics at a point $z = 300$ nm from the left boundary of the simulated structure for σ^+ and σ^- excitations is shown in Figure 49.41.

It would be interesting from an experimental point of view to test whether nanotube chirality could be determined from the ultrafast nonlinear response using ultrashort pulses with either helicity. We compare the Fourier spectra at the output boundary for each helicity of the optical pulse excitation in Figure 49.42. Both transmission spectra exhibit a sharp peak at the resonant wavelength, indicating

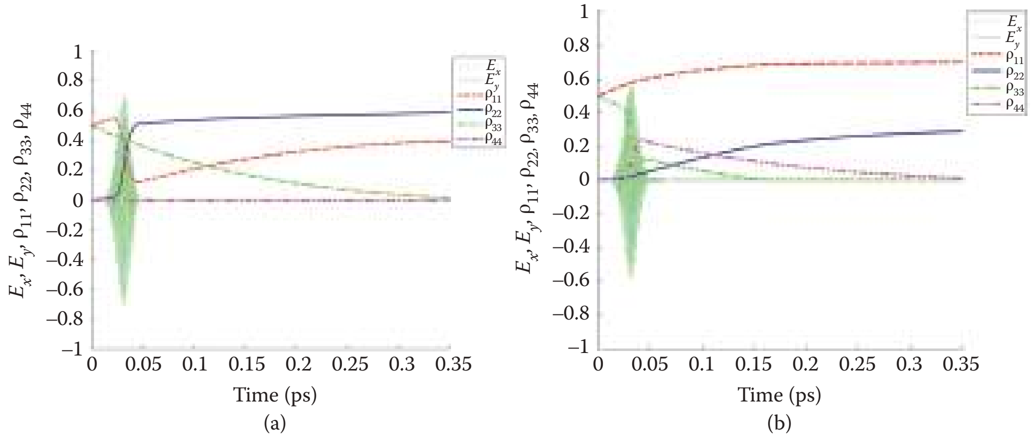


FIGURE 49.41 Time evolution of the electric-field components E_x and E_y, and the populations of all four energy levels of a SWCNT (a) for a circularly polarized σ^+ excitation pulse (note that level $|4\rangle$ is not involved in the dynamics) and (b) for a circularly polarized σ^- excitation pulse. The location is $z = 300$ nm. The left boundary of the SWCNT is at $z = 0$. The SWCNT is 500 nm long and is placed between two free-space regions, each 50 nm long.

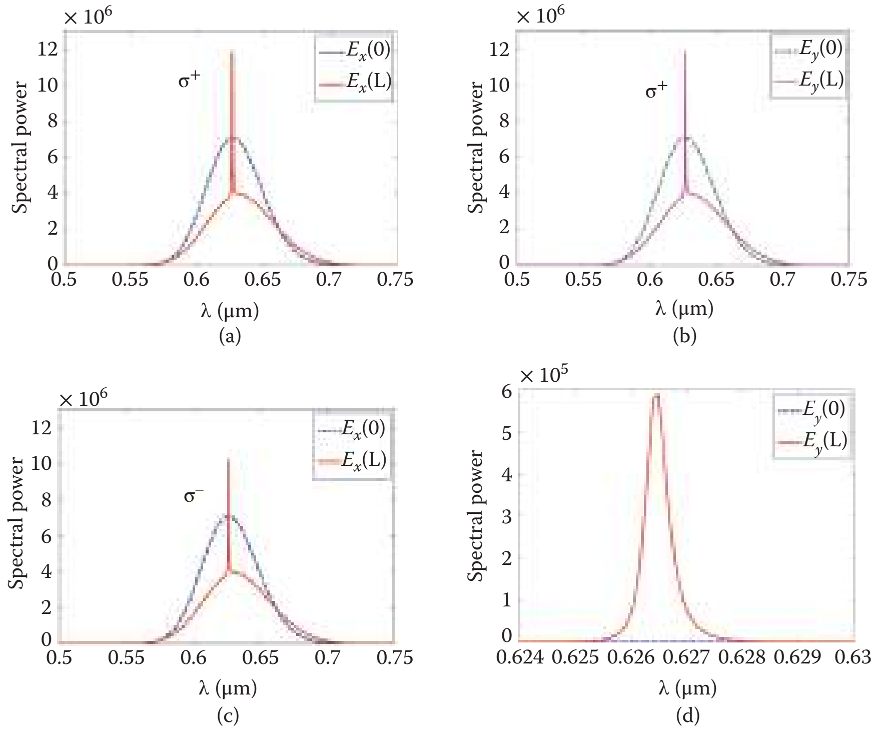


FIGURE 49.42 Transmission Fourier spectra plotted against the wavelength of the E_x and E_y electric-field components at the input, $E_{x,y}(0)$, and output, $E_{x,y}(L)$, of the structure for a (a), (b) σ^+ polarized pulse, and a (c), (d) σ^- polarized pulse. Note that the peak output amplitudes are approximately 20% higher in (a) and (b) compared to in (c) and (d). The simulated structure is as in Figure 49.41.

resonant amplification; however, the peak intensity corresponding to σ^+ excitation is nearly 20% higher. This difference can be exploited in an experiment aiming to determine unambiguously the chirality of a single CNT by first injecting a pulse of one helicity, then the other, and observing the consequences at the output. We show that the difference in the polarization-resolved transmission spectra of a linearly polarized pulse is much more pronounced and allows identification of the precise helical form of the SWCNT.

In order to demonstrate rotation of the electric-field polarization plane during the pulse propagation across the resonant four-level medium, we launch a linearly polarized pulse along x, expressed by:

$$X \begin{cases} E_x\,(z=0,t) = E_0 e^{-(t-t_0)^2 \big/ t_d^2} \cos(\omega_0 t) \\ E_y\,(z=0,t) = 0. \end{cases} \tag{49.60}$$

The system's temporal dynamics induced by the ultrashort linearly polarized pulse is shown in Figure 49.43a and b on an expanded scale, where the appearance of a second E_y component, and therefore also the appearance of optical rotation of the electric-field vector polarization, is clearly visible. The maximum amplitude of the E_y component continuously increases as the pulse propagates along the nanotube structure. We should therefore expect the maximum optical rotation angle to occur at the output boundary.

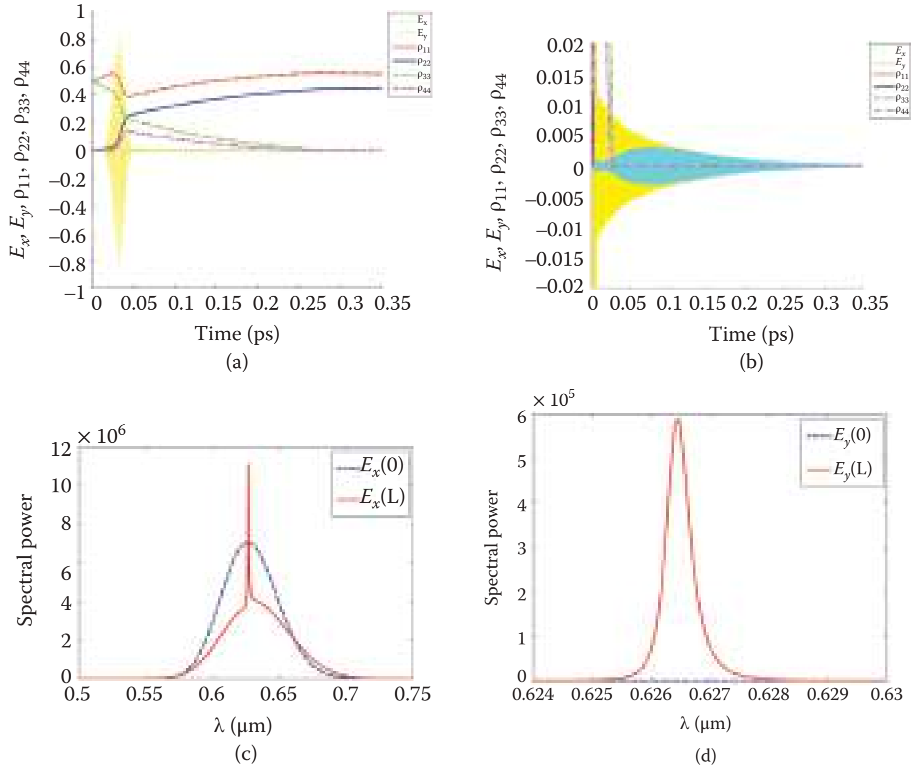
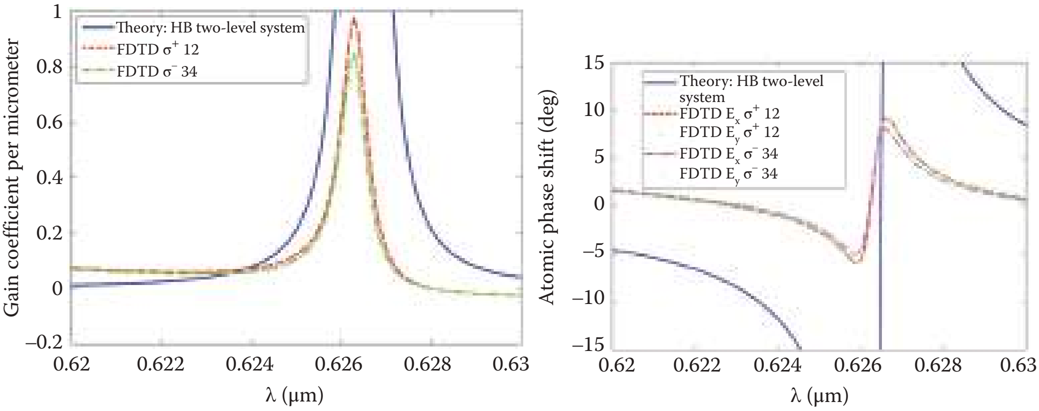


FIGURE 49.43 (a) Time evolution of a linearly polarized pulse with initial value $E_y = 0$, and the populations of all four levels at $z = 550$ nm from the left boundary. The simulated structure is as in Figure 49.41. (b) Magnified view of (a) showing the build-up of the E_y component with time, which is equivalent to a rotation of the electric-field polarization. The transmitted pulse is elliptically polarized and displays different power spectra for (c) the E_x component (sharp resonant peak superimposed on a broader line) and (d) the E_y component (single spectral line).

The polarization-resolved transmission spectra of the E_x and E_y components are quite distinct: while the former is similar to the spectrum obtained for circularly polarized optical excitation, the E_y spectrum exhibits a single spectral line. Conversely, for excitation with light of the opposite helicity (data not shown), the E_x and E_y profiles are interchanged. This difference in the polarization-resolved nonlinear optical response can therefore be used for unambiguous determination of the nanotube chirality.

In order to obtain a quantitative estimate of the natural optical activity in a single chiral CNT, we follow the general methodology developed in Section 49.3.3. We calculate the gain/absorption coefficient from the complex propagation factor, $e^{ik_c(z_2-z_1)}$, for the E_x and E_y electric-field components of a circularly polarized pulse over a distance of one dielectric wavelength, $l = z_2 - z_1 = \lambda_0/n$, where λ_0 is the resonant wavelength. The wavevector is $k_c = \beta + i\gamma$, where β is the phase shift per unit length induced in the optical pulse by the interaction with the resonant medium, and γ is the gain/absorption coefficient. The gain/absorption coefficient allows us to calculate the magnitude of the circular dichroism, whereas the phase shift represents a measure of the rotation angle. A comparison between the spatially resolved gain coefficients for σ^+ and σ^- pulse excitation is shown in Figure 49.44a; the analytically computed gain coefficient of a HB two-level system has also been plotted on the same graph, for reference. The maxima of the gain coefficients for both σ^+ and σ^- occur at the resonant wavelength. The circular dichroism, $\Delta A = G_R - G_L$, is therefore calculated as the average of the difference between the maximal gain coefficients for σ^+ and σ^- excitation (R and L subscripts, respectively) computed at several points along the nanotube. This yields a value for the circular dichroism of 0.083 μm^{-1}. By comparison, the absolute value of the circular dichroism of an artificial helicoidal bilayered structure varies in the range of 5–9 dB, which is equivalent to a linear amplitude gain/absorption coefficient of 0.58–1.04 μm^{-1}. The atomic phase shift ($\beta_{R(L),x(y)}l$) and the average rotation angle are plotted in Figure 49.44b. The specific rotatory power per unit length is calculated using $\rho = \pi(n_\text{L} - n_\text{R})/\lambda_0$, where n_L and n_R are the refractive indices for left and right circularly polarized light. Computing this value gives ≈2962.24°/mm.

Although the circular dichroism is low compared to the artificial chiral photonic structures, the optical rotation, as computed from the phase shift, is enormous compared to other birefringent materials. The



(a)

(b)

FIGURE 49.44 (a) Plot against wavelength of the spatially resolved gain coefficient per micron for a σ^+ (dashed curve) and σ^- (dash-dotted curve) circularly polarized ultrashort optical pulse, and the theoretical gain coefficient of a homogeneously broadened (HB) two-level system (solid curve) (Equation 49.30). The simulated structure is as in Figure 49.41. (b) Spectrum of the phase shift, averaged over the entire length of the structure, of the E_x (dashed curve) and E_y (dotted curve) electric-field components of σ^+ excitation, E_x (dash-dotted curve) and E_y (dotted curve) of σ^- optical excitation, and theoretical calculations of this phase shift (Equation 49.31), assuming a HB, resonant two-level system.

specific rotatory power of crystals varies widely from 2.24°/mm for $NaBrO_3$ to 522°/mm for $AgGaS_2$. Liquid substances exhibit much lower values of specific rotatory power—e.g., $\rho = -0.37°/\text{mm}$ for turpentine ($T = 10°$, $\lambda = 589.3$ nm); $\rho = 1.18°/\text{mm}$ for corn syrup, etc. Cholesteric liquid crystals and sculptured thin films exhibit large rotatory power in the visible spectrum: ~1000°/mm and ~6000°/mm, respectively, which is comparable with the calculated rotatory power. We should note that the dependence of the optical rotation angle on chirality opens up the exciting possibility to engineer this angle through manipulating the chirality of the structure.

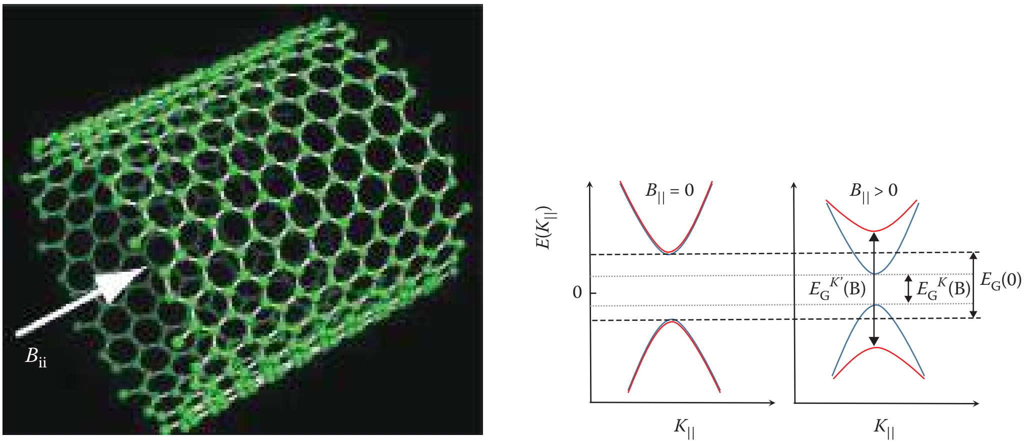
49.7.2 Faraday Rotation and Magneto-Chiral Effects

In this section, we shall develop a theoretical model of the resonant, coherent nonlinear optical activity when a static magnetic field, $B_{\|}$, threads the nanotube (Figure 49.45).

We shall be interested in the Faraday effect, or the rotation of the polarization of a plane-polarized electromagnetic wave propagating in a medium permeated by a static magnetic field, **B**, oriented along the direction of propagation. In the presence of an axial magnetic field, the electronic band structure of a single CNT—and the electronic states near the bandgap edge in particular—change significantly, owing to the combined action of two effects: the spin-B interaction, resulting in Zeeman splitting of the energy levels [93–95], and the appearance of the Aharonov–Bohm phase in the wavefunction [83,96,97]. The two symmetric subbands at the K (or K') point of the BZ are degenerate at $B_{\|} = 0$ (Figure 49.45b). An applied magnetic field along the nanotube axis lifts this degeneracy by shifting the energy levels. As a result, the bandgap of one of the subbands (K') becomes larger while the bandgap of the other subband (K) becomes smaller [94] (Figure 49.45c).

Without loss of generality, we shall consider the electronic states near the bandgap at the K point in the BZ. Therefore, the overall effect will be a bandgap reduction. We consider and calculate separately the contributions to the bandgap from the Zeeman splitting and the Aharonov–Bohm effect.

The effect of the Aharonov–Bohm flux on the bandgap is to induce oscillations in its energy. The energy bandgap thus varies between zero and a fixed value with a period of the flux quantum, $\Phi_0 = h/e$ [83,93],



(a) (b) (c)

FIGURE 49.45 (a) Single (20, 10) chiral nanotube threaded by an axial magnetic field (a so-called Faraday configuration). (b) Magnetic energy bands in the absence of a magnetic field, $B_{\|}$; the energy subbands at the point K (K') are degenerate. (c) When an axial magnetic field is applied, the degeneracy is lifted and the energy bandgap of the subband at K' becomes larger, while that of the other subband at K diminishes.

resulting in periodic oscillations of the magneto-optical absorption spectra given by

$$E_G(\Phi) = \begin{cases} 3E_G(0)\left|\frac{1}{3} - \frac{\Phi}{\Phi_0}\right|, & 0 \leq \Phi/\Phi_0 \leq 1/2 \\ 3E_G(0)\left|\frac{2}{3} - \frac{\Phi}{\Phi_0}\right|, & 1/2 \leq \Phi/\Phi_0 \leq 1, \end{cases} \tag{49.61}$$

where $E_G(0) = \hbar\omega_0$ is the energy bandgap at zero magnetic field and ω_0 is the resonant transition frequency.

For a magnetic field $B = 8\,\text{T}$ with a flux Φ threading a (5, 4) nanotube with a diameter 0.61145 nm, the ratio $\Phi/\Phi_0 = 0.00057$, and therefore the first of the equations (49.61) above holds. Due to the orbital Aharonov–Bohm effect, this leads to an energy-level shift, or bandgap renormalization of $E_{AB} = 3.37$ meV, corresponding to a resonant angular frequency, $\omega_{AB} = 5.12 \times 10^{12}$ rad/s. At a fixed value of the static magnetic field, the orbital Aharonov–Bohm effect leads to a uniform shift in the energy levels, so the resonant transition frequency ω_0 is replaced by $\omega_0 - \omega_{AB}$.

In an external magnetic field, the energy levels near the bandgap of Figure 49.40b split and the spin degeneracy is lifted. The resulting energy-level system can be split into two reduced systems of levels, each of which represents a mirror image of the other. The symmetry is broken only by the fact that the allowed optical transitions are different in each case (see Figure 49.46).

The Zeeman splitting, or the spin-B interaction energy, is given by

$$\mathrm{E}_z = \mu_B g_e \sigma B_{\|}, \tag{49.62}$$

where $\mu_B = \frac{e\hbar}{2m_e}$ is the Bohr magneton; the electron g-factor, g_e, is taken to be the same as that of pure graphite (≈ 2), $\sigma = \pm 1/2$ is the z-axis projection of the electron spin (spin-up/spin-down state), and m_e is the free electron mass.

We have performed simulations with magnetic fields in the range 2–100 T and found for the considered (5, 4) tube a nonmonotonous dependence of the optical rotation angle on the magnitude of **B**, with maximum rotation reached for magnetic fields in the range 8–10 T. We therefore selected $B_{\|} = 8\,\text{T}$ to demonstrate the maximum angle of Faraday rotation. Another reason for choosing this value is that it was the maximum magnetic field achievable by the static magnet available in the laboratory of our collaborators at the time when we performed these calculations. For $B_{\|} = 8\,\text{T}$, the Equation (49.62) gives an energy shift of $E_z \approx 0.46$ meV, which corresponds to a Zeeman resonant frequency, $\omega_z = 7.026 \times 10^{11}$ rad/s.

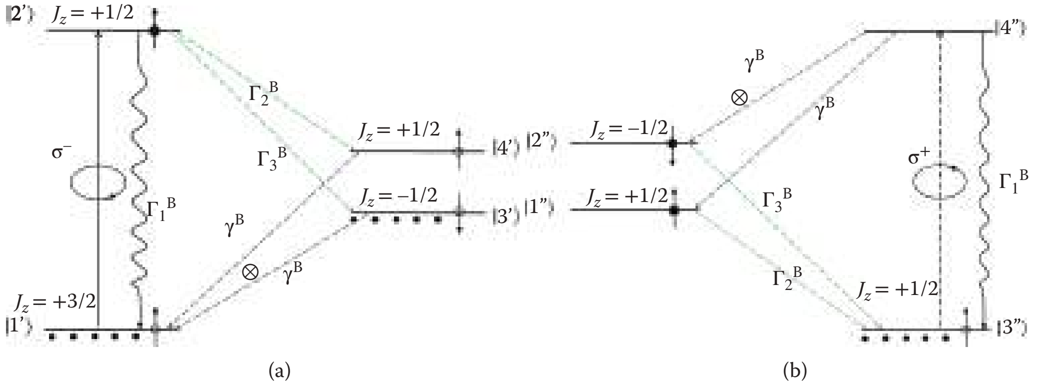


FIGURE 49.46 A four-level system at $B \neq 0$ splits into two nonsuperimposable reduced systems, involving levels energetically closest to the bandgap edge. (a) Left-handed reduced energy-level system excitable by a σ^- optical pulse. (b) Right-handed reduced energy-level system excitable by a σ^+ optical pulse. (Coherent and spontaneous transitions indicated as in Figure 49.40b. Forbidden dipole optical transitions are denoted by $\otimes$. In fact, there is a X-circle over the dashed lines in Figure 49.46, denoting forbidden transitions between levels).

The system Hamiltonians corresponding to the energy-level schemes displayed in Figure 49.46 are given by

$$\hat{H}' = \hbar \begin{pmatrix} \omega_z & -\frac{1}{2}\left(\Omega_x - i\Omega_y\right) & 0 & 0 \\ -\frac{1}{2}\left(\Omega_x + i\Omega_y\right) & \omega_0 - \omega_z & 0 & 0 \\ 0 & 0 & \Delta - \omega_z & 0 \\ 0 & 0 & 0 & \Delta + \omega_z \end{pmatrix} \tag{49.63}$$

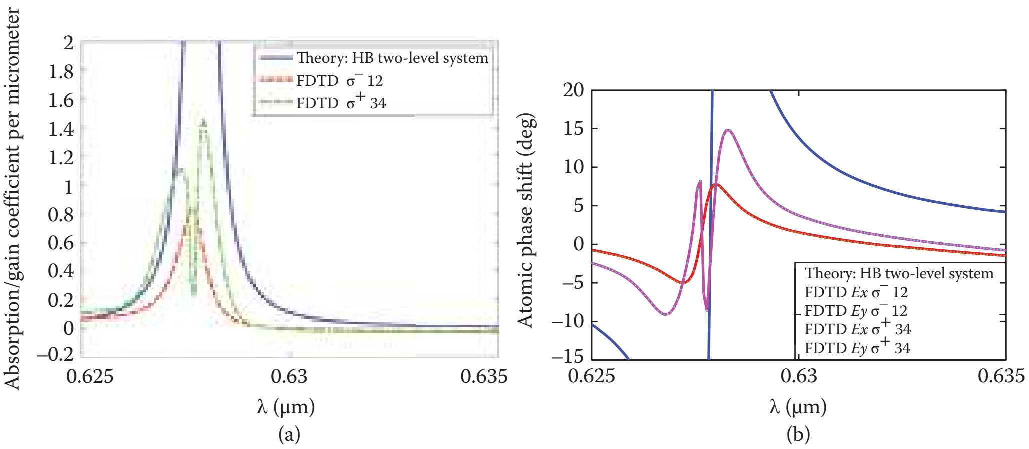
and

$$\hat{H}'' = \hbar \begin{pmatrix} \omega_0 - \omega_z & 0 & 0 & 0 \\ 0 & \omega_0 + \omega_z & 0 & 0 \\ 0 & 0 & \Delta + \omega_z & -\frac{1}{2}\left(\Omega_x + i\Omega_y\right) \\ 0 & 0 & -\frac{1}{2}\left(\Omega_x - i\Omega_y\right) & \Delta + \omega_0 - \omega_z \end{pmatrix}, \tag{49.64}$$

respectively.

We employ our vector Maxwell-pseudospin master equations model for the four-level systems above. The simulation structure is the same as the one described for the $B = 0$ case. Note that due to the presence of a magnetic field, the energy levels $|1\rangle$, $|2\rangle$, $|3\rangle$ and $|4\rangle$ are split into two sets, which we choose to distinguish by adding a $'$ and $''$ to the notation; the first set (denoted by $'$) is accessed only using σ^- light, whereas the second (denoted by $''$) is accessible only via σ^+ light, due to the optical selection rules. We then resonantly excite the $|1'\rangle \rightarrow |2'\rangle$ and $|3''\rangle \rightarrow |4''\rangle$ transitions using, respectively, a left (σ^-) and right (σ^+) circularly polarized π-pulse of duration 60 fs. The pulse's central frequency is tuned in resonance with the transition frequency, $\omega_0 - \omega_{AB} - 2\omega_z$, and the pulse envelope is a Gaussian function. Owing to the combined effect of the Aharonov-Bohm and Zeeman energy-level shift, the dipole matrix element is modified. An estimate of the optical dipole matrix element in an axial magnetic field can be obtained from the theory developed in [88], taking into account the bandgap reduction at $B = 8\,\mathrm{T}$, giving $\wp = 3.6205 \times 10^{-29}\,\mathrm{C} \cdot \mathrm{m}$. We recalculate the relaxation times using the above mentioned magnetic-field dipole coupling, thus obtaining the following relaxation rates: $\Gamma_1 = 2.914\,\mathrm{ns}^{-1}$, $\Gamma_2 = 9.79\,\mathrm{ns}^{-1}$ and $\Gamma_3 = 9.77\,\mathrm{ns}^{-1}$. The intraband relaxation rate, $\gamma = 130\,\mathrm{fs}^{-1}$, and the dephasing rates, $\Gamma_\mu = 800\,\mathrm{fs}^{-1}$ and $\Gamma_{\mu-1} = 1.6\,\mathrm{ps}^{-1}$, are taken to be the same as for the zero-field case. Note that the initial charge-carrier populations are different for excitation by the σ^- and σ^+ pulses: while in the former case the initial population is assumed to be equally distributed between the valence-band levels $\rho_{1'1'i} = \rho_{3'3'i} = 1/2$ nearest the bandgap edge, in the latter case the whole population is in the single valence-band state $\rho_{4'4'i} = 1$.

Similar to the zero-field case, the time evolution of the electric-field components and populations of all four levels are sampled at several points along the nanotube axis, and the magnetic circular dichroism spectra of the gain/absorption coefficients and phase shifts are calculated and displayed in Figure 49.47. Note that the resonance is shifted toward longer wavelengths due to the bandgap reduction. We should point out that the magnetic circular dichroism spectra at $B = 8$ T are quite distinct from the zero magnetic field ones for natural optical activity (Figure 49.44). While the σ^- polarized pulse is amplified during its propagation along the nanotube, the σ^+ polarized pulse is absorbed, resulting in a much larger net circular dichroism. The average value of the magnetic circular dichroism is $0.706\,\mu\mathrm{m}^{-1}$, an order of magnitude larger than the natural circular dichroism. The different behavior of the calculated gain and absorption spectra under σ^- compared to σ^+ polarized optical pulse excitation is a direct consequence of the different energy-level configuration describing the two cases. While the energy-level system for the σ^- excitation is a four-level system, the one corresponding to the σ^+ excitation is a three-level Λ-system, due to level $|2''\rangle$ being completely decoupled from the rest of the levels as a result of the dipole optical selection rules. The calculated spectra in the latter case are, in fact, a signature of EIT and coherent population trapping



(a)

(b)

FIGURE 49.47 (a) Spectrum of the gain coefficient per micron, averaged spatially over the entire length of the structure, of the E_x and E_y electric-field vector components of σ^- (dashed curve) and σ^+ (dash-dotted curve) circularly polarized ultrashort optical excitation. The theoretical gain coefficient of a homogeneously broadened (HB), resonant two-level system (solid curve) at $B = 8\,\text{T}$ is also shown. Note that, while the σ^- pulse is amplified during its propagation along the nanotube, the σ^+ pulse is absorbed, resulting in an absorption dip in the spectrum. (b) Spectrum of the phase shift, averaged spatially over the entire length of the structure, at $B = 8\,\text{T}$ of the E_x (dashed curve) and E_y (dots) electric-field components under σ^- excitation, and of E_x (dash-dotted curve) and E_y (dots) under σ^+ circularly polarized ultrashort optical excitation. The phase spectra exhibit double-peaked refraction characteristic of EIT. The theoretical phase shift of a HB resonant two-level system (solid curve), is also given. (FDTD, finite-difference time-domain.)

effects in a three-level system [98], as will be explained subsequently. The absorption at resonance is close to zero, and the spectrum exhibits a dip (Figure 49.47a) similar to the absorption dip characteristic of EIT. The dip is caused by destructive interference between the excitation and emission fields, and leads to the trapping of the charge-carrier population in the ground state (hence, the term "coherent population trapping"). Whereas the phase-shift spectrum for a σ^- pulse excitation is of the type of a two-level atomic phase shift, the spectrum of the phase shift for σ^+ excitation is double peaked (Figure 49.47b). The latter is characteristic of the real part of the dielectric susceptibility in a three-level system exhibiting EIT. The predicted destructive interference in an external axial magnetic field after the passage of the ultrashort pulse is a direct consequence of the timescales of the processes involved in the relaxation dynamics and the phase shift which they introduce in the emission. The calculated average specific rotatory power in a magnetic field $B = 8\,\text{T}$ is $-32580.4°/\text{mm}$, corresponding to an average refractive-index anisotropy $(n_R - n_L)$ of 6.497 (the purpose of the minus sign in the rotatory power is to denote counterclockwise rotation of the field when viewed from a direction opposite to that of its propagation). This is nearly an order of magnitude greater than the natural optical rotation. We should note that the calculated rotation is a combined effect of the chirality of the nanotube and the magnetic field–induced rotation, and thus can be considered as an estimate for the magneto-chiral effect in a single nanotube. Therefore, our model can be used to study the magnetic field and chirality dependence of the optical rotation angle.

49.8 Conclusion

We have developed a general methodology for description of ultrashort optical-pulse interaction with open, discrete multilevel quantum systems, based on a self-consistent solution of the vector Maxwell's equations and equations for the time evolution of the quantum system in the real-vector representation of the density-matrix geometrical picture. An advantage of this theoretical method is that the coupled vector Maxwell-pseudospin equations can be generated on the fly for a quantum system with an arbitrary number of energy levels using the corresponding unique SU(N) Lie algebra generators. Unlike the widely accepted approach of using simplified two-level systems for description of the system dynamics, our approach allows to model multiple optical transitions and time-dependent gain effects that occur in a realistic quantum system. Our theoretical framework goes beyond the usual rate equations approach, which is unable to model and predict coherent quantum effects that could arise due to couplings and interplay between multiple optical transitions in the system. Most other approaches use the SVEA and RWA which consider only events occurring on a relatively slow timescale—such as the pulse envelope—but do not take account of the pulse's fast carrier wave. By contrast, the proposed approach allows to model rigorously ultrashort optical pulses down to a few optical cycles. Contrary to BPM, no approximations are made to the electromagnetic field equations and their solutions.

We have shown that the model can describe coherent propagation effects: self-induced and EIT, and self-localization effects, such as soliton and pattern formation in cavities. The method can be applied to a wide range of quantum-system geometries by imposing macroscopic boundary conditions and, hence, it can be used to simulate the performance of not only individual idealized devices, but whole interacting systems. We have shown that the model can successfully describe the complex dynamics of an open quantum system, and at the same time provides the capability to observe the feedback mechanisms—including non-Markovian memory effects—as they occur in space and time. In addition, within the Langevin model extension, we were able to model quantum-optical effects such as spontaneous emission. This approach can be used to simulate quantum correlations, nonclassical light generation and manipulation, and few-photon optical nonlinearities.

The model can describe interactions of matter with circularly/elliptically polarized optical pulses and was successfully applied to the problem of generation, manipulation, and readout of single spins and their dynamics. This success is critical for the development of quantum information technologies—currently a very hot research topic which could potentially revolutionize technological applications. Possible application areas include—but are not limited to—quantum computing and information technologies, such as spin-qubit manipulation in QDs, cold atoms, and NV centers in nanodiamond. CNT and arrays of aligned nanotubes have the potential to serve as functional elements in polarization-sensitive devices and polarization switches, as well as for metamaterial applications in the visible part of the electromagnetic spectrum. Predicted spatio-temporal coherent effects, such as cavity solitons, can be used as integrated low-loss optical switches and for all-optical information processing. Finally, the proposed methodology can be used as a design tool for future quantum-optical devices and for the advancement of quantum-photonic technologies.

References

1. R. J. Glauber. The quantum theory of optical coherence. *Physical Review*, 130:2529, 1963.
2. R. J. Glauber. *Quantum Optics and Electronics*. New York, NY: Gordon Breach, 1965.
3. R. Hanbury Brown and R. Q. Twiss. *Nature*, 177:27, 1956.
4. L. Mandel and S. Wolf. *Optical Coherence and Quantum Optics*. Cambridge University Press, 1995.
5. B. Lounis and M. Orrit. Single-photon sources. *Report on Progress in Physics*, 68:1129, 2005.
6. A. Einstein, B. Podolsky, and N. Rosen. Can quantum mechanical description of physical reality be considered complete? *Physical Review*, 47:777, 1935.
7. J. L. O'Brien, A. Furusawa, and J. Vuckovic. Photonic quantum technologies. *Nature Photonics*, 3:687, 2009.
8. T. M. Babinec, B. J. M. Hausmann, M. Khan, Y. Zhang, J. R. Maze, Ph. R. Hemmer, and M. LonÄar. A diamond nanowire single-photon source. *Nature Nanotechnology*, 5:195, 2010.
9. J. Claudon, J. Bleuse, N. S. Malik, M. Bazin, P. Jaffrennou, N. Gregersen, C. Sauvan, P. Lalanne, and J.-M. Gerard. A highly efficient single-photon source based on a quantum dot in a photonic nanowire. *Nature Photonics*, 4:174, 2010.

10. D. Dalacu, K. Mnaymneh, V. Sazonova, P. J. Poole, G. C. Aers, J. Lapointe, R. Cheriton, A. J. SpringThorpe, and R. Williams. Deterministic emitter-cavity coupling using a single-site controlled quantum dot. *Physical Review B*, 82:033301, 2010.
11. K. G. Lee, X. W. Chen, H. Eghlidi, P. Kukura, R. Lettow, A. Renn, V. Sandoghdar, and S. Götzinger. A planar dielectric antenna for directional single-photon emission and near-unity collection efficiency. *Nature Photonics*, 5:166, 2011.
12. S. Strauf, N. G. Stoltz, M. T. Rakher, L. A. Coldren, P. M. Petroff, and D. Bouwmeester. High-frequency single-photon source with polarization control. *Nature Photonics*, 1:704, 2007.
13. D. Loss and D. P. diVincenzo. Quantum computation with quantum dots. *Physical Review A*, 57:120, 1998.
14. B. Meier and B. P. Zakharchenya editors. *Optical Orientation*. North-Holland, Amsterdam: Elsevier, 1984.
15. W. S. Warren, H. Rabitz, and M. Dahleh. Coherent control of quantum dynamics: The dream is alive. *Science*, 259:1581, 1993.
16. K. De Greve, D. Press, P. L. McMahon, and Y. Yamamoto. Ultrafast optical control of individual quantum dot spin qubits. *Reports on Progress in Physics*, 76:092501, 2013.
17. M. Sargent III, M. O Scully, and W. E. Lamb Jr. *Laser Physics*. Reading, MA: Addison-Wesley Publishing Company Advanced Book Program, 1974.
18. S. L. McCall and E. L. Hahn. Self-induced transparency by pulsed coherent light. *Physical Review Letters*, 18:908, 1967.
19. S. Schneider, P. Borri, W. Langbein, U. Woggon, J. Förstner, A. Knorr, R. L. Sellin, D. Ouyang, and D. Bimberg. Self-induced transparency in InGaAs quantum-dot waveguides. *Applied Physics Letters*, 83:3668, 2003.
20. S. Hughes. Breakdown of the area theorem: Carrier-wave Rabi flopping of femtosecond optical pulses. *Physical Review Letters*, 81:3363, 1998.
21. D. Pinotsi and A. Imamoglu. Single photon absorption by a single quantum emitter. *Physical Review Letters*, 100:093603, 2008.
22. M. Sondermann, R. Maiwald, H. Konermann, N. Lindlein, U. Peschel, and G. Leuchs. Design of a mode converter for efficient light-atom coupling in free space. *Applied Physics B*, 89:489, 2007.
23. S. J. van Enk. Atoms, dipole waves, and strongly focused light beams. *Physical Review A*, 69:043813, 2004.
24. M. Brune, F. Schmidt-Kaler, A. Maali, J. Dreyer, E. Hagley, J. M. Raimond, and S. Haroche. Quantum Rabi oscillation: A direct test of field quantization in a cavity. *Physical Review Letters*, 76:1800, 1996.
25. K. M. Birnbaum, A. Boca, R. Miller, A. D. Boozer, T. E. Nortup, and H. J. Kimble. Photon blockade in an optical cavity with one trapped atom. *Nature*, 436, 87:87, 2005.
26. R. Young, R. M. Stevenson, P. Atkinson, K. Cooper, D. A. Ritchie, A. J. Shields. *Advances in Solid State Physics*, Vol. 46:55. Berlin: Springer Verlag, 2008.
27. O. Gazzano, S. Michaelis de Vasconcellos, C. Arnold, A. Nowak, E. Galopin, I. Sagnes, L. Lanco, A. Lemaître, and P. Senellart. Bright solid-state sources of indistinguishable single photons. *Nature communications*, 4:1425, 2012.
28. K. Srinivasan and O. Painter. Linear and nonlinear optical spectroscopy of a strongly coupled micro disk-quantum dot system. *Nature*, 450:862, 2007.
29. N. Le Thomas, U. Woggon, O. Schöps, M. V. Artemyev, M. Kazes, and U. Banin. Cavity QED with semiconductor nanocrystals. *Nano Letters*, 6:557, 2006.
30. M. Heiss, Y. Fontana, A. Gustafsson, G. Wust, C. Magen, D. D. O'Regan, J. W. Luo, B. Ketterer, S. Conesa-Boj, A. V. Kuhlmann, J. Houel, E. Russo-Averchi, J. R. Morante, M. Cantoni, N. Mazari, J. Arbiol, A. Zunger, R. J. Warburton, and A. Fontcuberta i Morral. Self-assembled quantum dots in a nanowire system for quantum photonics. *Nature Materials*, 12:439, 2013.

31. H. Kim, R. Bose, T. C. Shen, G. S. Solomon, and E. Waks. A quantum logic gate between a solid-state quantum bit and a photon. *Nature Photonics*, 7:373, 2013.
32. A. Faraon, I. Fushman, D. Englund, N. Stoltz, P. Petroff, and J. Vučković. Coherent generation of non-classical light on a chip via photon-induced tunnelling and blockade. *Nature Physics*, 4:859, 2008.
33. J. Kasprzak, S. Reitzenstein, E. A. Muljarov, C. Kistner, C. Schneider, M. Strauss, S. Hing, A. Forchel, and W. Langbein. Up on the Jaynes-Cummings ladder of a quantum-dot/microcavity system. *Nature Materials*, 9:304, 2010.
34. P. Michler, A. Kiraz, C. Becher, W. V. Schoenfeld, P. M. Petroff, L. Zhang, E. Hu, and A. Imamoglu. A quantum dot single-photon turnstile device. *Science*, 290:2282, 2000.
35. C. Santori, D. Fattal, J. Vuckovic, G. S. Solomon, and Y. Yamamoto. Indistinguishable photons from a single-photon device. *Nature*, 419:594, 2002.
36. E. Knill, R. Laamme, and G. J. Milburn. A scheme for efficient quantum computation with linear optics. *Nature*, 409:46, 2001.
37. L. Allen and J. H. Eberly. *Optical Resonance and Two-Level Atoms*. New York, NY: Wiley, 1975.
38. R. P. Feynman, F. L. Vernon, and R. W. Hellwarth. Geometrical representation of the Schrödinger equation for solving laser problems. *Journal of Applied Physics*, 28:49, 1957.
39. R. J. Cook and B. W. Shore. Coherent dynamics of N-level atoms and molecules. III. An analytically soluble periodic case. *Physical Review A*, 20:539, 1979.
40. J. N. Elgin. Semiclassical formalism for the treatment of three-level systems. *Physics Letters*, 80A:140, 1980.
41. M. Gell-Mann and Y. Neeman. *The Eightfold Way*. New York, NY: Benjamin, 1964.
42. F. T. Hioe and J. H. Eberly. N-level coherence vectors and higher conservation laws in quantum optics and quantum mechanics. *Physical Review Letters*, 47:838, 1981.
43. S. L. McCall and E. L. Hahn. Self-induced transparency. *Physical Review*, 183:457, 1969.
44. G. Slavcheva. Model for the coherent optical manipulation of a single spin state in a charged quantum dot. *Physical Review B*, 77:115347, 2008.
45. G. Slavcheva, J. M. Arnold, I. Wallace, and R. W. Ziolkowski. Coupled Maxwell-pseudospin equations for investigation of self-induced transparency effects in a degenerate three-level quantum system in two dimensions: Finite-difference time-domain study. *Physical Review A*, 66:063418, 2002.
46. A. Taove. *Computational Electrodynamics: The Finite-Difference Time-Domain Method*. Norwood, MA: Artech House, 2000.
47. R. W. Ziolkowski, J. M. Arnold, and D. M. Gogny. Ultrafast pulse interactions with two-level atoms. *Physical Review*, 52:3082, 1995.
48. B. Engquist and A. Majda. Absorbing boundary conditions for the numerical simulation of waves. *Mathematics of Computation*, 31:629, 1977.
49. G. Mur. Absorbing boundary conditions for the finite-difference approximation of the time-domain electromagnetic-field equations. *IEEE Transactions on Electromagnetic Compatibility*, 23:377, 1981.
50. A. Yariv. *Quantum Electronics*. New York, NY: Wiley, 1989.
51. G. M. Slavcheva, J. M. Arnold, and R.W. Ziolkowski. FDTD simulation of the nonlinear gain dynamics in active optical waveguides and semiconductor microcavities. *IEEE Journal of Selected Topics in Quantum Electronics*, 10(5):1052, 2004.
52. A. E. Siegman. *Lasers*. Mill Valley, CA: University of Science Books, 1986.
53. G. Slavcheva, J. M. Arnold, and R. W. Ziolkowski. Ultrashort pulse lossless propagation through a degenerate three-level medium in nonlinear optical waveguides and semiconductor microcavities. *IEEE Journal of Selected Topics in Quantum Electronics*, 9(3):929, 2003.
54. M. Blaaboer, G. Kurizki, and B. A. Malomed. Spatiotemporally localized solitons in resonantly absorbing Bragg reflectors. *Physical Review E*, 62:R57, 2000.

55. A. Kozhekin and G. Kurizki. Self-induced transparency in Bragg reflectors: Gap solitons near absorption resonances. *Physical Review Letters*, 74:5020, 1995.
56. A. Kozhekin, G. Kurizki, and B. A. Malomed. Standing and moving gap solitons in resonantly absorbing Bragg gratings. *Physical Review Letters*, 81:3647, 1998.
57. G. Slavcheva and O. Hess. Dynamical model of coherent circularly polarized optical pulse interactions with two-level quantum systems. *Physical Review A*, 72:053804, 2005.
58. A. I. Maimistov, A. M. Besharov, O. Elyutin, and Yu. M. Sklyarov. Present state of self-induced transparency theory. *Physics Reports*, 191:1–108, 1990.
59. G. Slavcheva and O. Hess. Spin-dependent dynamics of ultrafast polarised optical pulse propagation in coherent semiconductor quantum systems. *Physica Status Solidi C*, 3:2414, 2006.
60. G. Slavcheva and O. Hess. All-optical coherent control of spin dynamics in semiconductor quantum dots. *Optical and Quantum Electronics*, 31(12–14):973, 2007.
61. A. Wojs, P. Hawrylak, S. Fafard, and L. Jacak. Electronic structure and magneto-optics of self-assembled quantum dots. *Physical Review B*, 54:5604, 1996.
62. A. Greilich, R. Oulton, E. A. Zhukov, I. A. Yugova, D. R. Yakovlev, M. Bayer, A. Shabaev, A. L. Efros, I. A. Merkulov, V. Stavarache, D. Reuter, and A. Wieck. Optical control of spin coherence in singly charged (In,Ga)As/GaAs quantum dots. *Physical Review Letters*, 96:227401, 2006.
63. P. Chen, C. Piermarocchi, L. J. Sham, D. Gammon, and D. G. Steel. Theory of quantum optical control of a single spin in a quantum dot. *Physical Review B*, 69:075320, 2004.
64. A. Shabaev, A. L. Efros, D. Gammon, and I. A. Merkulov. Optical readout and initialization of an electron spin in a single quantum dot. *Physical Review B*, 68:201305(R), 2003.
65. A. V. Khaetskii, D. Loss, and L. Glazman. Electron spin decoherence in quantum dots due to interaction with nuclei. *Physical Review Letters*, 88:186802, 2002.
66. I. A. Merkulov, A. L. Efros, and M. Rosen. Electron spin relaxation by nuclei in semiconductor quantum dots. *Physical Review B*, 65:205309, 2002.
67. D. V. Bulaev and V. Loss. Spin relaxation and decoherence of holes in quantum dots. *Physical Review Letters*, 95:076805, 2005.
68. T. Flissikowski, I. A. Akimov, A. Hundt, and F. Henneberger. Single-hole spin relaxation in a quantum dot. *Physical Review B*, 68:161309(R), 2003.
69. S. Economou, R.-B. Liu, L. J. Sham, and D. G. Steel. Unified theory of consequences of spontaneous emission in a λsystem. *Physical Review B*, 71:195327, 2005.
70. M. Taylor, E. Harbord, P. Spencer, E. Clarke, G. Slavcheva, and R. Murray. Optical spin-filtering effect in charged InAs/GaAs quantum dots. *Applied Physics Letters*, 97:171907, 2010.
71. K. Kavokin. Fine structure of the quantum-dot trion. *Physica Status Solidi: A*, 195:592, 2003.
72. P. F. Braun, X. Marie, L. Lombez, B. Urbaszek, T. Amand, P. Renucci, V. K. Kalevich, K. V. Kavokin, O. Krebs, P. Voisin, and Y. Masumoto. Direct observation of the electron spin relaxation induced by nuclei in quantum dots. *Physical Review Letters*, 94:116601, 2005.
73. B. Eble, C. Testelin, P. Desfonds, F. Bernardot, A. Balocchi, T. Amand, A. Miard, A. Lemaître, X. Marie, and M. Chamarro. Hole-nuclear spin interaction in quantum dots. *Physical Review Letters*, 102:146601, 2009.
74. G. Slavcheva. Model of coherent optical spin manipulation through hot trion states in p-doped InAs/GaAs quantum dots. *ArXiv:1301.7018v1*, 2013.
75. T. Warming, E. Siebert, A. Schliwa, E. Stock, R. Zimmermann, and D. D. Bimberg. Hole-hole and electron-hole exchange interactions in single InAs/GaAs quantum dots. *Physical Review B*, 79:125316, 2009.
76. G. A. Narvaez, G. Bester, and A. Zunger. Carrier relaxation mechanisms in self-assembled (In, Ga) As/GaAs quantum dots: Efficient P→S auger relaxation of electrons. *Physical Review B*, 74:075403, 2006.
77. C. W. Gardiner and P. Zoller. *Quantum Noise*, 3rd ed. Berlin: Springer, 2004.

78. V. Weisskopf and E. Wigner. Berechnung der natürlichen linienbreite auf grund der diracschen lichttheorie. *Zeitschrift für Physik*, 63:54, 1930.
79. P. A. M. Dirac. The quantum theory of the emission and absorption of radiation. *Proceedings of the Royal Society London A*, 114:243, 1927.
80. G. Slavcheva and Ph. Roussignol. Nonlinear coherent magneto-optical response of a single chiral carbon nanotube. *New Journal of Physics*, 12:103004, 2010.
81. L. Yang, M. P. Anantram, J. Han, and J. P. Lu. Band-gap change of carbon nanotubes: Effect of small uniaxial and torsional strain. *Physical Review B*, 60:13874, 1999.
82. G. L. A. Rikken and E. Raupach. Observation of magneto-chiral dichroism. *Nature*, 390:493, 390.
83. A. Ajiki and T. Ando. Electronic states of carbon nanotubes. *Journal of the Physical Society of Japan*, 62:1255, 1993.
84. Y. Zhang, A. Chang, J. Cao, Q. Wang, W. Kim, Y. Li, N. Morris, E. Yenilmez, J. Kong, and H. Daia. Electric-field-directed growth of aligned single-walled carbon nanotubes. *Applied Physics Letters*, 79:3155, 2001.
85. L. Jin, C. Bower, and O. Zhou. Alignment of carbon nanotubes in a polymer matrix by mechanical stretching. *Applied Physics Letters*, 73:1197, 1998.
86. Ge. G. Samsonidze, A. Grüneis, R. Saito, A. Jorio, A. G. Souza Filho, G. Dresselhaus, and M. S. Dresselhaus. Interband optical transitions in left- and right-handed single-wall carbon nanotubes. *Physical Review B*, 69:205402, 2004.
87. J.-S. Lauret, C. Voisin, G. Cassabois, C. Delalande, Ph. Roussignol, O. Jost, and L. Capes. Ultrafast carrier dynamics in single-wall carbon nanotubes. *Physical Review Letters*, 90:057404, 2003.
88. E. L. Ivchenko and B. Spivak. Chirality effects in carbon nanotubes. *Physical Review B*, 66:155404, 2002.
89. F. Wang, G. Dukovic, L. E. Brus, and T. F. Heinz. Time-resolved fluorescence of carbon nanotubes and its implication for radiative lifetimes. *Physical Review Letters*, 92:177401, 2000.
90. H. Bao, X. Ruan, and T. S. Fisher. Optical properties of ordered vertical arrays of multi-walled carbon nanotubes from FDTD simulations. *Optics Express*, 18:6347, 2010.
91. V. N. Popov and L. Henrard. Comparative study of the optical properties of single-walled carbon nanotubes within orthogonal and nonorthogonal tight-binding models. *Physical Review B*, 70:115407, 2004.
92. V. N. Popov. Curvature effects on the structural, electronic and optical properties of isolated single-walled carbon nanotubes within a symmetry-adapted non-orthogonal tight-binding model. *New Journal of Physics*, 6:17, 2004.
93. J. Jiang, J. Ding, and D. Y. Xing. Zeeman effect on the electronic spectral properties of carbon nanotubes in an axial magnetic field. *Physical Review B*, 62:13209, 2000.
94. E. D. Minot, Y. Yaish, V. Sazonova, and P. L. McEuen. Determination of electron orbital magnetic moments in carbon nanotubes. *Nature*, 428:536, 2004.
95. F. L. Shuy, C. P. Chang, R. B. Chen, C. W. Chiu, and M. F. Lin. Magnetoelectronic and optical properties of carbon nanotubes. *Physical Review B*, 67:045405, 2003.
96. H. Ajiki. Magnetic-field effects on the optical spectra of a carbon nanotube. *Physical Review B*, 65:233409, 2002.
97. R. Saito, G. Dresselhaus, and M. S. Dresselhaus. Magnetic energy bands of carbon nanotubes. *Physical Review B*, 50:14698, 1994.
98. K. J. Boller, A. Imamoglu, and S. E. Harris. Observation of electromagnetically induced transparency. *Physical Review Letters*, 66:2593, 1991.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

X

Mathematical Methods

50 Drift-Diffusion Models *Patricio Farrell, Nella Rotundo, Duy Hai Doan, Markus Kantner, Jürgen Fuhrmann, and Thomas Koprucki* 733
Introduction • How to Get Started: Numerical Solution of the Stationary 1D Semiconductor Equations • Van Roosbroeck System • Discretizing the van Roosbroeck System • Nonlinear Solvers • Mesh Generation for the Finite Volume Method • Time Discretization • Contact Terminal Currents • An Alternative: The Finite Element Method • Extensions and Outlook

51 Monte Carlo Device Simulations *Katerina Raleva, Abdul R. Shaik, Raghuraj Hathwar, Akash Laturia, Suleman S. Qazi, Robin Daugherty, Dragica Vasileska, and Stephen M. Goodnick* ... 773
Introduction • Bulk Monte Carlo Method Description • Particle-Based Device Simulation • Representative Simulation Results • Outlook

52 Photonics *Frank Schmidt* 807
Landscape of Numerical Methods • Maxwell's Equations in Nondispersive, Chargeless Media • FDTD Method • Fourier Modal Method • Finite Element Method • DG Method

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

50

Drift-Diffusion Models

50.1 Introduction 733
50.2 How to Get Started: Numerical Solution of the Stationary 1D Semiconductor Equations 734
1D van Roosbroeck Model • Thermodynamic Equilibrium and Local Charge Neutrality • Discretizing the Poisson Problem • Solving the Discrete Nonlinear Poisson Problem • Discretizing the Continuity Equations Using the Scharfetter–Gummel Scheme • Solving the Full Discretized van Roosbroeck System • Why Not Use Central Finite Differences for the Fluxes?
50.3 Van Roosbroeck System 745
The van Roosbroeck System of Equations • Initial and Boundary Conditions • Recombination Processes • Thermodynamic Equilibrium • Non-Boltzmann Statistics and Generalized Einstein Relation • Free Energy and Dissipation Rate • Existence and Uniqueness Results • Maximum Principle
50.4 Discretizing the van Roosbroeck System 753
Finite Volume Method in Higher Dimensions • Scharfetter–Gummel Fluxes and Their Non-Boltzmann Generalizations • Flux Expressions Consistent with the Thermodynamic Equilibrium • Free Energy and Dissipation Rate • Existence, Uniqueness, and Convergence • Maximum Principle
50.5 Nonlinear Solvers 759
Newton's Method • Gummel's Method
50.6 Mesh Generation for the Finite Volume Method 761
Boundary Conforming Delaunay Meshes on Polyhedral Domains • Tensor Product Approaches • Open Problems
50.7 Time Discretization 763
50.8 Contact Terminal Currents 763
50.9 An Alternative: The Finite Element Method 765
50.10 Extensions and Outlook 766
Additional Physical Models • Coupling the van Roosbroeck System to Other Models • Methods for Doping Optimization • Alternative Modeling Approaches for Carrier Transport

Patricio Farrell
Nella Rotundo
Duy Hai Doan
Markus Kantner
Jürgen Fuhrmann
and
Thomas Koprucki

50.1 Introduction

The semiconductor technology is undeniably one of the most important branches in modern industry. Apart from its obvious significance to our daily lives, this technology is an excellent example of a broad collaboration among various disciplines. The development of novel technologies and devices has not only been driven by engineers, but also by physicists, mathematicians, and numerical analysts. On the one hand, rigorous mathematical models allow sophisticated predictions, which might be difficult to observe experimentally. On the other hand, numerical simulations have the potential to optimize device designs without the costly and time-consuming development of prototypes.

In 1950, van Roosbroeck introduced the fundamental semiconductor device equations [1] as a system of three nonlinearly coupled partial differential equations (PDEs). They describe the semiclassical transport of free electrons and holes in a self-consistent electric field using a drift-diffusion approximation. Since then, the so-called *van Roosbroeck system* (frequently also called *drift-diffusion system*) became the standard model to describe the current flow in semiconductor devices at a macroscopic scale. Typical devices modeled by these equations range from diodes and transistors to LEDs, solar cells, and lasers [2,3]. In recent years, the emergence of quantum nanostructures [4–7] and organic semiconductors [8–10] has invigorated the research activities in this area.

In this chapter, we focus on solving the van Roosbroeck system numerically. From a mathematical point of view, the challenge lies in the strong nonlinearities, the drift dominated nature of the underlying physics, the formation of internal and boundary layers as well as the need to accurately mirror qualitative physical properties, such as nonnegativity of carrier densities and consistency with thermodynamical principles. The finite difference scheme invented by Scharfetter and Gummel [11] was the first to deal appropriately with most of these difficulties for charge carrier densities described by Boltzmann statistics. Later on, Fichtner, Rose, and Bank [12] extented it to higher dimensions as a finite volume scheme, providing geometric flexibility. Selberherr combined a lot of these results along with his own contributions in his textbook [13] which has become one of the standard reference works in the field.

However, often the Boltzmann approximation is insufficient and one has to impose Fermi–Dirac statistics or other constitutive laws for the charge carrier densities. For this reason, we present recent attempts to generalize the Scharfetter–Gummel approach while still reflecting the physics correctly.

In order to enable readers not familiar with semiconductor simulations to quickly grasp the main points in the discretization process of the van Roosbroeck system, we start by studying a one-dimensional (1D) problem in Section 50.2. Even though we have to postpone some of the solid-state physics to the third section, we can already introduce many key concepts of the standard finite difference ansatz proposed by Scharfetter and Gummel, which we will interpret as a 1D finite volume scheme.

In Section 50.3, we introduce the full van Roosbroeck system in higher spatial dimensions along with its physical concepts. In Section 50.4, we look at its numerical solution. We present the two-point flux finite volume method for the stationary case, which allows to use the Scharfetter–Gummel scheme in higher dimensions and discuss its generalization to more general carrier statistics. The remaining sections are devoted to special numerical aspects: nonlinear solvers, mesh generation, time stepping, the correct computation of terminal currents, and the finite element method. In the final section, we discuss generalizations of the approach and its embedding into more complex physical models.

50.2 How to Get Started: Numerical Solution of the Stationary 1D Semiconductor Equations

This section serves as a quick introduction to convey the main idea behind the van Roosbroeck system and its discretization. The van Roosbroeck system consists of three equations modeling the charge carrier flow as well as the electrostatic potential distribution in a semiconductor device. However, since the physics behind these equations is rather cumbersome and the focus of this chapter are numerical methods, we omit some of these physical details for now and start with a relatively simple 1D problem. Thus, we can highlight the key elements (and difficulties!) in the discretization of the van Roosbroeck system. This approach may appeal to newcomers to the field of optoelectronics. Experienced readers may jump directly to Section 50.3, where we will dive into the physical meaning of this model.

Here, we wish to model the stationary state of a 1D semiconductor device consisting of a homogeneous material on an interval $\Omega = [0, L]$ with Ohmic contacts at each end. This configuration already covers the simulation of p-n and p-i-n junctions.

50.2.1 1D van Roosbroeck Model

The stationary 1D van Roosbroeck system consists of three nonlinear ordinary differential equations for the unknown electrostatic potential $\psi(x)$, the quasi-Fermi potential for electrons $\varphi_n(x)$ and the quasi-Fermi potential for holes $\varphi_p(x)$. It links Poisson's equation for the electric field to the continuity equations for the carrier densities as follows:

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\varepsilon_s \frac{\mathrm{d}}{\mathrm{d}x}\psi\right) = q\Big(C + p(\psi, \varphi_p) - n(\psi, \varphi_n)\Big), \tag{50.1}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} j_n = \quad qR(\psi, \varphi_n, \varphi_p), \tag{50.2}$$

$$\frac{\mathrm{d}}{\mathrm{d}x} j_p = -qR(\psi, \varphi_n, \varphi_p). \tag{50.3}$$

The fluxes or current densities are given by

$$j_n = -q\mu_n n(\psi, \varphi_n)\frac{\mathrm{d}}{\mathrm{d}x}\varphi_n \qquad \text{and} \qquad j_p = -q\mu_p p(\psi, \varphi_p)\frac{\mathrm{d}}{\mathrm{d}x}\varphi_p. \tag{50.4}$$

The electron density n and the hole density p are related to the electric potential ψ as well as the quasi-Fermi potentials of electrons and holes via

$$n(\psi, \varphi_n) = N_c \exp\left(\frac{q(\psi - \varphi_n) - E_c}{k_\mathrm{B}T}\right) \quad \text{and} \quad p(\psi, \varphi_p) = N_v \exp\left(\frac{q(\varphi_p - \psi) + E_v}{k_\mathrm{B}T}\right). \tag{50.5}$$

We stress here that in (50.5) we have used the so-called Boltzmann approximation. In general, the exponentials will be replaced by some monotonically increasing *statistical distribution function* $\mathcal{F}$. This function will be discussed in greater detail in Section 50.3.

Next, we explain the constants and functions in the van Roosbroeck system. The elementary charge q and the Boltzmann constant k_B are universal constants. The (absolute) dielectric permittivity $\varepsilon_\mathrm{s} = \varepsilon_0\varepsilon_\mathrm{r}$ is given as the product of the vacuum dielectric permittivity ε_0 and the relative permittivity of the semiconductor ε_r in static (low frequency) limit. The carrier mobilities μ_n and μ_p, the conduction and valence band densities of states N_c and N_v as well as the conduction and valence band-edge energies E_c and E_v are assumed to be constant even though in general they can vary with the material (e.g., due to abrupt or graded heterojunctions). The temperature T is also assumed to be constant; in general it can be space and even time dependent. The doping profile $C = C(x)$ describes material properties and the function $R(\psi, \varphi_n, \varphi_p)$ models the recombination and generation of electron–hole pairs. The electric field $\mathbf{E}$ is determined by the derivative (in higher dimensions the gradient) of the electrostatic potential

$$\mathbf{E} = -\frac{\mathrm{d}}{\mathrm{d}x}\psi\, \mathbf{e}_x,$$

where $\mathbf{e}_x$ denotes the unit normal along the x axis. A more detailed discussion on the physics behind these different quantities will be given in Section 50.3.

In the literature, authors frequently use a different set of unknowns for the van Roosbroeck system by replacing the quasi-Fermi potentials with the carrier densities n and p. However, we prefer to use the quasi-Fermi potentials for several reasons: they allow to easily model abrupt or graded heterostructures and appear naturally in the thermodynamic description of the van Roosbroeck system since the negative gradients of the quasi-Fermi potentials are the driving force of the currents [14]. Even mathematically, they are more beautiful as they make it possible to write the whole van Roosbroeck system in a gradient form [15] consistent with the principles of nonequilbrium thermodynamics [16]. Also carrier densities as opposed to quasi-Fermi potentials vary drastically in magnitude, which may lead to numerical difficulties.

The system of Equations 50.1 through 50.3 needs to be supplied with boundary conditions at $x = 0$ and $x = L$. For applied external voltages U_1 and U_2, we require Dirichlet boundary conditions at the *Ohmic contacts*, i.e.,

$$\psi(0) = \psi_0(0) + U_1, \qquad \psi(L) = \psi_0(L) + U_2, \tag{50.6}$$

$$\varphi_n(0) = U_1, \qquad \varphi_n(L) = U_2, \tag{50.7}$$

$$\varphi_p(0) = U_1, \qquad \varphi_p(L) = U_2. \tag{50.8}$$

We discuss the meaning of the potential ψ_0 in Section 50.2.2.

50.2.2 Thermodynamic Equilibrium and Local Charge Neutrality

The goal is to enforce that the van Roosbroeck system of (50.1) through (50.3) is consistent with the *thermodynamic equilibrium*, which is a physical state defined by vanishing currents, namely

$$j_n = j_p = 0 \qquad \text{implying} \qquad \varphi_0 := \varphi_n = \varphi_p = \text{const.} \tag{50.9}$$

Without loss of generality, we can set $\varphi_0 = 0$. We also discuss the thermodynamic equilibrium in Section 50.3.4. As a consequence, the three differential equations (50.1) through (50.3) reduce to the 1D nonlinear Poisson equation

$$-\frac{\mathrm{d}}{\mathrm{d}x}\left(\varepsilon_s \frac{\mathrm{d}}{\mathrm{d}x}\psi\right) = q\left(N_v \exp\left(\frac{E_v - q\psi}{k_B T}\right) - N_c \exp\left(\frac{q\psi - E_c}{k_B T}\right) + C\right). \tag{50.10}$$

We supply it with Dirichlet boundary conditions (50.6) for zero voltages $U_1 = U_2 = 0$. The solution of Equation 50.10 with these boundary conditions is the *built-in potential* denoted with ψ_{eq}. In general, we cannot expect to find an analytic expression for ψ_{eq}.

Another important physical concept is *local charge neutrality*, characterized by a vanishing left-hand side in the Poisson equation. Combined with the equilibrium condition (50.9), this leads to

$$0 = q\left(N_v \exp\left(\frac{E_v - q\psi_0}{k_B T}\right) - N_c \exp\left(\frac{q\psi_0 - E_c}{k_B T}\right) + C\right), \tag{50.11}$$

which can be solved for ψ_0 by solving a quadratic equation and omitting the unphysical solution, yielding

$$\psi_0(x) = \frac{E_c + E_v}{2q} - \frac{1}{2}U_T \log\left(\frac{N_c}{N_v}\right) + U_T \operatorname{arcsinh}\left(\frac{C}{2N_{\text{intr}}}\right), \tag{50.12}$$

where the intrinsic carrier density N_{intr} and the *thermal voltage* U_T are given by

$$N_{\text{intr}}^2 = N_c N_v \exp\left(-\frac{E_c - E_v}{k_B T}\right) \qquad \text{and} \qquad U_T = \frac{k_B T}{q}. \tag{50.13}$$

The potential ψ_0 serves two purposes. In Section 50.2.4, it will be used to obtain an initial guess for the nonlinear iteration to solve a discretized version of Equation 50.10. On the other hand, modeling (ideal) Ohmic contacts requires local charge neutrality. Hence, the boundary values in (50.6) are obtained by evaluating the function ψ_0 at the contacts $x = 0$ and $x = L$. The voltage difference between the boundary values $\psi_0(0)$ and $\psi_0(L)$ is called *built-in voltage*.

50.2.3 Discretizing the Poisson Problem

Our goal is to find a discrete solution to Equation 50.1. For this, we will use the *finite volume method* which naturally preserves many important physical properties. To see this, the reader may have a quick look in the table of contents and compare the next two sections devoted to the van Roosbroeck system and its numerical solution. It will become apparent that for many aspects the continuous model and its discretized counterpart are closely related.

We discretize the interval using a (possibly) nonuniform mesh of the form

$$0 =: x_1 < \cdots < x_N := L.$$

We associate each node x_k with a *control volume* ω_k defined by the interior subintervals

$$\omega_k = \left[x_{k-1,k}, x_{k,k+1}\right] \quad \text{for} \quad k = 2, \dots, N-1$$

and the boundary subintervals

$$\omega_1 = \left[x_1, x_{1,2}\right] \quad \text{and} \quad \omega_N = \left[x_{N-1,N}, x_N\right].$$

The boundaries of the kth control volume are given by

$$x_{k-1,k} := \frac{x_{k-1} + x_k}{2} \qquad \text{and} \qquad x_{k,k+1} := \frac{x_k + x_{k+1}}{2}$$

for $k = 2, \dots, N-1$. We point out that we have generated a disjoint partition of the interval, $\Omega = \bigcup_{k=1}^{N} \omega_k$. This discretization along with its notation is illustrated in Figure 50.1.

We introduce a finite volume discretization for Poisson's equation. Using the fundamental theorem of calculus, we integrate Equation 50.1 over the interior control volumes ω_k and obtain Gauss's law of electrodynamics:

$$-\varepsilon_s \left(\left.\frac{\mathrm{d}\psi}{\mathrm{d}x}\right|_{x_{k,k+1}} - \left.\frac{\mathrm{d}\psi}{\mathrm{d}x}\right|_{x_{k-1,k}} \right) = \int_{\omega_k} q\Big(C + p(\psi, \varphi_p) - n(\psi, \varphi_n)\Big)\, \mathrm{d}x. \tag{50.14}$$

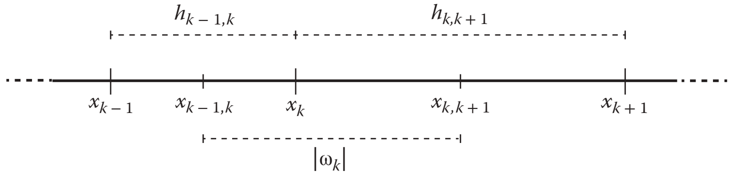


FIGURE 50.1 1D discretization around the kth node along with the corresponding notation.

This equation constitutes a balance law that can be interpreted as follows: the total charge in the kth control volume is given by the difference of the electric displacement $D = -\varepsilon_s \frac{\mathrm{d}}{\mathrm{d}x}\psi$ at its interfaces.

So far, we have derived an integral form of Poisson's equation for every interior control volume. We need approximations of the electric displacements at the boundary of each control volume. For this we employ central finite differences:

$$\left.\frac{\mathrm{d}\psi}{\mathrm{d}x}\right|_{x_{k,k+1}} \approx \frac{\psi_{k+1} - \psi_k}{h_{k,k+1}}, \tag{50.15}$$

with the distance between two neighboring nodes given by

$$h_{k,k+1} := x_{k+1} - x_k.$$

Moreover, the integral over the net charge density is approximated by the rectangle method, evaluating the integrand at node x_k times the size of the control volume $|\omega_k| = x_{k,k+1} - x_{k-1,k}$. That is, the balance equation (50.14) is approximated by

$$-\varepsilon_s\left(\frac{\psi_{k+1} - \psi_k}{h_{k,k+1}} - \frac{\psi_k - \psi_{k-1}}{h_{k-1,k}}\right) = q\Big(C_k + p(\psi_k, \varphi_{p;k}) - n(\psi_k, \varphi_{n;k})\Big)|\omega_k|$$

or, slightly rearranged,

$$\begin{aligned} 0 = -\varepsilon_s\left(\frac{1}{h_{k,k+1}}\psi_{k+1} - \left[\frac{1}{h_{k,k+1}} + \frac{1}{h_{k-1,k}}\right]\psi_k + \frac{1}{h_{k-1,k}}\psi_{k-1}\right)& \\ - q\Big(C_k + p(\psi_k, \varphi_{p;k}) - n(\psi_k, \varphi_{n;k})\Big)|\omega_k|&, \end{aligned} \tag{50.16}$$

where $k = 2, \ldots, N-1$. The subindices denote evaluation at a given node. So, for example, ψ_k means $\psi(x_k)$ as well as $\varphi_{n;k}$ indicates $\varphi_n(x_k)$. We use this nodal notation frequently within this chapter. To incorporate the Dirichlet boundary conditions (50.6), we simply impose two more equations for the endpoints, namely

$$\psi_1 = \psi_0(0) + U_1 \qquad \text{and} \qquad \psi_N = \psi_0(L) + U_2. \tag{50.17}$$

50.2.4 Solving the Discrete Nonlinear Poisson Problem

Using the finite volume discretization scheme we have just described, we solve the nonlinear Poisson equation (50.16) with Dirichlet boundary conditions (50.17) in thermodynamic equilibrium ($U_1 = U_2 = 0$) to obtain the built-in potential ψ_{eq}. Recalling that there are $N-2$ interior and two boundary nodes, we end up with a system of N variables which can be summarized by the nonlinear discrete system

$$\mathbf{0} = \mathbf{F}_1(\boldsymbol{\psi}_{\mathrm{eq}}),$$

where $\boldsymbol{\psi}_{\mathrm{eq}} = (\psi_{\mathrm{eq}}(x_k))_{k=1}^{N}$ is the vector of the nodal values. Solving this system is not straightforward due to its inherent nonlinearity. One approach to obtain a solution is Newton's method, which we discuss in more detail in Section 50.5. Newton's method converges quadratically—if one has a starting guess sufficiently close to the solution [17].

50.2.4.1 Solution Procedure: Nonlinear Poisson Problem

1. The easiest way to obtain a starting guess for the solution of Equation 50.16 with Dirichlet boundary conditions (50.17) is to neglect the left-hand side. That is, we use the local charge neutrality condition (50.11) for each node, which we can solve explicitly for ψ_0 via (50.12). With this function, we determine the components (nodal values) of our starting guess for the nonlinear Poisson problem $\boldsymbol{\psi_0}$ via

$$\boldsymbol{\psi_0} = (\psi_0(x_k))_{k=1}^{N}.$$

2. Solving the discrete nonlinear Poisson problem via Newton's method with the starting guess $\boldsymbol{\psi_0}$ yields the built-in potential vector solution $\boldsymbol{\psi}_{\text{eq}}$.

This solution procedure for the nonlinear Poisson equation for a p-i-n structure is illustrated in Figure 50.2. In the following, we address how to discretize the continuity equations. A task which turns out to be nontrivial.

50.2.5 Discretizing the Continuity Equations Using the Scharfetter–Gummel Scheme

Now, we turn our attention to the continuity equations (50.2) and (50.3). We proceed just as for Poisson's equation: we integrate over the control volumes $\omega_k = [x_{k-1,k}, x_{k,k+1}]$, apply the fundamental theorem of calculus for the left-hand sides and approximate the integral of the recombination term as before. For the interior control volumes we obtain

$$\begin{aligned} 0 &= j_{n;k,k+1} - j_{n;k-1,k} - qR(\psi_k, \varphi_{n;k}, \varphi_{p;k})|\omega_k|, \\ 0 &= j_{p;k,k+1} - j_{p;k-1,k} + qR(\psi_k, \varphi_{n;k}, \varphi_{p;k})|\omega_k|, \end{aligned} \tag{50.18}$$

where $k = 2, \ldots, N-1$. Similar to before, these equations constitute a balance law in integral form which we can interpret as follows: the carrier densities within the control volume ω_k change either due to in- and outflow or by recombination.

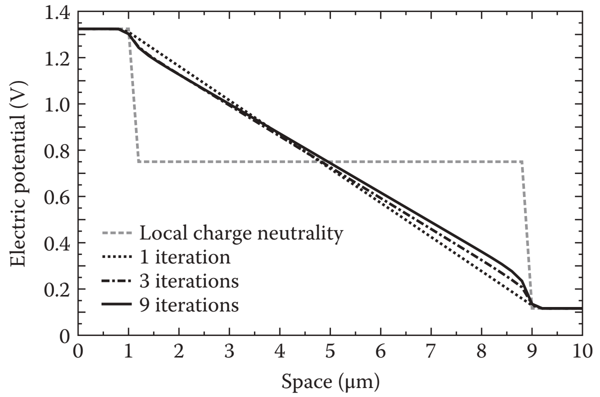


FIGURE 50.2 Convergence of the electric potential toward the thermodynamic equilibrium solution (built-in potential) for a GaAs p-i-n diode on an equidistant mesh with 51 nodes. The locally electroneutral initialization quickly converges to the built-in potential by means of a Newton iteration. The device has total length of 10 μm (1 μm n-doped with $N_{\text{D}} = 10^{16}\,\text{cm}^{-3}$, 8 μm intrinsic, 1 μm p-doped with $N_{\text{A}} = 10^{17}\,\text{cm}^{-3}$. Standard parameters for GaAs at $T = 300\,\text{K}$: $E_v = 0\,\text{eV}$, $E_c = 1.424\,\text{eV}$, $N_v = 9.0 \times 10^{18}\,\text{cm}^{-3}$, $N_c = 4.7 \times 10^{17}\,\text{cm}^{-3}$, and $\varepsilon_r = 12.9$.

The fluxes (50.18) need to be expressed in terms of our set of unknowns, namely the nodal values of the electrostatic and the quasi-Fermi potentials. Deriving a suitable approximation for the current expressions $j_{n;k,k+1}$ and $j_{p;k,k+1}$ describing the carrier exchange between neighboring control volumes ω_k and ω_{k+1} is crucial for numerical simulations. Since standard discretization techniques such as central finite differences may lead to unphysical behavior (see Section 50.2.7), the derivation of such an approximation requires particular care.

For the Boltzmann approximation (50.5), a numerically stable scheme was invented in 1969 by Scharfetter and Gummel [11]. In fact, it has been discovered independently several times [18–20]. Its derivation is based on integrating the flux along the interval $[x_k, x_\ell]$ between two neighboring discretization nodes x_k and x_ℓ under the assumption of constant current density and constant electric field. That is, either $\ell = k-1$ or $\ell = k+1$. We illustrate the approach for the electron current density j_n.

In order to obtain a flux approximation, we will make two important assumptions, namely

1. The electric field along each edge is constant. This implies that the derivative of the electric potential $\frac{\mathrm{d}\psi}{\mathrm{d}x}$ along the edge $\left[x_k, x_\ell\right]$ is approximated by $\frac{\psi_\ell - \psi_k}{h_{k,\ell}}$, where again $\psi_\ell = \psi(x_\ell)$ and $\psi_k = \psi(x_k)$. This is consistent with the central difference approximation for the electric field in the Poisson equation, see (50.15).
2. The current density j_n is assumed to be constant along the edge.

The second assumption means that the current fluxes through the boundary from one control volume to another are conserved, which implies that the derivative of the flux along the interval $[x_k, x_\ell]$ vanishes. This means that we assume that there is no recombination along the edge between neighboring nodes. Hence, we obtain for the electron flux (the hole flux follows analogously)

$$\frac{\mathrm{d}}{\mathrm{d}x} j_n = \frac{\mathrm{d}}{\mathrm{d}x}\left(-q\mu_n N_c \exp\left(\frac{q(\psi(x) - \varphi_n(x)) - E_c}{k_\mathrm{B} T}\right)\frac{\mathrm{d}}{\mathrm{d}x}\varphi_n(x)\right) = 0. \tag{50.19}$$

We supply the differential equation with the boundary conditions

$$\varphi_n(x_k) = \varphi_{n;k} \qquad \text{and} \qquad \varphi_n(x_\ell) = \varphi_{n;\ell}. \tag{50.20}$$

The values $\varphi_{n;k}$ and $\varphi_{n;\ell}$ have to be determined. But for now we assume they are given. Equations 50.19 and 50.20 constitute a two-point boundary value problem. Upon integrating once and multiplication by the integrating factor

$$M(x) = \exp\left(-\frac{\psi_\ell - \psi_k}{h_{k,\ell} U_\mathrm{T}} x\right),$$

we obtain for the integration constant $j_{n;k,\ell}$ the equation

$$j_{n;k,\ell} M(x) = q\mu_n N_c U_\mathrm{T} \frac{\mathrm{d}}{\mathrm{d}x}\left(M(x) \exp\left(\frac{q(\psi(x) - \varphi_n(x)) - E_c}{k_\mathrm{B}\mathrm{T}}\right)\right).$$

Integrating once more from x_k to x_ℓ and using the boundary conditons (50.20) yields after some algebraic manipulations the Scharfetter–Gummel expression for the current flux along the edge $[x_k, x_\ell]$ between neighboring control volumes

$$j_{n;k,\ell} = j_{n;k,\ell}(\psi_k, \psi_\ell, \varphi_{n;k}, \varphi_{n;\ell}) = -\frac{q\mu_n U_{\mathrm{T}}}{h_{k,\ell}} \left[B\left(-\frac{\psi_\ell - \psi_k}{U_{\mathrm{T}}}\right) N_c \exp\left(\frac{q\left(\psi_k - \varphi_{n;k}\right) - E_c}{k_{\mathrm{B}} T}\right) \right.$$
$$\left. - B\left(\frac{\psi_\ell - \psi_k}{U_{\mathrm{T}}}\right) N_c \exp\left(\frac{q\left(\psi_\ell - \varphi_{n;\ell}\right) - E_c}{k_{\mathrm{B}} \mathrm{T}}\right) \right], \tag{50.21}$$

where we have introduced the *Bernoulli function*

$$B(x) = \frac{x}{\exp(x) - 1}.$$

This function smoothly interpolates between zero and $-x$, see Figure 50.3. Remembering the relationships (50.5), the flux can be brought into the simpler form

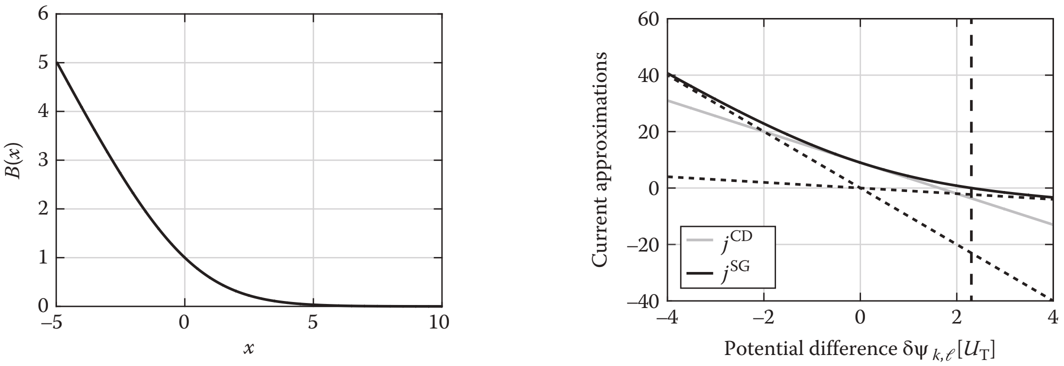
$$j_{n;k,\ell} = -\frac{q\mu_n U_{\mathrm{T}}}{h_{k,\ell}} \left[B\left(-\frac{\psi_\ell - \psi_k}{U_{\mathrm{T}}}\right) n_k - B\left(\frac{\psi_\ell - \psi_k}{U_{\mathrm{T}}}\right) n_\ell \right]. \tag{50.22}$$

An important property of this scheme is its consistency with the thermodynamic equilibrium. That is, similarly to (50.9), we have

$$j_{n;k,\ell} = 0 \quad \text{for} \quad \varphi_{n;k} = \varphi_{n;\ell} = \text{const.} \tag{50.23}$$

For vanishing electric fields corresponding to $\delta\psi_{k,\ell} := \psi_\ell - \psi_k = 0$, the flux $j_{n;k,\ell}$ reduces to a purely diffusive flux

$$j_{n;k,\ell} = -q\left(-D_n \frac{n_\ell - n_k}{h_{k,\ell}}\right).$$



(a) (b)

FIGURE 50.3 On the left the Bernoulli function $B(x)$ is shown. The picture on the right shows different current approximations for $q\mu_n U_{\mathrm{T}}/h_{k,\ell} = 1$, $n_\ell = 10$, and $n_k = 1$ in terms of the potential difference $\delta\psi_{k,\ell}$. The dashed black line denotes the equilibrium potential difference. The dotted lines show the drift currents, i.e., the asymptotics of the Scharfetter–Gummel flux (50.24) for large positive and for large negative arguments. The central-difference flux is denoted with j^{CD} and the Scharfetter–Gummel flux with j^{SG}.

The diffusion constant D_n is defined via the so-called *Einstein relation* $D_n = \mu_n U_\mathrm{T}$, which we discuss in Section 50.3.5 in more detail. For large electric fields, we obtain asymptotically a drift flux

$$j_{n;k,\ell} = \begin{cases} -q\mu_n n_k \frac{\psi_\ell - \psi_k}{h_{k,\ell}}, & \text{for } \delta\psi_{k,\ell} \gg +U_\mathrm{T}, \\ -q\mu_n n_\ell \frac{\psi_\ell - \psi_k}{h_{k,\ell}}, & \text{for } \delta\psi_{k,\ell} \ll -U_\mathrm{T}. \end{cases} \tag{50.24}$$

These two extreme cases show that the Scharfetter–Gummel scheme naturally encompasses drift and diffusive currents depending on the strength of the local electric field, see Figure 50.3.

Using analogous considerations, we can derive a current expression for the holes

$$j_{p;k,\ell} = \frac{q\mu_p U_\mathrm{T}}{h_{k,\ell}} \left[B\left(\frac{\psi_\ell - \psi_k}{U_\mathrm{T}}\right) p_k - B\left(-\frac{\psi_\ell - \psi_k}{U_\mathrm{T}}\right) p_\ell \right].$$

Since we have an approximation for the fluxes at the control volume boundary, we can derive from the discrete system (50.18) nonlinear equations for the interior nodes in terms of our set of unknowns $(\psi, \varphi_n, \varphi_p)$.

We stress once more that the previous discussion is only valid for the Boltzmann approximation made in (50.5). A discussion on more general distribution functions follows in Sections 50.3.5 and 50.4.2.

50.2.6 Solving the Full Discretized van Roosbroeck System

So far we have described how to derive a set of discrete and nonlinear equations from the 1D van Roosbroeck system ((50.1)–(50.3)), namely

$$\begin{aligned} 0 &= -\varepsilon_s \left(\frac{1}{h_{k,k+1}} \psi_{k+1} - \left[\frac{1}{h_{k,k+1}} + \frac{1}{h_{k-1,k}} \right] \psi_k + \frac{1}{h_{k-1,k}} \psi_{k-1} \right) \\ &\qquad - q\Big(C_k + p(\psi_k, \varphi_{p;k}) - n(\psi_k, \varphi_{n;k}) \Big) |\omega_k|, \\ 0 &= j_{n;k,k+1}(\psi_k, \psi_{k+1}, \varphi_{n;k}, \varphi_{n;k+1}) - j_{n;k-1,k}(\psi_{k-1}, \psi_k, \varphi_{n;k-1}, \varphi_{n;k}) - qR(\psi_k, \varphi_{n;k}, \varphi_{p;k})|\omega_k|, \\ 0 &= j_{p;k,k+1}(\psi_k, \psi_{k+1}, \varphi_{p;k}, \varphi_{p;k+1}) - j_{p;k-1,k}(\psi_{k-1}, \psi_k, \varphi_{p;k-1}, \varphi_{p;k}) + qR(\psi_k, \varphi_{n;k}, \varphi_{p;k})|\omega_k|, \end{aligned} \tag{50.25}$$

with $k = 2, \ldots, N-1$. For the boundary nodes x_1 and x_N, the equations are given by the boundary conditions ((50.6)–(50.8)):

$$\begin{aligned} \psi_1 &= \psi(0) = \psi_0(0) + U_1, & \varphi_{n,1} &= \varphi_n(0) = U_1, & \varphi_{p,1} &= \varphi_p(0) = U_1, \\ \psi_N &= \psi(L) = \psi_0(L) + U_2, & \varphi_{n,N} &= \varphi_n(L) = U_2, & \varphi_{p,N} &= \varphi_p(L) = U_2. \end{aligned}$$

Just as for the Poisson problem we assume there are N control volumes in total. We end up with a system of $3N$ variables which we can summarize by the nonlinear discrete system

$$\mathbf{0} = \mathbf{F}(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) := \begin{pmatrix} \mathbf{F}_1(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) \\ \mathbf{F}_2(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) \\ \mathbf{F}_3(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) \end{pmatrix},$$

where

$$\boldsymbol{\psi} = (\psi(x_k))_{k=1}^N, \qquad \boldsymbol{\varphi}_n = (\varphi_n(x_k))_{k=1}^N, \qquad \boldsymbol{\varphi}_p = (\varphi_p(x_k))_{k=1}^N.$$

Again, we employ Newton's method to solve this system. We aim to guarantee an appropriate starting guess by starting from thermodynamic equilibrium where the three equations reduce to one. The solution of the full van Roosbroeck system with an applied bias is then obtained via an iteration technique that we state here.

50.2.6.1 Solution Procedure: Full van Roosbroeck System

1. We assume thermodynamic equilibrium and set the constant quasi-Fermi potentials to $\varphi_0 = 0$. With the solution procedure introduced in Section 50.2.4, we obtain a solution vector $\boldsymbol{\psi}_{\text{eq}}$ for the nonlinear Poisson problem with Dirichlet boundary conditions.
2. We denote with $\boldsymbol{\varphi}_{\boldsymbol{n}}^{\boldsymbol{0}}$ and $\boldsymbol{\varphi}_{\boldsymbol{p}}^{\boldsymbol{0}}$ the constant vectors, consisting of the equilibrium value for the quasi-Fermi potentials, namely

   $$\boldsymbol{\varphi}_{\boldsymbol{n}}^{\boldsymbol{0}} = (\varphi_0)_{i=1}^N \qquad \text{and} \qquad \boldsymbol{\varphi}_{\boldsymbol{p}}^{\boldsymbol{0}} = (\varphi_0)_{i=1}^N.$$

   We choose zero boundary voltages $U_1 = U_2 = 0$ and remember that we have obtained $\psi_0(0)$ and $\psi_0(L)$ via the charge neutrality condition. Thus we have determined the boundary conditions ((50.6)–(50.8)). Using the equilibrium vector $(\boldsymbol{\psi}_{\text{eq}}, \boldsymbol{\varphi}_{\boldsymbol{n}}^{\boldsymbol{0}}, \boldsymbol{\varphi}_{\boldsymbol{p}}^{\boldsymbol{0}})$ as a new starting guess, we can now solve the full discrete system (50.25) with these boundary conditions via Newton's method, yielding a solution $(\boldsymbol{\psi}^{\boldsymbol{1}}, \boldsymbol{\varphi}_{\boldsymbol{n}}^{\boldsymbol{1}}, \boldsymbol{\varphi}_{\boldsymbol{p}}^{\boldsymbol{1}})$. That is, we compute the equilibrium state of the full van Roosbroeck system using the equilibrium of the Poisson equation as a starting guess. We point out that solution and starting guess should agree up to machine precision by construction.
3. We slightly increase U_1 and U_2 to generate a "small" bias $\delta U = U_2 - U_1$. What small here means is not very precise and needs to be determined via trial and error. Now we solve the full discrete system (50.25) via Newton's method with the starting guess $(\boldsymbol{\psi}^{\boldsymbol{1}}, \boldsymbol{\varphi}_{\boldsymbol{n}}^{\boldsymbol{1}}, \boldsymbol{\varphi}_{\boldsymbol{p}}^{\boldsymbol{1}})$ to obtain a new solution $(\boldsymbol{\psi}^{\boldsymbol{2}}, \boldsymbol{\varphi}_{\boldsymbol{n}}^{\boldsymbol{2}}, \boldsymbol{\varphi}_{\boldsymbol{p}}^{\boldsymbol{2}})$.
4. Now we iterate the last step: gradually increase the bias and use the old solution as starting guess for the new one.

Figures 50.4 and 50.5 show the numerical results for the p-i-n structure, already introduced in Section 50.2.4. They depict the band edge energies, the quasi-Fermi energy levels $E_{F_n} = -q\varphi_n$ and $E_{F_p} = -q\varphi_p$ as well as the carrier densities for different bias values. Additionally, the current voltage characteristics for different n-dopings are shown.

50.2.7 Why Not Use Central Finite Differences for the Fluxes?

Thus far, we have discussed the Scharfetter–Gummel scheme to approximate the flux at the control volume interface. One might ask why not use a simpler flux approximation, for example, central finite differences. This question we address now.

For simplicity we assume a linear electrostatic potential. In this case

$$\frac{\mathrm{d}}{\mathrm{d}x}\psi = \psi(1) - \psi(0) =: \delta\psi.$$

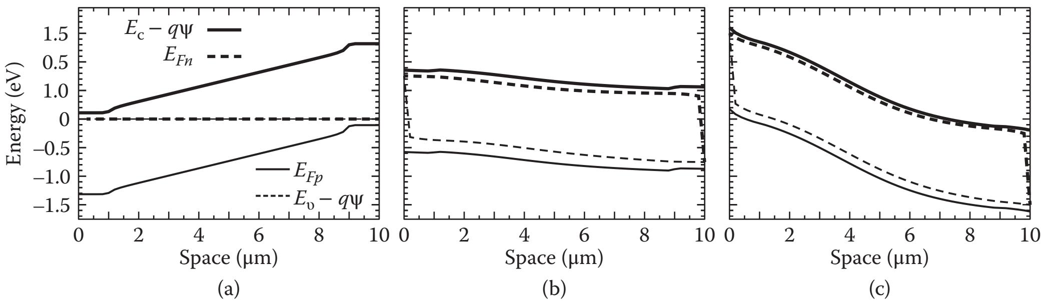


FIGURE 50.4 Computed band edge energy and quasi-Fermi levels for a GaAs p-i-n diode at different values of the applied voltage. (a) Thermodynamic equilibrium (off-state, 0V), (b) $U = 1.5$ V (flat band), (c) $U = 3.0$ V. Device has total length of 10 μm (1 μm n-doped with $N_D = 10^{16}$ cm^{-3}, 8 μm intrinsic, 1 μm p-doped with $N_A = 10^{17}$ cm^{-3}. The parameters used are standard parameters for GaAs as given in the caption of Figure 50.2 and $\mu_n = 8500$ cm^2 V^{-1} s^{-1}, $\mu_p = 400$ cm^2 V^{-1} s^{-1}, $\tau_n = \tau_p = 1$ ns, $C_n = C_p = 1 \times 10^{-29}$ cm^6 s^{-1}, and $r_{\mathrm{spont}} = 1 \times 10^{-10}$ cm^3 s^{-1}.

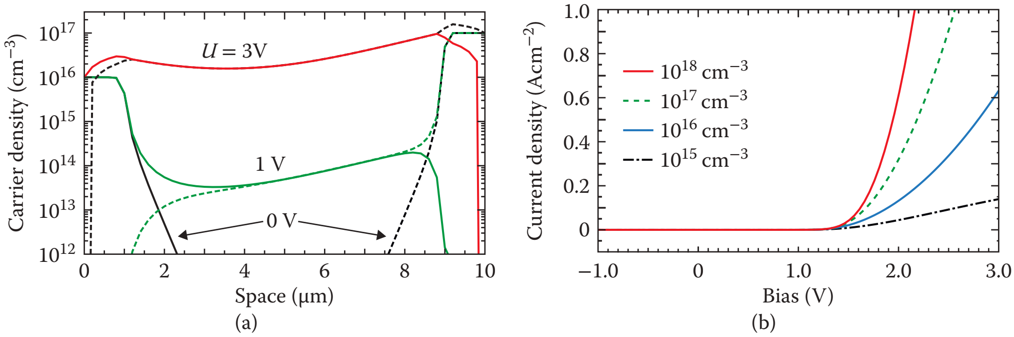


FIGURE 50.5 (a) Carrier densities in the p-i-n-diode at different applied voltages. Electron densities are shown as solid lines, holes are illustrated by dashed lines. One observes how the intrinsic region gets populated by free carriers while increasing the applied bias. (b) Current–voltage curves for the same device with altered values of the donor density N_{D}. Same parameters as in Figure 50.4.

We focus on the continuity equation for electrons on the domain $\Omega = [0, 1]$ which results in a linear differential equation of second order,

$$\frac{\mathrm{d}}{\mathrm{d}x} j_n = qD_n \frac{\mathrm{d}^2}{\mathrm{d}x^2} n - q\mu_n \delta\psi \frac{\mathrm{d}}{\mathrm{d}x} n = 0. \tag{50.26}$$

In this equation, we have introduced the drift-diffusion form of the (electron) flux

$$j_n = q\left(D_n \frac{\mathrm{d}}{\mathrm{d}x} n - n\mu_n \frac{\mathrm{d}}{\mathrm{d}x}\psi\right), \tag{50.27}$$

where we have used again the Einstein relation $D_n = \mu_n U_{\mathrm{T}}$. We omit the physical derivation of the flux here. It is only important to know that this way of characterizing the flux is equivalent to Equation 50.4.

We have already seen how to solve this type of equation numerically by solving *local* two-point boundary value problems for the flux, which yields the Scharfetter–Gummel scheme (50.22). Of course, there are many different ways to discretize the flux (50.27) at a control volume interface $x_{k,\ell}$, needed in (50.18). For example, looking at Equation 50.27 one might consider replacing the derivative of the electron density $\frac{\mathrm{d}}{\mathrm{d}x} n$ with a central finite difference and the density n with an average. That is, at the control volume interface

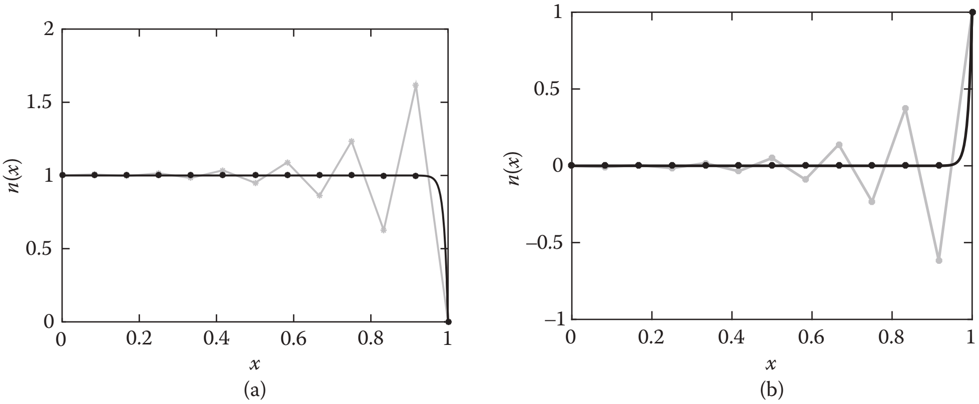


FIGURE 50.6 The exact solution (black) of Equation 50.26 and its central difference approximation (gray) on a uniform mesh with mesh width $h = 1/12$. The boundary conditions (50.28) lead to a numerical solution which violates the maximum principle (left). On the other hand, the boundary conditions (50.29) produce a numerical solution which violates the positivity of the density (right). Here $qD_n = 0.01$ and $q\mu_n\delta\psi = 1$.

we make the approximation

$$j_{n;k,\ell}^{\mathrm{CD}} = qD_n\frac{n_\ell - n_k}{h_{k,\ell}} - q\mu_n\delta\psi\,\frac{n_\ell + n_k}{2} = -\frac{q\mu_n U_{\mathrm{T}}}{h_{k,\ell}}\left(n_k - n_\ell + \frac{\delta\psi_{k,\ell}}{U_{\mathrm{T}}}\,\frac{n_\ell + n_k}{2}\right).$$

This, however, is a very bad idea since the scheme will become unstable, resulting under certain conditions in large oscillations. To show this, we supply Equation 50.26 with two different types of boundary conditions:

$$n(0) = 1 \qquad \text{and} \qquad n(1) = 0 \tag{50.28}$$

as well as

$$n(0) = 0 \qquad \text{and} \qquad n(1) = 1. \tag{50.29}$$

Then, the discretization using central finite differences shows large oscillations for $\mu_n\delta\psi/D_n = 100$, which can be seen in Figure 50.6. For the first set of boundary conditions, we can see that the *maximum principle* is violated, i.e., the solution exceeds both boundary values [21,22]. For the second set of boundary conditions, the numerically computed density becomes negative. Both observations constitute huge violations of basic physical laws and are unacceptable in applications.

The problem here is that the diffusion constant D_n is small compared to the drift constant $\mu_n\delta\psi$. We point out, that in this special 1D case of vanishing recombination the Scharfetter–Gummel scheme, on the other hand, yields *exact* nodal values due to its construction. The Scharfetter–Gummel and the central difference flux are compared in Figure 50.3.

50.3 Van Roosbroeck System

In this section, we make several generalizations to the van Roosbroeck system ((50.1)–(50.3)). In particular, we make the model time dependent and allow higher spatial dimensions. Before stating the equations, we discuss the physical concepts we want to encode in the model.

As in the 1D case, the van Roosbroeck system consists of three equations: the Poisson equation and two continuity equations for the carrier densities. The Poisson equation describes the electric field $\mathbf{E} = -\nabla\psi$ which is generated by a scalar electric potential $\psi(\mathbf{x}, t)$ in the presence of a free charge carrier density. In a (doped) bipolar semiconductor device, this charge density has three ingredients: the density of free (negatively charged) electrons $n(\mathbf{x}, t)$ occupying the conduction band, the density of (positively charged) holes $p(\mathbf{x}, t)$ occupying the valence band and the density of ionized built-in dopants $C(\mathbf{x}) = N_{\mathrm{D}}^{+}(\mathbf{x}) - N_{\mathrm{A}}^{-}(\mathbf{x})$, where N_{D}^{+} denotes the density of singly ionized donor atoms and N_{A}^{-} is the density of singly ionized acceptor atoms.

The continuity equations (in differential form), on the other hand, model the flow of the charge carrier densities due to diffusion and drift governed by the self-consistent electric field, which is generated by the net charge density. Furthermore, the recombination and generation of electron–hole pairs influences the electron and hole densities.

50.3.1 The van Roosbroeck System of Equations

For a bounded spatial domain $\Omega \subset \mathbb{R}^d$ where $d \in \{1, 2, 3\}$, the van Roosbroeck system consists of three coupled nonlinear PDEs of the form

$$-\nabla \cdot (\varepsilon_s \nabla\psi) = q\,(p - n + C), \tag{50.30}$$

$$-q\partial_t n + \nabla \cdot \mathbf{j}_n = qR, \tag{50.31}$$

$$q\partial_t p + \nabla \cdot \mathbf{j}_p = -qR, \tag{50.32}$$

for $\mathbf{x} \in \Omega$ and $t \in [0, T]$. The current densities in Equations 50.31 and 50.32 are given by the usual expressions

$$\mathbf{j}_n = -q\mu_n n \nabla\varphi_n \qquad \text{and} \qquad \mathbf{j}_p = -q\mu_p p \nabla\varphi_p. \tag{50.33}$$

That is, the negative gradients of the quasi-Fermi potentials are the driving forces of the currents [14]. These relationships correspond directly to Equation 50.4.

Within the framework of effective mass approximation [2], the densities of free carriers in a solid are given by

$$n = N_c \mathcal{F}\left(\frac{q(\psi - \varphi_n) - E_c}{k_{\mathrm{B}}T}\right) \qquad \text{and} \qquad p = N_v \mathcal{F}\left(\frac{E_v - q(\psi - \varphi_p)}{k_{\mathrm{B}}T}\right). \tag{50.34}$$

Equations 50.34 indicate that the electric potential effectively leads to a bending of the energy band edge levels and thus a nonlinear, self-consistent coupling to the carrier densities is achieved. We assume a globally constant temperature for both carrier species and the crystal lattice.

The function $\mathcal{F}$ describes the occupation of energy states in the semiconductor. Since it plays an important role in the following discussion, we will define and discuss it later in Section 50.3.5. We only point out here that Equations 50.34 is analogous to Equations 50.5 when choosing for $\mathcal{F}$ the exponential function, the so-called *Boltzmann approximation*.

50.3.2 Initial and Boundary Conditions

The system ((50.30)–(50.32)) must be supplemented with initial and boundary conditions. The initial conditions at time $t = 0$ are given by the initial distributions ψ^I, φ_n^I and φ_p^I, i.e.,

$$\psi(\mathbf{x},0) = \psi^I(\mathbf{x}), \qquad \varphi_n(\mathbf{x},0) = \varphi_n^I(\mathbf{x}), \qquad \varphi_p(\mathbf{x},0) = \varphi_p^I(\mathbf{x}) \quad \text{for } \mathbf{x} \in \Omega.$$

Regarding the boundary conditions, we briefly discuss the most important case where the boundary of the domain Ω can be decomposed into Ohmic contacts, a gate contact and *artificial* interfaces, i.e.,

$$\partial\Omega = \left(\bigcup_{\alpha=1}^{N_O} \Gamma_{O,\alpha}\right) \cup \Gamma_G \cup \Gamma_A.$$

Semiconductor–metal interfaces, such as Ohmic contacts, are modeled by Dirichlet boundary conditions. For any Ohmic contact $\Gamma_{O,\alpha}$ with $\alpha = 1, \dots, N_O$, we set for all $\mathbf{x} \in \Gamma_{O,\alpha}$ and $t \in [0, T]$

$$\psi(\mathbf{x},t) = \psi_0(\mathbf{x}) + U_\alpha(t), \tag{50.35}$$

$$\varphi_n(\mathbf{x},t) = U_\alpha(t), \tag{50.36}$$

$$\varphi_p(\mathbf{x},t) = U_\alpha(t), \tag{50.37}$$

where U_α denotes the corresponding externally applied contact voltage. The value ψ_0 at the boundary is defined by local charge neutrality similar to (50.11), where the exponential is now replaced by $\mathcal{F}$:

$$0 = N_v \mathcal{F}\left(\frac{E_v - q\psi_0(\mathbf{x})}{k_B T}\right) - N_c \mathcal{F}\left(\frac{q\psi_0(\mathbf{x}) - E_c}{k_B T}\right) + C(\mathbf{x}).$$

We just remark that, in general, this equation yields no analytical solution, and therefore its solution needs to be obtained by a nonlinear solver. The boundary conditions for the more advanced nonlinear semiconductor–metal interfaces (Schottky contacts) can be found in [13,23].

Gate contacts are modeled by Robin boundary conditions for the electrostatic potential and homogeneous Neumann boundary conditions for the quasi-Fermi potentials

$$\left.\begin{aligned} \varepsilon_s \nabla\psi(\mathbf{x},t)\cdot\boldsymbol{\nu} + \tfrac{\varepsilon_{ox}}{d_{ox}}(\psi(\mathbf{x},t) - U_G(t)) = 0, \\ \mathbf{j}_n(\mathbf{x},t)\cdot\boldsymbol{\nu} = \mathbf{j}_p(\mathbf{x},t)\cdot\boldsymbol{\nu} = 0, \end{aligned}\right\} \quad \text{for all } \mathbf{x} \in \Gamma_G \text{ and } t \in [0,T], \tag{50.38}$$

where ε_{ox} and d_{ox} are the absolute dielectric permittivity and the thickness of the oxide, respectively. The function $U_G(t)$ is the applied voltage at the outside of the insulating gate oxide at Γ_G. Here $\boldsymbol{\nu}$ denotes the outer normal vector on the interface.

On the remaining (artificial) interfaces, one typically imposes homogeneous Neumann boundary conditions (natural boundary conditions), namely

$$\nabla\psi(\mathbf{x},t)\cdot\boldsymbol{\nu} = \mathbf{j}_n(\mathbf{x},t)\cdot\boldsymbol{\nu} = \mathbf{j}_p(\mathbf{x},t)\cdot\boldsymbol{\nu} = 0 \qquad \text{for all } \mathbf{x} \in \Gamma_A \text{ and } t \in [0,T]. \tag{50.39}$$

50.3.3 Recombination Processes

The net recombination rate R on the right-hand side of Equations 50.31 and 50.32 describes the radiative and nonradiative generation or recombination of carriers due to thermal excitation and various scattering effects. We assume that the recombination rate $R(n, p)$ is given by the sum of the most common processes, namely the Shockley–Read–Hall recombination R_{SRH}, the spontaneous radiative recombination R_{rad} and the Auger recombination R_{Auger}. All of these rates are of the form

$$R(n, p) = r(n, p)np\left(1 - \exp\left(\frac{q\varphi_n - q\varphi_p}{k_{\mathrm{B}}T}\right)\right), \tag{50.40}$$

where $r(n, p)$ is a model-dependent generation-recombination rate [2,13,24]. In Figure 50.7, one finds the definitions of these rates together with a schematic illustration of the corresponding processes. For Boltzmann statistics, Equation 50.40 is equivalent to the widely used $R(n, p) = r(n, p)(np - N_{\mathrm{intr}}^2)$, where N_{intr} is the intrinsic carrier density defined in (50.13).

50.3.4 Thermodynamic Equilibrium

The thermodynamic equilibrium is characterized by vanishing current fluxes $\mathbf{j}_n(\mathbf{x}, t) = \mathbf{j}_p(\mathbf{x}, t) = \mathbf{0}$. As a consequence, $R(n, p) = 0$ and the quasi-Fermi potentials assume constant values which due to (50.40) are equal:

$$\varphi_0 := \varphi_n = \varphi_p = \text{const.}$$

Without loss of generality, we set $\varphi_0 = 0$. Therefore, the van Roosbroeck system of (50.30)–(50.32) reduces to the nonlinear Poisson equation

$$-\nabla \cdot \varepsilon_s \nabla \psi = q\left(N_v \mathcal{F}\left(\frac{E_v - q\psi}{k_{\mathrm{B}}T}\right) - N_c \mathcal{F}\left(\frac{q\psi - E_c}{k_{\mathrm{B}}T}\right) + C\right). \tag{50.41}$$

We supply it with Dirichlet boundary conditions (50.35) for zero voltages $U_\alpha = 0$ for $\alpha = 1, \ldots, N_O$ and homogeneus Neumann boundary conditions (50.39) elsewhere. The solution of (50.41) with these boundary conditions defines the *built-in potential* denoted with ψ_{eq}.

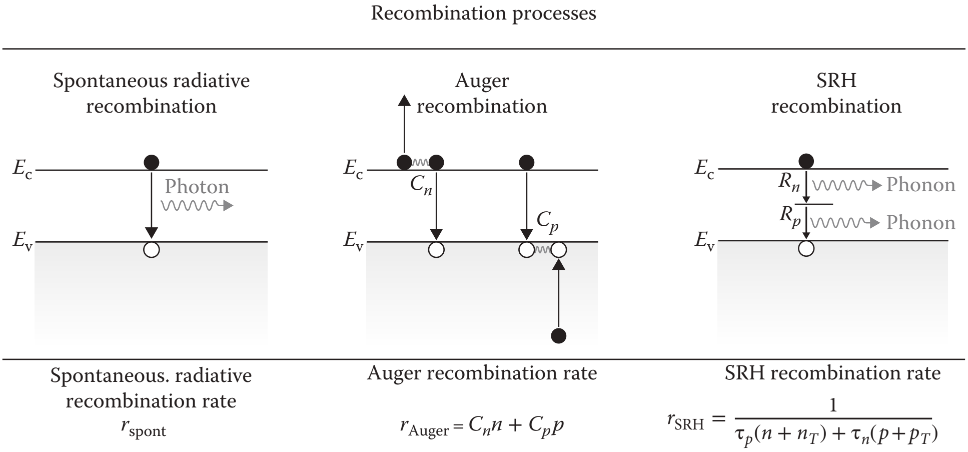

Recombination processes
Spontaneous radiative recombination
Auger recombination
SRH recombination
E_c
E_v
Photon
C_n
C_p
R_n
R_p
Phonon
Phonon
Spontaneous. radiative recombination rate
r_{spont}
Auger recombination rate
$r_{\mathrm{Auger}} = C_n n + C_p p$
SRH recombination rate
$r_{\mathrm{SRH}} = \frac{1}{\tau_p(n + n_T) + \tau_n(p + p_T)}$

FIGURE 50.7 This figure illustrates three different recombination processes. The carrier life times τ_n, τ_p, the reference carrier densities n_T, p_T, and the coefficients r_{spont}, C_n, C_p are material-dependent parameters.

50.3.5 Non-Boltzmann Statistics and Generalized Einstein Relation

Now, we address the statistical distribution function $\mathcal{F}$ which appears in Equation 50.34. It describes the relationship between the density of the free carriers as well as the electrostatic potential and the quasi-Fermi potentials [2].

Assuming that the electrons in the conduction band are in quasi equilibrium, i.e., they are described by the quasi-Fermi level E_{F_n}, the electron density can be introduced as a convolution integral of the density of states DOS(E) with the Fermi–Dirac distribution function

$$n = \int_{-\infty}^{\infty} \mathrm{DOS}(E) \frac{1}{\exp\left(\frac{E-E_{F_n}}{k_\mathrm{B}T}\right)+1} \mathrm{d}E. \tag{50.42}$$

It is possible to express this convolution as a product of the effective density of states N_c and a nondimensionalized statistical distribution function $\mathcal{F}$, i.e.,

$$n(\eta) = N_c\mathcal{F}(\eta). \tag{50.43}$$

We consider organic and inorganic semiconductors. For inorganic, three-dimensional (3D) bulk semiconductors with parabolic bands the statistical distribution function is given by the Fermi–Dirac integral of order 1/2 [2], which can be approximated by the Blakemore [25] or the Boltzmann distribution in the low density limit. For organic semiconductors, it is given by the Gauss–Fermi integral [10], which reduces to a Boltzmann type distribution function (up to some normalizing factor) in the low-density limit or a Blakemore distribution function for vanishing variance σ. The latter corresponds to a δ-shaped density of states [10,26], describing a single electronic transport level. See Figure 50.8 for definitions and a relationship between these functions and Figure 50.9 for a graphical representation.

For inorganic, 3D bulk semiconductors assuming parabolic bands with effective mass m_e^* the effective density of states is given by

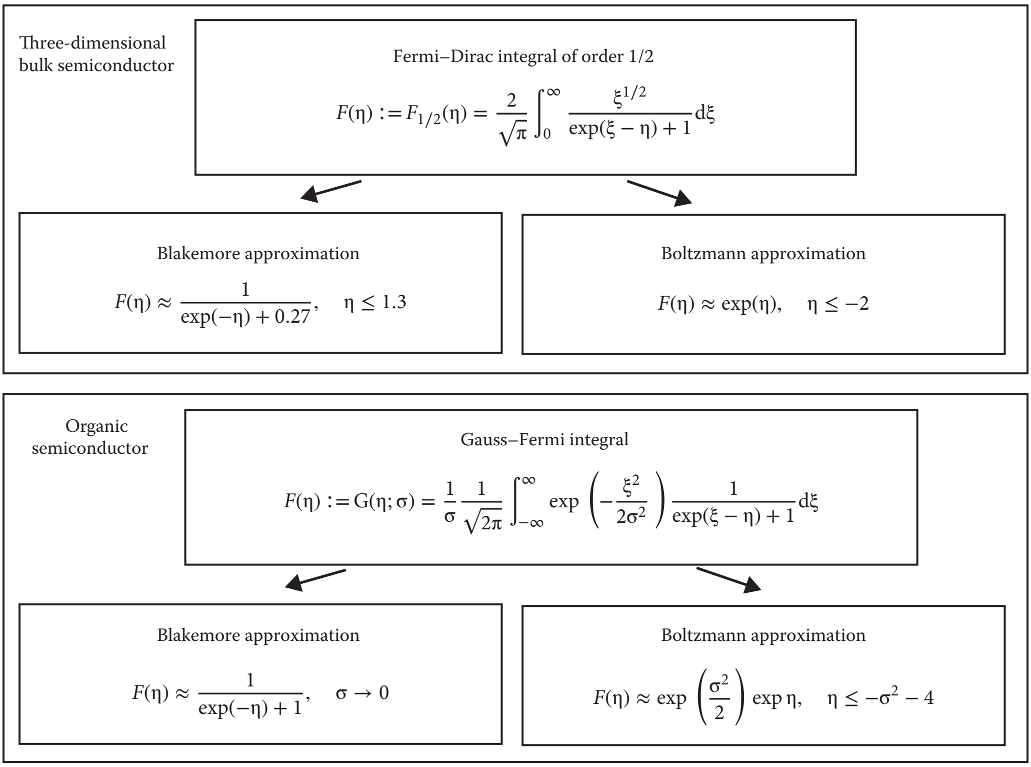
$$N_c = 2\left(\frac{m_e^* k_\mathrm{B}T}{2\pi\hbar^2}\right)^{3/2},$$

where T is the temperature and $\hbar$ the Planck constant. For organic semiconductors, the effective density of states N_c is given by the density of transport states N_t.

Introducing a general relationship between carrier density as well as quasi-Fermi and electrostatic potential via Equation 50.34 has quite a few implications. The nonlinear Poisson equation (50.41) becomes more complicated. In Section 50.2.4, we have studied how to design useful starting guesses for the Boltzmann approximation. For general distribution functions $\mathcal{F}$, it is no longer possible to obtain an explicit expression for ψ_0, like we have achieved in Equation 50.12. One rather needs to solve a nonlinear equation, for example, via Newton's methods at a given point $\mathbf{x}$. For highly doped regions, the Boltzmann approximation strongly overestimates the carrier densities if $\eta > 0$, see Figure 50.9. This implies that the more realistic Fermi–Dirac distribution leads to a higher built-in voltage. Moreover, from a numerical point of view, the Scharfetter–Gummel method needs to be adjusted. We discuss this in Section 50.4.2.

Finally, the distribution function $\mathcal{F}$ influences the ratio between diffusion and drift. To see this, we write the currents in *drift-diffusion* form. Using the relations for the carrier densities (50.34), we obtain

$$\mathbf{j}_n = -q\mu_n n\nabla\psi + qD_n\nabla n, \qquad \mathbf{j}_p = -q\mu_p p\nabla\psi - qD_p\nabla p.$$



Three-dimensional bulk semiconductor

Fermi–Dirac integral of order 1/2

$$F(\eta) := F_{1/2}(\eta) = \frac{2}{\sqrt{\pi}} \int_0^{\infty} \frac{\xi^{1/2}}{\exp(\xi - \eta) + 1} d\xi$$

Blakemore approximation

$$F(\eta) \approx \frac{1}{\exp(-\eta) + 0.27}, \quad \eta \leq 1.3$$

Boltzmann approximation

$$F(\eta) \approx \exp(\eta), \quad \eta \leq -2$$

Organic semiconductor

Gauss–Fermi integral

$$F(\eta) := \mathrm{G}(\eta; \sigma) = \frac{1}{\sigma} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(-\frac{\xi^2}{2\sigma^2}\right) \frac{1}{\exp(\xi - \eta) + 1} d\xi$$

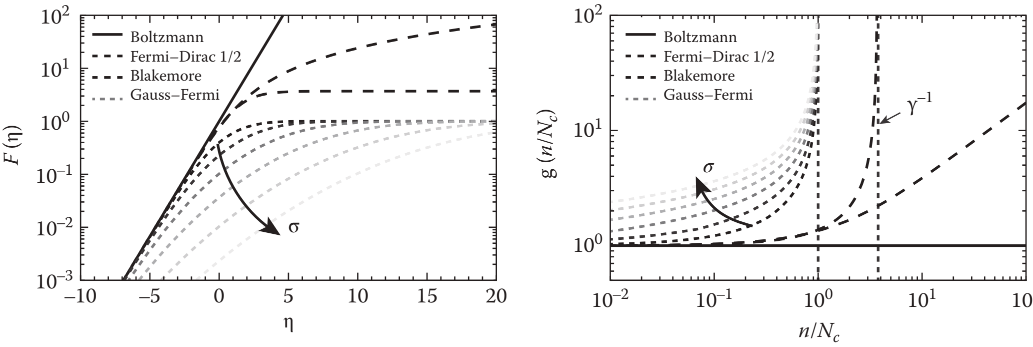
Blakemore approximation

$$F(\eta) \approx \frac{1}{\exp(-\eta) + 1}, \quad \sigma \to 0$$

Boltzmann approximation

$$F(\eta) \approx \exp\left(\frac{\sigma^2}{2}\right) \exp \eta, \quad \eta \leq -\sigma^2 - 4$$

FIGURE 50.8 This figure shows two important choices of the statistical distribution function: for three-dimensional bulk semiconductors and for organic semiconductors [10] described by a Gaussian density of states with variance σ. Both can be approximated using the Blakemore or Boltzmann distribution function.



(a)

(b)

FIGURE 50.9 (a) Some frequently considered distribution functions for the density of free carriers. For the Blakemore and the Gauss–Fermi functions one obtains a saturation of the carrier density at high values of η. (b) Illustration of the diffusion enhancement factors for the distribution functions shown in (a). In the case of the Blakemore and the Gauss–Fermi functions the saturation values of the carrier density directly lead to a divergent behavior of the nonlinear diffusion factor.

The diffusion coefficients D_n and D_p are connected to the carrier mobilities by a *generalized* Einstein relation

$$\frac{D_n}{\mu_n} = \frac{k_B T}{q} g\left(\frac{n}{N_c}\right), \qquad \frac{D_p}{\mu_p} = \frac{k_B T}{q} g\left(\frac{p}{N_v}\right) \tag{50.44}$$

with a density-dependent nonlinear factor

$$g(\xi) = \xi \left(\mathcal{F}^{-1}\right)'(\xi),$$

leading in general to a nonlinear diffusion coefficient. For the Boltzmann approximation, we immediately see that $g(\xi) = 1$, which gives the classical Einstein relation $D_n = k_B T\mu_n/q$. For non-exponential distribution function and $\xi \geq 1$, however, we note $g(\xi) \geq 1$, see [27]. For this reason, we call this factor *diffusion enhancement* as proposed in [9]. The dependence of the diffusion enhancement g on the density is depicted in Figure 50.9 for various distribution functions.

50.3.6 Free Energy and Dissipation Rate

We consider consistency with fundamental principles of thermodynamics to be a quality measure for models describing natural processes. So it is of significant interest to study the consistency for both continuous and discretized models. Due to our simplifying assumption of constant temperature, we cannot have energy conservation in the proper sense. The second law of thermodynamics for nonequilibrium processes [16,28] requires nonnegativity of the local entropy production which, multiplied by the temperature T, gives the *dissipation rate*.

For the sake of readability, we discuss these concepts for Boltzmann statistics [29] here. See [30] and [31] for more general statistics. For a triple (ψ, n, p) and a thermodynamic equilibrium solution $(\psi_{eq}, n_{eq}, p_{eq}) := (\psi_{eq}, n(\psi_{eq}), p(\psi_{eq}))$ of the full van Roosbroeck system the *free energy* is defined as

$$\begin{aligned}\mathbb{F}(\psi, n, p) = & \int_\Omega \left(n \log \frac{n}{n_{eq}} - n + p \log \frac{p}{p_{eq}} - p + n_{eq} + p_{eq} \right) \mathrm{d}\mathbf{x} \\ & + \int_\Omega \frac{\varepsilon_s}{2} \left| \nabla \left(\psi - \psi_{eq}\right)\right|^2 \mathrm{d}\mathbf{x} + \frac{1}{2} \int_{\Gamma_G} \frac{\varepsilon_{ox}}{d_{ox}} \left(\psi - \psi_{eq}\right)^2 \mathrm{d}s.\end{aligned} \tag{50.45}$$

For a transient solution $(\psi(t), n(t), p(t))$ of ((50.30)–(50.32)), the function $\mathbb{L}(t) = \mathbb{F}(\psi(t), n(t), p(t))$ decays exponentially as t tends to infinity and one has

$$\mathbb{L}(t) = \mathbb{L}(0) - \int_0^t \mathbb{D}(\tau)\mathrm{d}\tau,$$

where $\mathbb{D}(t)$ is the nonnegative dissipation rate [29]

$$\mathbb{D}(t) = \int_\Omega \Big(n\mu_n |\nabla\varphi_n|^2 + p\mu_p |\nabla\varphi_p|^2 + r(n,p)(np-1)\log(np) \Big) \mathrm{d}\mathbf{x} \geq 0 \tag{50.46}$$

depending on t via the time evolution of n and p. This result confirms the consistency with the second law of thermodynamics. Incidentally, the function $\mathbb{L}$ is a Lyapunov function, allowing in certain situations to prove the global stability of the thermodynamical equilibrium. Furthermore, it can be used as a tool to prove uniqueness of solutions to the system ((50.30)–(50.32)). This leads us naturally to the following subsection.

50.3.7 Existence and Uniqueness Results

When modeling complex physical phenomena, it is often necessary to simplify the underlying physical model. Therefore, a sound mathematical investigation is necessary to assess the implications of such simplifications. Take, for example, the electrostatic potential of a device. Physically, we know that such a potential exists. Hence, a sound model should guarantee its existence—also from a mathematical point of view. If the existence cannot be shown, then the model does not represent the physical world accurately. Apart from existence, uniqueness of the solution is often desirable as well as continuous dependence on the initial data.

The mathematical technique used to prove existence and uniqueness depends on the device geometry and the model. The first existence result on the van Roosbroeck system was shown by Mock [32]. Since then several results have been obtained by Gajewski and Gröger, we refer the interested reader to [33,34]. In [34], the key tool to show the existence and uniqueness of the time-dependent system is based on finding a Lyapunov function. Moreover, Gajewski and Gröger presented the first result which considered Fermi–Dirac statistics instead of Boltzmann statistics. There are other important existence results. We would like to mention explicitly the results studied by Markowich see, for example, [35], Jüngel [36], and Jerome [37].

We would like to show the reader that these results are not of purely analytical interest but can also be used to design numerical methods. Gummel's method [38], for example (see Section 50.5), is based on the same fixed point iteration technique used to obtain the existence result which we discuss next. The connection to numerics can be found in [37]. We would like to give the reader an idea regarding the proof of a standard result without going into the tricky details. We follow ideas from Markowich's textbook [39].

We consider the stationary van Roosbroeck system with Boltzmann statistics. Using a scaling that can be found in [39, Section 2.4] and making a change of variables, we can rewrite the steady-state van Roosbroeck system in the following way:

$$-\lambda^2 \Delta\psi = \delta^2 e^{-\psi} v - \delta^2 e^{\psi} u + C, \tag{50.47}$$

$$\nabla \cdot (\mu_n e^{\psi} \nabla u) = \tilde{R}, \tag{50.48}$$

$$\nabla \cdot (\mu_p e^{-\psi} \nabla v) = \tilde{R}. \tag{50.49}$$

The densities n and p are related to the so-called *Slotboom variables* u and v via $n = \delta^2 e^{\psi} u$ and $p = \delta^2 e^{-\psi} v$. In Equations 50.47 through 50.49, the parameter λ denotes the normalized characteristic Debye length of the device, δ^2 indicates the scaled intrinsic carrier density and $\tilde{R}$ is the scaled recombination term. It is of the form $\tilde{R} = c(\psi, u, v)(uv - 1)$ with $c(\psi, u, v) > 0$. The chosen scaling and change of variables make it possible to rewrite the boundary conditions. We consider only Ohmic contacts and homogeneous Neumann boundaries.

In order to prove the existence of solutions for the system ((50.47)–(50.49)) equipped with suitable boundary conditions, we need to make several technical assumptions. We refer the interested reader to [39] where these assumptions are well explained as well as physically and mathematically justified. Here, we would rather like to focus on the key idea of the proof which is exploited numerically, omitting mathematical detail. The existence proof is based on an iteration scheme which considers the Poisson equation (50.47) decoupled from the continuity equations (50.48) and (50.49). The proof consists of the following steps:

1. We fix some $u_0, v_0 > 0$, and consider the semilinear elliptic problem

$$-\lambda^2 \Delta\psi = \delta^2 e^{-\psi} v_0 - \delta^2 e^{\psi} u_0 + C, \text{ in } \Omega, \qquad \text{(plus BC for } \psi \text{ on } \partial\Omega). \tag{50.50}$$

   Using standard analytical results for semilinear elliptic equations (namely the Leray–Schauder fixed point theorem and the maximum principle [40]) we prove that there exists a unique solution of the problem (50.50) that we denote with ψ_1.

2. We insert the solution ψ_1 of Equation 50.50 into the decoupled linear elliptic equations

$$\nabla \cdot (\mu_n e^{\psi_1} \nabla u) = c(\psi_1, u_0, v_0)(uv_0 - 1), \text{ in } \Omega, \qquad \text{(plus BC for } u \text{ on } \partial\Omega), \tag{50.51}$$

$$\nabla \cdot (\mu_p e^{-\psi_1} \nabla v) = c(\psi_1, u_0, v_0)(u_0 v - 1), \text{ in } \Omega, \qquad \text{(plus BC for } v \text{ on } \partial\Omega). \tag{50.52}$$

   We point out that in the right-hand side of Equation 50.51 the rate c depends only on ψ_1 determined from the previous step and on the fixed u_0, v_0. The other term, $uv_0 - 1$, depends linearly on the unknown u. Analogous considerations are valid for Equation 50.52. Physically, this means that in the factor describing relaxation to equilibrium, $uv - 1$, we freeze the hole density, more precisely the Slotboom variable v_0, in the continuity equation for the electrons and vice versa. Thanks to the nonnegativity of $c(\psi_1, u, v)$ there exist unique solutions u_1 and v_1 to Equations 50.51 and 50.52. This is ensured by standard results for linear elliptic equations (see, e.g., [40]).

3. Based on the first two steps, we can now define a map H which maps (u_0, v_0) onto (u_1, v_1). This map is known as *Gummel map*. It is possible to prove that this map has a fixed point (u^*, v^*) which determines a (weak) solution (ψ^*, u^*, v^*) of the coupled system ((50.47)–(50.49)). The electrostatic potential ψ^* that solves Poisson's equation (50.47) can be determined using the first step by substituting (u_0, v_0) with (u^*, v^*).

Finally, we discuss the uniqueness of solutions to the van Roosbroeck system. It is well known that for the steady-state system, the uniqueness of the solution cannot be shown without additional assumptions, for example, on the applied voltage. This is no surprise, as in fact, some semiconductor devices (e.g., thyristors [41]) are designed to have multiple steady states. However, for a device in thermal equilibrium (see Section 50.3.4), we have $u = v = 1$. In this case, there is a unique function which satisfies the Poisson equation (50.47). Hence, the solution $(\psi_e, 1, 1)$ is the unique equilibrium solution of the system ((50.47)–(50.49)), see [32]. For sufficiently small bias voltages, the uniqueness of the solution is shown in [39] under some specific assumptions on the recombination rates. All recombination mechanisms appearing in Figure 50.3.3 fulfill these assumptions. However, impact ionization rates, for example, are excluded.

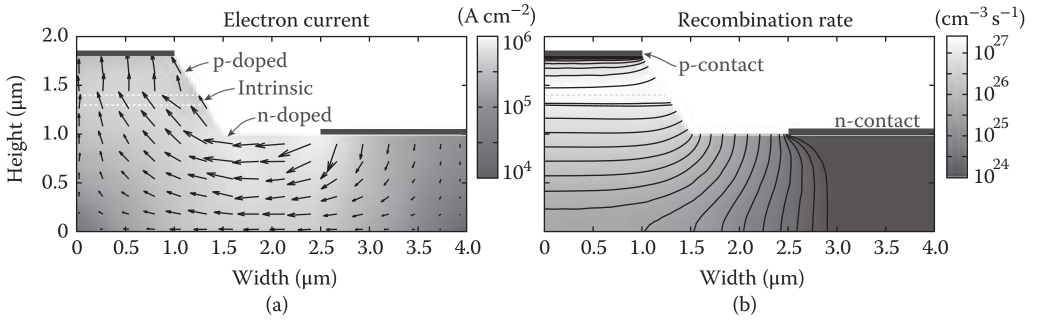
50.3.8 Maximum Principle

An important mathematical tool, often used for proofs similar to the ones in the previous section, is the maximum principle. Intuitively, the maximum principle states that if the domain is bounded and the right-hand side of the elliptic equation is positive, then the maximum of the solution is attained on the boundary of the domain [40].

50.4 Discretizing the van Roosbroeck System

In this section, we introduce a method for discretizing the van Roosbroeck system which is close to the physicist's approach to derive PDEs based on a subdivision of the computational domain into *representative elementary volumes* or *control volumes*. The *two-point flux* finite volume method described here can be seen as a straightforward generalization, preserving the properties of the 1D Scharfetter–Gummel scheme in higher dimensions. Figure 50.10 shows a 2D simulation. The 2D variant of this approach was introduced as *box method* in [12]. Historically, it goes back to [42]. The 3D variant of this method was probably first investigated in [29,43].

The method has two main ingredients: a geometry-based approach to obtain a system describing communicating control volumes and a consistent description of the fluxes between two adjacent control volumes. These are discussed in the following two subsections. We finish this section by describing various properties of the finite volume scheme.



(a) (b)

FIGURE 50.10 Exemplary results of a 2D simulation. The example is inspired by a ridge-waveguide laser with Ohmic contacts on the top of a mesa and at the side. For simplicity, we restrict to a homogeneous material which is here GaAs with the parameters as given in Figure 50.2. Within the mesa, an intrinsic domain is enclosed by a p-doped top-layer ($N_A = 2 \times 10^{18}$ cm^{-3}) and the n-doped substrate ($N_D = 2 \times 10^{18}$ cm^{-3}). The boundaries of the doped domains are shown in the pictures as dashed lines. (a) At an applied voltage of 1.5 V (flat band conditions) a significant electron current flow can be observed. The absolute value of the current density is shown by the gray scale, the arrows indicate the direction of particle flow (in the case of electrons the current density vector points in the opposite direction). (b) Plot of the total recombination rate at the same bias. The maximal recombination rate is observed in the vicinity of the intrinsic domain.

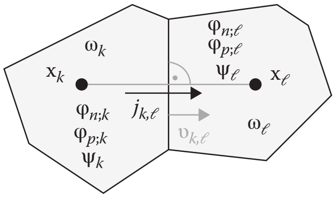


FIGURE 50.11 Two adjacent control volumes ω_k and ω_ℓ with corresponding data.

50.4.1 Finite Volume Method in Higher Dimensions

We partition the domain into N nonintersecting, convex polyhedral control volumes ω_k such that $\Omega = \bigcup_{k=1}^{N} \omega_k$. We associate with each control volume a node $\mathbf{x}_k \in \omega_k$. If the control volume intersects with the boundary of our domain, we demand that the node lies on the boundary, that means $\mathbf{x}_k \in \partial\Omega \cap \omega_k$. We assume that the partition is *admissible* [44], i.e., the *edge* $\overline{\mathbf{x}_k\mathbf{x}_\ell}$ with length $h_{k,\ell}$ is orthogonal to $\partial\omega_k \cap \partial\omega_\ell$. Thus, the normal vectors to $\partial\omega_k$ can be calculated by $\boldsymbol{\nu}_{k,\ell} = (\mathbf{x}_\ell - \mathbf{x}_k)/\|\mathbf{x}_\ell - \mathbf{x}_k\|$. See Figure 50.11 for details.

The set of all nodes and control volumes is called the *mesh* of the domain Ω. For a discussion of constructive ways to obtain such mesh partitions, see Section 50.6. In order to keep the equations simple, we introduce two abbreviations,

$$\eta_n\left(\psi, \varphi_n\right) = \frac{q\left(\psi - \varphi_n\right) - E_c}{k_B T} \quad \text{and} \quad \eta_p\left(\psi, \varphi_p\right) = \frac{E_v - q\left(\psi - \varphi_n\right)}{k_B T},$$

and assume that we have only a single gate contact with gate voltage U_G. We integrate the van Roosbroeck system ((50.30)–(50.32)) over ω_k and apply Gauss's divergence theorem, resulting in the integral equations

$$-\int_{\partial\omega_k} \varepsilon_s \nabla\psi \cdot \boldsymbol{\nu} \mathrm{d}s + \int_{\partial\omega_k \cap \Gamma_G} \frac{\varepsilon_{ox}}{d_{ox}}(\psi - U_G)\mathrm{d}s = q \int_{\omega_k} \left(C - N_c\mathcal{F}\left(\eta_n\left(\psi, \varphi_n\right)\right) + N_v\mathcal{F}\left(\eta_p\left(\psi, \varphi_p\right)\right)\right)\mathrm{d}\mathbf{x},$$

$$-q\frac{\partial}{\partial t}\int\limits_{\omega_k} N_c\mathcal{F}\left(\eta_n\left(\psi,\varphi_n\right)\right)\mathrm{d}\mathbf{x} + \int\limits_{\partial\omega_k} \mathbf{j}_n\cdot\boldsymbol{\nu}\mathrm{d}s = q\int\limits_{\omega_k} R d\mathbf{x},$$

$$q\frac{\partial}{\partial t}\int\limits_{\omega_k} N_v\mathcal{F}\left(\eta_p\left(\psi,\varphi_p\right)\right)\mathrm{d}\mathbf{x} + \int\limits_{\partial\omega_k} \mathbf{j}_p\cdot\boldsymbol{\nu}\mathrm{d}s = -q\int\limits_{\omega_k} R\mathrm{d}\mathbf{x},$$

for $k = 1, \ldots, N$. Here, $\boldsymbol{\nu}$ is outward-pointing unit normal to the control volume ω_k; see Figure 50.11. These equations represent an integral form of the van Roosbroeck system on every control volume. In particular, the first equation is Gauss's law of electrodynamics. The other two equations are balance laws for the carrier densities. The densities in each control volume change only due to in- and outflow through the boundary or recombination.

Next, the surface integrals are split into the sum of integrals over the planar interfaces between the control volume ω_k and its neighbors. Employing one point quadrature rules for the surface and volume integrals, we deduce the finite volume scheme

$$\begin{aligned}\sum_{\omega_\ell\in\mathcal{N}(\omega_k)} |\partial\omega_k\cap\partial\omega_\ell| D_{k,\ell} &= q|\omega_k|\left(C_k - N_c\mathcal{F}\left(\eta_n\left(\psi_k,\varphi_{n;k}\right)\right) + N_v\mathcal{F}\left(\eta_p\left(\psi_k,\varphi_{p;k}\right)\right)\right)\\ &\quad + |\partial\omega_k\cap\Gamma_G|\frac{\varepsilon_{\text{ox}}}{d_{\text{ox}}}(U_G - \psi_k),\end{aligned} \tag{50.53}$$

$$-q|\omega_k|N_c\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}\left(\eta_n\left(\psi_k,\varphi_{n;k}\right)\right) + \sum_{\omega_\ell\in\mathcal{N}(\omega_k)} |\partial\omega_k\cap\partial\omega_\ell| j_{n;k,\ell} = q|\omega_k|R_k, \tag{50.54}$$

$$q|\omega_k|N_v\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{F}\left(\eta_p\left(\psi_k,\varphi_{p;k}\right)\right) + \sum_{\omega_\ell\in\mathcal{N}(\omega_k)} |\partial\omega_k\cap\partial\omega_\ell| j_{p;k,\ell} = -q|\omega_k|R_k. \tag{50.55}$$

In the above mentioned formulas, $\mathcal{N}\left(\omega_k\right)$ denotes the set of all control volumes neigboring ω_k. In 2D, the measure $|\partial\omega_k\cap\partial\omega_\ell|$ corresponds to the length of the boundary line segment and in 3D to the area of the intersection of the boundary surfaces. The measure $|\omega_k|$ is in 2D given by the area and in 3D by the volume of the control volume ω_k. The unknowns ψ_k, $\varphi_{n;k}$, and $\varphi_{p;k}$ are approximations of the electric potential as well as the quasi-Fermi potentials for electrons and holes evaluated at node $\mathbf{x}_k$ like already introduced for the 1D case in Section 50.2. Accordingly, R_k is defined as

$$R_k = R\left(N_c\mathcal{F}\left(\eta_n\left(\psi_k,\varphi_{n;k}\right)\right), N_v\mathcal{F}\left(\eta_p\left(\psi_k,\varphi_{p;k}\right)\right)\right).$$

The doping is defined by the integral average

$$C_k = \frac{1}{|\omega_k|}\int_{\omega_k} C(\mathbf{x})\mathrm{d}\mathbf{x},$$

which can be estimated by its nodal value $C(\mathbf{x}_k)$. The numerical fluxes $D_{k,\ell}$, $j_{n;k,\ell}$ and $j_{p;k,\ell}$ approximate $-\varepsilon\nabla\psi\cdot\boldsymbol{\nu}_{k\ell}$, $\mathbf{j}_n\cdot\boldsymbol{\nu}_{k\ell}$ and $\mathbf{j}_p\cdot\boldsymbol{\nu}_{k\ell}$, respectively, on the interfaces between two adjacent control volumes ω_k and ω_ℓ, see Figure 50.11. As in the 1D case, we assume that these fluxes can be expressed as nonlinear functions depending on the values $\psi_k, \varphi_{n;k}, \varphi_{p;k}$ and $\psi_\ell, \varphi_{n;\ell}, \varphi_{p;\ell}$.

The electric displacement flux is approximated by

$$D_{k,\ell} = -\varepsilon_s\frac{\psi_\ell - \psi_k}{h_{k,\ell}},$$

where

$$h_{k,\ell} = \|\mathbf{x}_\ell - \mathbf{x}_k\|$$

is the edge length, compare with Equation 50.15.

For the Boltzmann approximation, the numerical charge carrier fluxes can be approximated via the Scharfetter–Gummel expression (50.21). In theory, any classical technique such as central differences or upwinding could be used to discretize the numerical fluxes. However, the former suffers from instability issues (as seen in Section 50.2.7) and neither scheme is consistent with the thermodynamic equilibrium. We will explain what this means in Section 50.4.3. In the next section, we discuss thermodynamically consistent schemes for non-Boltzmann statistics.

The finite volume scheme ((50.53)–(50.55)) yields a nonlinear system of $3N$ equations depending on $3N$ variables. We can directly substitute Dirichlet boundary values in the equations. In practice, however, this way of handling Dirichlet values is very technical to implement. Exploiting floating point arithmetic, the Dirichlet penalty method [45] provides a reasonable alternative. Physically, it replaces the Dirichlet boundary conditions by gate boundary conditions with very high oxide permittivity.

50.4.2 Scharfetter–Gummel Fluxes and Their Non-Boltzmann Generalizations

In 1D, we have already discussed the highly effective flux discretization scheme proposed by Scharfetter and Gummel for Boltzmann statistics [11]. They derive the numerical flux by locally solving a linear two-point boundary value problem. The 1D idea immediately carries over to higher dimensions if we insert the Scharfetter–Gummel (50.21) into the discrete system ((50.53)–(50.55)). However, if the Boltzmann approximation is not valid anymore, we can no longer derive the flux like Scharfetter and Gummel suggested.

This motivated the work in [46] where the Scharfetter–Gummel idea was generalized to a large class of nonlinear convection-diffusion problems, allowing to define consistent numerical fluxes from nonlinear two-point boundary value problems. Unfortunately, these *generalized Scharfetter–Gummel schemes* cannot be expressed by closed formulas. Sometimes, however, the local fluxes can be obtained iteratively [27]. It is also possible to approximate the two-point boundary value problems by simpler ones (e.g., by freezing some coefficients). This leads to *modified Scharfetter–Gummel schemes*. We address some of these schemes now.

For the sake of readability, we provide the formulas for electrons only. The formulas for holes follow similarly.

50.4.2.1 Generalized Scharfetter–Gummel Schemes

By the construction of our mesh, it suffices to study the 1D flux j_n along the edge $\overline{\mathbf{x}_k\mathbf{x}_\ell}$. For general distribution functions $\mathcal{F}$, we want to solve the ordinary differential equation

$$\frac{\mathrm{d}}{\mathrm{d}x} j_n = \frac{\mathrm{d}}{\mathrm{d}x}\Big(-q\mu_n N_c \mathcal{F}\left(\eta_n\left(\psi, \varphi_n\right)\right) \nabla\varphi_n \Big) = 0$$

on the interval $[0, h_{k,\ell}]$ with boundary conditions

$$\varphi_n(0) = \varphi_{n;k} \quad \text{and} \quad \varphi_n\left(h_{k,\ell}\right) = \varphi_{n;\ell},$$

where $\varphi_{n;k}$ and $\varphi_{n,\ell}$ are the values of the quasi-Fermi potentials at nodes $\mathbf{x}_k$ and $\mathbf{x}_\ell$, respectively, see Figure 50.11. We assume now that the flux is constant between both nodes and denote it with $j_{n;k,\ell}$. Integrating twice leads to an integral equation [27] for the unknown current, namely

$$\int_{\eta_{n;k}}^{\eta_{n;\ell}} \left(\frac{j_{n;k,l}/j_0}{\mathcal{F}(\eta)} + \frac{\psi_\ell - \psi_k}{U_\mathrm{T}} \right)^{-1} d\eta = 1, \tag{50.56}$$

where $j_0 = q\mu_n N_c \dfrac{U_\mathrm{T}}{h_{k,\ell}}$ and the limits are given by $\eta_{n;k} = \eta_n \left(\psi_k, \varphi_{n;k}\right)$ and $\eta_{n;\ell} = \eta_n \left(\psi_\ell, \varphi_{n;\ell}\right)$. For strictly monotonously increasing $\mathcal{F}(\eta)$ this equation has always a unique solution [47].

For the Boltzmann approximation $\mathcal{F}(\eta) = \exp(\eta)$, this integral equation can be solved analytically and yields the classical Scharfetter–Gummel expression for the flux (50.21). In [27], it was shown that for the Blakemore distribution function $\mathcal{F}(\eta) = \frac{1}{\exp(-\eta)+\gamma}$, the integral equation yields a fixed point equation

$$\widehat{j}_{n;k,\ell} = B\left(\gamma \widehat{j}_{n;k,\ell} + \frac{\psi_\ell - \psi_k}{U_\mathrm{T}} \right) e^{\eta_{n;\ell}} - B\left(- \left[\gamma \widehat{j}_{n;k,\ell} + \frac{\psi_\ell - \psi_k}{U_\mathrm{T}} \right] \right) e^{\eta_{n;k}} \tag{50.57}$$

for the nondimensionalized edge current $\widehat{j}_{n;k,\ell} = j_{n;k,l}/j_0$. The right-hand side is a Scharfetter–Gummel expression where the argument of the Bernoulli function is shifted by $\gamma \widehat{j}_{n;k,\ell}$. Hence, this reduces to the classical Scharfetter–Gummel scheme for $\gamma = 0$.

Since the Bernoulli function is strictly decreasing, this fixed point equation possesses a unique solution $\widehat{j}_{n;k,\ell}$. If we want to use the flux given by (50.57) in the discrete system ((50.53)–(50.55)) , we need to solve for the fluxes twice (once for electrons and once for holes) on each discretization edge $\overline{\mathbf{x}_k \mathbf{x}_\ell}$. A few Newton steps are sufficient to solve this equation iteratively.

Even though (50.57) is restricted to the Blakemore approximation, it provides a useful scheme in the context of organic semiconductors. There it arises naturally as a model for materials with δ-shaped density of states [10,26], describing a single electronic transport level, see Figures 50.8 and 50.9. This has been described in Section 50.3.5.

Unfortunately, for a general statistical distribution function, no corresponding equation has been derived so far. Therefore, in [47] it was proposed to use piecewise approximations of $\mathcal{F}$ by Blakemore type or rational approximations of Padé type in order to obtain piecewise integrable expression from the local boundary value problem.

50.4.2.2 Modified Scharfetter–Gummel Schemes

Bessemoulin-Chatard [48] derived a finite volume scheme for convection-diffusion problems by averaging the nonlinear diffusion term appropriately. This idea was generalized to more general distribution functions in [49], introducing a logarithmic average of the nonlinear diffusion enhancement

$$g_{n;k,\ell} = \frac{\eta_{n;k} - \eta_{n;\ell}}{\log \mathcal{F}\left(\eta_{n;k}\right) - \log \mathcal{F}\left(\eta_{n;\ell}\right)}$$

along the discretization edge. Using the generalized Einstein relation (50.44), one immediately observes that the diffusion enhancement g can be seen as a modification factor of the thermal voltage U_T. Replacing U_T in the Scharfetter–Gummel expression (50.21) by $U_\mathrm{T}^* = U_\mathrm{T} g_{n;k,\ell}$, we deduce the following modified Scharfetter–Gummel scheme

$$j_{n;k,\ell} = -\frac{q\mu_n U_\mathrm{T}}{h_{k,\ell}} g_{n;k,\ell} \left(N_c \mathcal{F}\left(\eta_{n;k}\right) B\left(-\frac{\psi_\ell - \psi_k}{U_\mathrm{T} g_{n;k,\ell}} \right) - N_c \mathcal{F}\left(\eta_{n;\ell}\right) B\left(\frac{\psi_\ell - \psi_k}{U_\mathrm{T} g_{n;k,\ell}} \right) \right), \tag{50.58}$$

approximating the current along the edge.

Previously, we have replaced the thermal voltage by a suitable average along the edge. Now, we want to approximate $\mathcal{F}(\eta)$ along the edge by an exponential (Boltzmann approximation) and adjusting the effective density of states N_c accordingly, i.e.,

$$N_c\mathcal{F}(\eta) \approx N^*_{c;k,\ell} \exp(\eta).$$

This choice makes it possible to use the original Scharfetter–Gummel flux (50.21) by replacing N_c with $N^*_{c;k,\ell}$. One choice for the modified density of states is

$$N^*_{c;k,\ell}(\eta^*) = N_c \frac{\mathcal{F}(\eta^*)}{\exp(\eta^*)},$$

where $\eta^* \in [\eta_k, \eta_\ell]$, assuming $\eta_k \leq \eta_\ell$. In practice, we might consider taking the geometric average between $N_c(\eta_k)$ and $N_c(\eta_\ell)$, which leads to another modified Scharfetter–Gummel scheme:

$$j_{n;k,\ell} = -\frac{q\mu_n U_{\mathrm{T}}}{h_{k,\ell}} N_c \sqrt{\frac{\mathcal{F}\left(\eta_{n;k}\right)\mathcal{F}\left(\eta_{n;\ell}\right)}{\exp(\eta_{n;k})\exp(\eta_{n;\ell})}} \left(B\left(-\frac{\psi_\ell - \psi_k}{U_{\mathrm{T}}}\right) e^{\eta_{n;k}} - B\left(\frac{\psi_\ell - \psi_k}{U_{\mathrm{T}}}\right) e^{\eta_{n;\ell}} \right). \tag{50.59}$$

The idea behind this scheme was introduced in [50] for the numerical solution of the generalized Nernst–Planck system which is similar to the van Roosbroeck system ((50.30)–(50.33)). A variant of this scheme for Fermi–Dirac statistics is described in [14,51] and numerically implemented in the semiconductor device simulation package WIAS-TeSCA [52].

50.4.3 Flux Expressions Consistent with the Thermodynamic Equilibrium

As introduced in Section 50.3.4, the currents of holes and electrons vanish in the thermodynamic equilibrium when no bias is applied. If a numerical scheme for zero bias boundary conditions results into vanishing numerical fluxes, we call it *consistent with the thermodynamical equilibrium*. In practice, one can examine this consistency by checking whether $j_{n;k,\ell} = j_{p;k,\ell} = 0$ if $\varphi_{n;k} = \varphi_{n;\ell}$ and $\varphi_{p;k} = \varphi_{p;\ell}$. Violating this property causes unphysical dissipation (spurious Joule heating) in the steady state attained for zero bias boundary conditions, which by definition is supposed to be the thermodynamic equilibrium [48].

All schemes introduced in Section 50.4.2 are consistent with the thermodynamical equilibrium. Indeed, let us assume that the quasi-Fermi potentials φ_n and φ_p between two adjacent control volumes ω_k and ω_ℓ are equal. The consistency of the classical Scharfetter–Gummel (50.21) and the scheme using a modified density of states (50.59) are obvious since the Bernoulli function satisfies $B(-x) = \exp(x)\, B(x)$.

Due to the definition from the solution of the two-point boundary value problem, the generalized Scharfetter–Gummel scheme defined in Equation 50.56 and its specialization for the Blakemore approximation (50.57) are consistent with the thermodynamic equilibrium. And finally, the logarithmic average of the diffusion enhancement is the only possible average which guarantees consistency with thermodynamics in the scheme (50.58). For details see [27,47,49].

50.4.4 Free Energy and Dissipation Rate

For Boltzmann statistics, it is rather straightforward to define discrete analoga of the free energy (50.45) and the dissipation rate (50.46). The positivity of the discrete dissipation rate was shown in [29], and the exponential decay of the free energy to its equilibrium value was proven in [53]. An overview of the entropy method for finite volume schemes has been given in [54]. First results on more general statistics functions in this respect have been obtained in [55].

These pioneering works strongly indicate that the chosen discretization approach results in discrete models, which are consistent with the structural assumptions of nonequilibrium thermodynamics. A full account of these issues in the context of general statistics functions remains an open research topic.

50.4.5 Existence, Uniqueness, and Convergence

There are very few existence proofs for the solutions of the discretized system ((50.53)–(50.55)). For the Boltzmann approximation, Gärtner [56] proved that the discretized steady-state system has a solution, which becomes unique if a small bias is applied. A similar result for Fermi statistics has been obtained in [47].

A convergence theory for the finite volume scheme for the full discrete system ((50.53)–(50.55)) and general flux functions is still missing. However, practical experience and a number of results make its convergence plausible.

For example, in one space dimension, second-order convergence in the discrete maximum norm for the Scharfetter–Gummel scheme has been shown in [19]. Under the assumption that second derivatives of the continuous solution exist, in [57] for moderately sized drift terms and two-dimensional (2D), square grids, first-order convergence for the simple upwind scheme (see e.g., [21]), and second-order convergence for the exponential fitting scheme in the L_2-norm has been shown. Reinterpretations of the finite volume Scharfetter–Gummel scheme as a nonstandard finite element method allowed to obtain convergence estimates for Scharfetter–Gummel schemes on Delaunay grids (see Section 50.6) [58,59]. For a general approach to the convergence theory of finite volume schemes, see [44]. In [46], weak convergence (no order estimate) for a generalization of the Scharfetter–Gummel scheme to nonlinear convection-diffusion problems has been shown. A convergence proof for a variant of the van Roosbroeck system discretized with the simple upwind scheme was given in [60].

50.4.6 Maximum Principle

When applied to the drift-diffusion formulation, compared to various variants of the stabilized finite element method, the two-point flux finite volume scheme is outstanding in the sense that it guarantees positivity of densities and absense of unphysical oscillations [61,62]. It allows to carry over the discussions of Section 50.2.7 to the higher dimensional case.

50.5 Nonlinear Solvers

In Section 50.2, we have already briefly mentioned that we need to solve a nonlinear discrete system of the form

$$\mathbf{0} = \mathbf{F}(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) := \begin{pmatrix} \mathbf{F}_1(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) \\ \mathbf{F}_2(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) \\ \mathbf{F}_3(\boldsymbol{\psi}, \boldsymbol{\varphi}_n, \boldsymbol{\varphi}_p) \end{pmatrix}.$$

Here, we discuss two ways to do this: Newton's method and *Gummel's iteration method* [38], which is based on a blockwise decoupling of the discrete system.

50.5.1 Newton's Method

Assuming that a good starting guess has been provided, for example, by following the ideas in Section 50.2.6, Newton's method constructs a new $(k+1)$th iterate from the kth one by solving the linear system

$$\begin{pmatrix} \frac{\partial \mathbf{F}_1}{\partial \boldsymbol{\psi}} & \frac{\partial \mathbf{F}_1}{\partial \boldsymbol{\varphi}_n} & \frac{\partial \mathbf{F}_1}{\partial \boldsymbol{\varphi}_p} \\ \frac{\partial \mathbf{F}_2}{\partial \boldsymbol{\psi}} & \frac{\partial \mathbf{F}_2}{\partial \boldsymbol{\varphi}_n} & \frac{\partial \mathbf{F}_2}{\partial \boldsymbol{\varphi}_p} \\ \frac{\partial \mathbf{F}_3}{\partial \boldsymbol{\psi}} & \frac{\partial \mathbf{F}_3}{\partial \boldsymbol{\varphi}_n} & \frac{\partial \mathbf{F}_3}{\partial \boldsymbol{\varphi}_p} \end{pmatrix}^k \begin{pmatrix} \boldsymbol{\delta\psi}^k \\ \boldsymbol{\delta\varphi}_n^k \\ \boldsymbol{\delta\varphi}_p^k \end{pmatrix} = - \begin{pmatrix} \mathbf{F}_1(\boldsymbol{\psi}^k, \boldsymbol{\varphi}_n^k, \boldsymbol{\varphi}_p^k) \\ \mathbf{F}_2(\boldsymbol{\psi}^k, \boldsymbol{\varphi}_n^k, \boldsymbol{\varphi}_p^k) \\ \mathbf{F}_3(\boldsymbol{\psi}^k, \boldsymbol{\varphi}_n^k, \boldsymbol{\varphi}_p^k) \end{pmatrix} \tag{50.60}$$

for the update vector

$$\begin{pmatrix} \boldsymbol{\delta\psi}^k \\ \boldsymbol{\delta\varphi}_n^k \\ \boldsymbol{\delta\varphi}_p^k \end{pmatrix} := \begin{pmatrix} \boldsymbol{\psi}^{k+1} - \boldsymbol{\psi}^k \\ \boldsymbol{\varphi}_n^{k+1} - \boldsymbol{\varphi}_n^k \\ \boldsymbol{\varphi}_p^{k+1} - \boldsymbol{\varphi}_p^k \end{pmatrix}.$$

Then the new iterate is obtained by adding the previous vector to the update. The advantage of Newton's method is that it converges quadratically if the starting guess is sufficiently close to the solution [17]. This allows to obtain highly accurate discrete solutions at low additional cost. The major drawback is that it might converge very slowly or even fail to converge if the starting guess is too far from the actual solution. Damping—multiplying the update with a factor less than 1—is known to increase the convergence region. Another remedy is parameter embedding where one slowly changes a parameter, always using the old solution as a new starting guess. We have already employed this embedding technique in Section 50.2.6 when gradually increasing the bias voltage in the solution procedure.

An important advantage of finite element and finite volume discretizations is the fact that they create only next-neighbor couplings in the discretized systems. The resulting linearized systems are therefore sparse, i.e., the maximum number of nonzero elements in a matrix row is bounded by a constant independent of the number of discretization cells, making it possible to use highly economic storage schemes. For 2D applications, one can usually solve the resulting sparse linear system (50.60) via sparse direct solvers, such as PARDISO [63–66] or UMFPACK [67,68]. Direct solvers calculate a representation of the system matrix as a product of easily solvable lower and upper triangular matrices, which unfortunately are not anymore sparse, increasing memory consumption and computational time especially in large 3D applications. This fill-in phenomenon is avoided by preconditioned Krylov subspace methods [69], which compared to the direct solvers need significant effort in tuning and adaptation to the problem at hand.

50.5.2 Gummel's Method

Gummel [13,38] suggested decoupling the three equations in the van Roosbroeck system. He devised an iterative method at the continuous level, which for Boltzmann statistics and drift-diffusion form leads to alternating between solving linear differential equations for the electric potential as well as the charge carrier densities. Suppose, one already knows the iterate (ψ^k, n^k, p^k). In order to obtain the electric potential at the new level, ψ^{k+1}, one solves

$$0 = -\frac{1}{q}\nabla \cdot \left(\varepsilon_s \nabla \psi^{k+1}\right) + \left(n^k + p^k\right)\left(\frac{\psi^{k+1} - \psi^k}{U_\mathrm{T}}\right) + n^k - p^k - C.$$

This formulation is motivated by linearizing a fixed point problem that one obtains from the Poisson equation [13]. Once one has solved for the electrostatic potential at the new level, one can successively solve the continuity equations from

$$0 = \nabla \cdot \mathbf{j}_n(\psi^{k+1}, n^{k+1}) - qR(\psi^{k+1}, n^{k+1}, p^k)$$
$$0 = \nabla \cdot \mathbf{j}_p(\psi^{k+1}, p^{k+1}) + qR(\psi^{k+1}, n^{k+1}, p^{k+1}).$$

Gummel's method is known to have a larger convergence region and a slower asymptotic convergence rate compared to Newtons's method. Depending on the software environment chosen, it also may be easier to implement. Note that standard existence and uniqueness proofs for the van Roosbroeck system rely on a similar decoupling strategy, see Section 50.3.7.

50.6 Mesh Generation for the Finite Volume Method

While appearing to be just a technical footnote compared to the questions of modeling, analysis and discretization, mesh generation in reality is a hard problem which deserves special attention. We focus on methods to construct admissible subdivisions of the computational domain Ω as used in Section 50.4.

50.6.1 Boundary Conforming Delaunay Meshes on Polyhedral Domains

Assume a partition (triangulation in 2D, tetrahedralization in 3D) of the polyhedral domain $\Omega = \bigcup_{k=1}^{N_\Sigma} \Sigma_k$ into nonoverlapping simplices Σ_k as it is commonly used for finite element methods [70]. We require that this simplicial partition of Ω has the *boundary conforming Delaunay property* [71]. For a triangulation of a 2D domain this property is equivalent to:

(1) For any two triangles with a common edge, the sum of their respective angles opposite to that edge is less or equal to 180°.
(2) For any triangle sharing an edge with $\partial\Omega$, its angle opposite to that edge is less or equal to 90°.

For a given vertex $x_i \in X$, where X is the set of vertices, the *Voronoi cell* around X is the set

$$V_i = \left\{ \mathbf{x} \in \mathbb{R}^d \; : \; \|\mathbf{x} - \mathbf{x}_i\| < \|\mathbf{x} - \mathbf{x}_j\| \text{ for all } \mathbf{x}_j \in X \text{ with } \mathbf{x}_j \neq \mathbf{x}_i \right\}.$$

The *Voronoi diagram*—the set of Voronoi cells for all vertices in X—is dual to the *Delaunay triangulation* of the point set X in the sense that for each edge $\overline{\mathbf{x}_i\mathbf{x}_j}$ in the Delaunay triangulation, $\partial V_i \cap \partial V_j \neq \emptyset$. If the simplicial partition is boundary conforming Delaunay, the restricted Voronoi cells $\omega_i = \Omega \cap V_i$ are well defined and can be obtained by joining the circumcenters of the simplices adjacent to the vertex $\mathbf{x}_i$. These restricted Voronoi cells provide an admissible control volume partition as required in Section 50.4. We note that in order to implement the finite volume method as described in Section 50.4, there is no need for an explicit construction of the control volumina ω_i as geometrical objects. Given the simplical partition, it is sufficient to base the calculations on the simplicial contributions $s_{ij}^k = |\partial\omega_i \cap \partial\omega_j \cap \Sigma_k|$ and $|\omega_i \cap \Sigma_K|$, and to use a simplex-based assembly loop as common for finite elements [29]. Figure 50.12 shows (boundary conforming) Delaunay triangulations.

There are several efficient algorithms to construct Delaunay triangulations for a given point set X [73,74]. These are good starting points of devising meshing algorithms in general. Many different problems complicate the generation of a mesh. For example, the boundary conforming Delaunay property is rather difficult to achieve, in particular it requires the careful insertion of additional points on the boundary. In 3D, slivers (very flat tetrahedra) must be avoided. And finally, it may be very complicated to fulfill additional requirements like constraints on the minimum angle or the local element size. Though there are still unsolved problems, the triangle ([72], 2D, free for noncommercial use) and TetGen ([75], 3D, open source) mesh generators help to create boundary conforming Delaunay meshes based on algorithms which are proven to

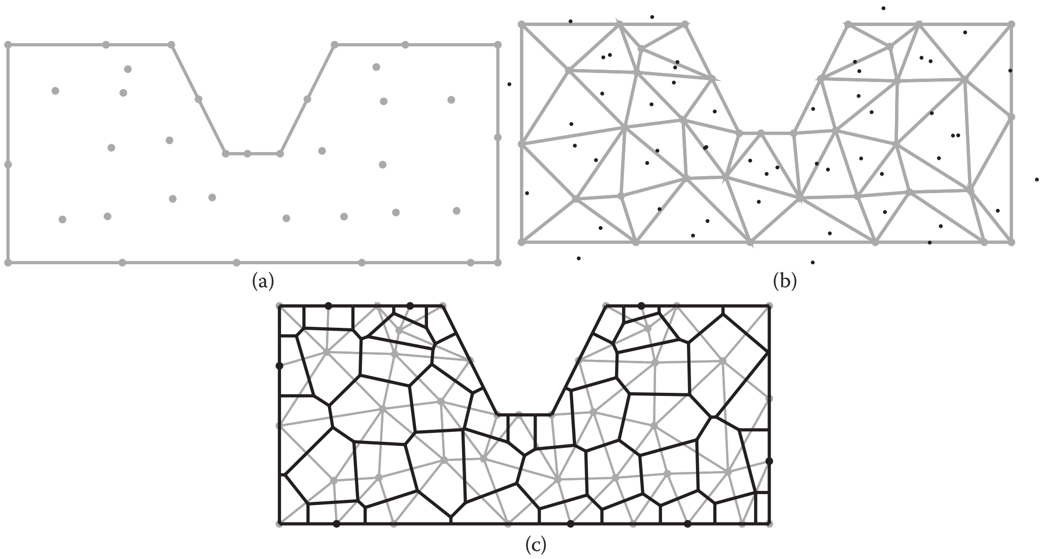


FIGURE 50.12 (a) Piecewise linear description of computational domain with given point cloud. (b) Delaunay triangulation of domain and triangle circumcenters. As some boundary triangles have angles larger than 90° angles opposite to the boundary, their circumcenters (small dots) are outside of the domain. (c) Boundary conforming Delaunay triangulation with automatically inserted additional points at the boundary (large dots), and Voronoi and restricted Voronoi cells (black). Created using triangle [72].

deliver meshes with the desired properties in finite time for a broad class of geometries given by a piecewise linear description of the boundary.

Other mesh generation approaches, in particular the advancing front [76] and the octree method [77], are similarly widespread. However, while popular in finite element community, their design makes it much harder to use them as a starting point for generating boundary conforming Delaunay meshes.

50.6.2 Tensor Product Approaches

In simplified geometrical situations, rectangular and cuboid mesh structures are a straightforward way to generate admissible finite volume partitions of the computational domain [12]. Extrusion of a 2D boundary conforming Delaunay base mesh into a 3D mesh consisting of prisms (and optional subsequent subdivision of these prisms into tetrahedra) provides another method to create admissible finite volume partitions [78].

50.6.3 Open Problems

Charge densities and potentials in semiconductors exhibit interior and boundary layers, i.e., thin regions charaterized by rapid gradients in one direction and slow gradients in the orthogonal direction. This may happen in space charge regions in the vicinity of gate contacts and at p-n junctions. These layers and other possible "hotspots" due to recombination, for example, call for local mesh adaptation. Element size control via the mesh generators using *a priori* knowledge is often used in this situation [72,75]. The development of reliable a posteriori mesh refinement criteria is an open problem in the context of the strongly coupled van Roosbroeck system. A physically motivated heuristic approach based on the equidistribution of the dissipation rate was suggested in [29] and investigated in [79].

Mesh generation tools mostly fail to create boundary conforming Delaunay meshes which use thin, anisotropic simplices to resolve interior and boundary layers with an optimal number of points. A partial remedy in some situations can be provided by node offsetting [80], a technique still under development. In certain cases, tensor product approaches [78] give reasonable results.

50.7 Time Discretization

In order to discretize the time derivative in the van Roosbroeck system ((50.30)–(50.32)), we define a series of discrete time values $0 = t^0, t^1, \ldots, t^M = T$ with step lengths $\tau^m = t^{m+1} - t^m$. We describe the temporal semi-discretization of the (spatially) continuous problem. For practical purposes it needs to be combined with the space discretization approach described in Section 50.4. The implict Euler time discretization scheme assumes given values $\psi^m, \varphi_n^m, \varphi_p^m, n^m, p^m$ of the variables at time t^m. Values at the subsequent time t^{m+1} are calculated by solving the nonlinear system of equations

$$-\nabla \cdot (\varepsilon_s \nabla \psi^{m+1}) = q(C - n^{m+1} + p^{m+1}), \tag{50.61}$$

$$-q\frac{n^{m+1} - n^m}{\tau^m} - q\nabla \cdot \mu_n n^{m+1} \nabla \varphi_n^{m+1} = qR(n^{m+1}, p^{m+1}), \tag{50.62}$$

$$q\frac{p^{m+1} - p^m}{\tau^m} - q\nabla \cdot \mu_p p^{m+1} \nabla \varphi_p^{m+1} = -qR(n^{m+1}, p^{m+1}), \tag{50.63}$$

where

$$n^{m+1} = N_c \mathcal{F}(\eta_n(\psi^{m+1}, \varphi_n^{m+1})),$$
$$p^{m+1} = N_v \mathcal{F}(\eta_p(\psi^{m+1}, \varphi_p^{m+1})).$$

As a consequence, the nonlinear system ((50.61)–(50.63)), which is structurally similar to Equations 50.53 through 50.55, has to be solved for each time step by iterative methods described in Section 50.5. An initial guess for these methods can be obtained from the old time step t^m or a linear extrapolation involving data from several previous time steps [81].

It is advisable to choose the size of the time steps τ^m according to multiple criteria, including the convergence radius of Newton's method, the free energy, and local changes of the potentials.

At first glance, the computational cost of having to solve a nonlinear system for each time step appears to be rather high. However, this scheme is stable independent of the time step size and allows to carry over all advantages of the finite volume scheme from the stationary to the transient case. The transient discrete solution converges to the stationary one, in particular, for the corresponding boundary condition to the thermodynamic equilibrium solution. The free energy of the system decays during this approach to the equilibrium. In several cases, it has been proven that this decay is exponential [55,82]. By preserving the physics, the implicit Euler scheme fits well to the thermodynamically motivated discretization schemes. The use of other schemes implies giving up some of these physical properties. If the resulting deviations from thermodynamic properties can be kept under control, higher order schemes (e.g., backward differentiation formula (BDF) methods [83]) might help to reduce computational costs due to a higher temporal convergence order.

Due to the perceived complexity and high computational costs of solving coupled nonlinear PDE systems, linear implicit methods, mixing linearization, and time discretizations are alternatives. Most prominent is the scheme devised in [84].

50.8 Contact Terminal Currents

In simulations, one is usually interested in IV-curves, i.e., the dependency of terminal currents on applied voltages. Therefore, calculating terminal currents accurately is crucial to a successful postprocessing of the simulated field data. The *total current density* is given by

$$\mathbf{j} = \mathbf{j}_d + \mathbf{j}_n + \mathbf{j}_p,$$

where $\mathbf{j}_d = -\varepsilon_s \nabla(\partial_t \psi) = \partial_t \mathbf{D}$ is the *displacement current density*. Taking the time derivative of the Poisson equation (50.30) yields

$$\begin{aligned} -\nabla \cdot \varepsilon_s \nabla(\partial_t \psi) &= q(\partial_t n - \partial_t p) = -\nabla \cdot \mathbf{j}_n - qR - \nabla \cdot \mathbf{j}_p + qR \\ \nabla \cdot \partial_t \mathbf{D} &= -\nabla \cdot \mathbf{j}_n - \nabla \cdot \mathbf{j}_p, \end{aligned}$$

resulting in $\nabla \cdot \mathbf{j} = 0$, which physically implies charge conservation.

Given a set of contacts $\Gamma_\alpha \subset \partial\Omega$ for $\alpha = 1, \ldots, N_C$, the terminal current through contact α is defined as

$$I_\alpha = \int_{\Gamma_\alpha} \mathbf{j} \cdot \boldsymbol{\nu} \, \mathrm{d}s.$$

The correct calculation of this integral presupposes a regularity of the solution and its derivative which is not supported by analytical theory. Therefore, in [51,52] a test function–based technique was proposed, which uses the weak formulation of the problem. For a physical motivation of this approach and an interpretation as a generalized Shockley–Ramo theorem we refer to [85]. For any α, let the function $T_\alpha : \Omega \to \mathbb{R}$ solve the boundary value problem

$$-\nabla^2 T_\alpha = 0 \qquad \text{in } \Omega, \tag{50.64}$$

$$\nabla T_\alpha \cdot \boldsymbol{\nu} = 0 \qquad \text{on } \partial\Omega \setminus \bigcup_{\beta=1}^{N_C} \Gamma_\beta, \tag{50.65}$$

$$T_\alpha = \delta_{\alpha\beta} \qquad \text{on } \Gamma_\beta \text{ (for } \beta = 1, \ldots, N_C), \tag{50.66}$$

where $\delta_{\alpha\beta}$ denotes the Kronecker delta. As a consequence,

$$\begin{aligned} I_\alpha = \int_{\Gamma_\alpha} \mathbf{j} \cdot \boldsymbol{\nu} \, \mathrm{d}s &= \int_{\partial\Omega} T_\alpha \mathbf{j} \cdot \boldsymbol{\nu} \, \mathrm{d}s = \int_\Omega \nabla \cdot (T_\alpha \mathbf{j}) \, \mathrm{d}\mathbf{x} \\ &= \int_\Omega \nabla T_\alpha \cdot \mathbf{j} \, \mathrm{d}\mathbf{x} + \int_\Omega T_\alpha \nabla \cdot \mathbf{j} \, \mathrm{d}\mathbf{x} \\ &= \int_\Omega \nabla T_\alpha \cdot (\mathbf{j}_n + \mathbf{j}_p) \, \mathrm{d}\mathbf{x} + \partial_t \int_\Omega \nabla T_\alpha \cdot \mathbf{D} \, \mathrm{d}\mathbf{x}. \end{aligned}$$

Given the semi-discretization in time from Section 50.7, one arrives at

$$I_\alpha^{m+1} = \int_\Omega \nabla T_\alpha \cdot (\mathbf{j}_n^{m+1} + \mathbf{j}_p^{m+1}) \, \mathrm{d}\mathbf{x} + \frac{1}{\tau^m} \int_\Omega \nabla T_\alpha \cdot \left(\mathbf{D}^{m+1} - \mathbf{D}^m\right) \, \mathrm{d}\mathbf{x}.$$

Like in Section 50.4, we approximate the system (50.64)–(50.66) by a finite volume discretization. Using partial integration and the notation introduced in Section 50.4 we come up with an expression for the total current through the α-th contact of the form

$$I_{\alpha}^{m+1} \approx \sum_{\substack{\omega_\ell \in \mathcal{N}(\omega_k) \\ k<\ell}} |\partial\omega_k \cap \partial\omega_\ell| \left(j_{n;k,\ell}^{m+1} + j_{p;k,\ell}^{m+1} + \frac{D_{k,\ell}^{m+1} - D_{k,\ell}^{m}}{\tau^m} \right) (T_{\alpha;k} - T_{\alpha;\ell}). \tag{50.67}$$

50.9 An Alternative: The Finite Element Method

We have put special emphasis on the finite volume method. However, it is worth pointing out that there are other approaches. The finite difference method that has been used to solve the van Roosbroeck system [13] on tensor product meshes is equivalent to the previously introduced finite volume method.

The most popular ansatz to deal with unstructured meshes is the finite element method [70]. It starts with a *weak formulation* of the steady-state van Roosbroeck problem ((50.30)–(50.32)). Each equation of the system is multiplied with a so-called test function and integrated in space. Integrating by parts, each equation of the steady-state system can be restated as the following problem:

$$\text{find } u \in U \text{ such that } a(u, v) = f(v) \text{ for all } v \in V,$$

where U is a function space containing the ansatz functions u and V is a function space containing the test functions v. The map $a: U \times V \to \mathbb{R}$ has the structure

$$a(u, v) = \int_{\Omega} \alpha(u) \nabla u \cdot \nabla v \, d\mathbf{x} - \int_{\partial\Omega} \nabla u \cdot \mathbf{n} \, v \, ds \tag{50.68}$$

and the functional $f: V \to \mathbb{R}$ is continuous. The idea of a finite element method is to approximate this continuous problem with a discrete version. We approximate the infinite dimensional spaces U and V with finite dimensional subspaces U_h and V_h spanned by basis functions with localized support. Then we obtain a finite dimensional version of Equation 50.68, namely

$$\text{find } u_h \in U_h \text{ such that } a_h(u_h, v_h) = f(v_h) \text{ for all } v_h \in V_h.$$

As it is sufficient to fulfill this condition for finitely many localized basis functions in V_h, we obtain a sparse (containing many zeros) nonlinear system of equations for the basis coefficients of u_h, assuming we are able to calculate the integrals. One usually approximates these integrals by quadrature rules.

Auf der Maur [86] suggested to write the stationary van Roosbroeck system in this framework, using the quasi-Fermi potential formulation of the flux. In this case one has

$$a^{\psi}(\psi, v) = \int_{\Omega} \varepsilon_s \nabla\psi \cdot \nabla v \, d\mathbf{x} + \int_{\Gamma_G} \frac{\varepsilon_{\text{ox}}}{d_{\text{ox}}} (\psi - U_{\text{G}}) v \, ds, \qquad f^{\psi}(v) = q \int_{\Omega} (p - n + C) v \, d\mathbf{x}, \tag{50.69}$$

$$a^{\varphi_n}(\varphi_n, v) = \int_{\Omega} \mu_n n \nabla\varphi_n \cdot \nabla v \, d\mathbf{x}, \qquad f^{\varphi_n}(v) = \int_{\Omega} R v \, d\mathbf{x}, \tag{50.70}$$

$$a^{\varphi_p}(\varphi_p, v) = \int_{\Omega} \mu_p p \nabla\varphi_p \cdot \nabla v \, d\mathbf{x}, \qquad f^{\varphi_p}(v) = -\int_{\Omega} R v \, d\mathbf{x}. \tag{50.71}$$

At the Dirichlet parts of the boundary, namely at the Ohmic contacts $\Gamma_{\text{O},\alpha}$ introduced in Section 50.3, we assume that test functions v vanish and the ansatz functions denoted by ψ, φ_n, and φ_p fulfill the Dirchlet boundary conditions ((50.35)–(50.37)). Note, that the gate boundary conditions (50.38) have been incorparated into the boundary part of the quadratic form (50.69).

We point out that for Poisson's equation $\alpha^{\psi}(\psi) = \varepsilon_s$ and for the continuity equations we have $\alpha^{\varphi_n}(\varphi_n) = \mu_n n$ and $\alpha^{\varphi_p}(\varphi_p) = \mu_p p$, respectively. Due to the formulation of the problem with quasi-Fermi potentials as unknowns, this means that $\alpha^{\varphi_n}(\varphi_n)$ and $\alpha^{\varphi_p}(\varphi_p)$ introduce nonlinearities in a^{φ_n} and a^{φ_p}. The integrals on the left-hand side of Equations 50.69 through 50.71 have no analytical expressions and are approximated using quadrature rules [70].

Essential for this finite element scheme is the underlying potential-based formulation of the flux in terms of φ_n and φ_p. If one attempts to apply the finite element method using the density-based, drift-diffusion formulation of the flux, one encounters a lot of technical difficulties and one has to stabilize the finite element method [61,87]. A convergence proof of the finite element method for a similar system in a somewhat different context (and without recombination) is available in [88].

50.10 Extensions and Outlook

In this chapter, we discussed the most important numerical solution techniques of the van Roosbroeck system, describing the transport of electrons and holes in semiconductors in a self-consistent electric field. This approach provides the core functionalities for a solver, which can be enhanced by models describing additional physical phenomena and more complex device structures including, for example, heterostructures. We mention a number of additional extensions and point to the corresponding literature.

50.10.1 Additional Physical Models

Many devices require a more accurate description of the carrier mobilities. For example, models of ionized impurity scattering, high-field drift velocity saturation and similar effects [13,89] introduce dependencies of the mobilities on carrier densities and the electric field strength, resulting in even more nonlinear couplings in the system.

The carrier transport in organic semiconductors is governed by hopping transport between the energy states of neighboring molecules. The van Roosbroeck system can be used to describe this effectively [8,9]. This may require the use of statistical distribution functions that reflect the distribution of the energy transport levels such as Blakemore for δ-shaped densities of states or the Gauss–Fermi integral [10]. In this context, the correct treatmeant of the diffusion enhancement is vital. One also has to account for the nonlinear mobilities [8,90] related to the energetic disorder characteristic for organic semiconductors.

Spin-polarized drift-diffusion models have been proposed in [91] for the description of spintronic devices. They generalize the van Roosbroeck system, introducing spin-resolved densities for electron and holes and additional mechanisms describing the spin relaxation. They have been recently studied also from a numerical point of view in [92–94].

50.10.2 Coupling the van Roosbroeck System to Other Models

When modeling lasers one has to couple the van Roosbroeck system to equations for the optical modes. This introduces additional recombination processes, describing stimulated emission. Additionally, one has to consider balance equations for the photon number. A comprehensive model (also including heating effects) along with a numerical solution strategy is given in [14].

In semiconductor devices with embedded nanostructures (such as quantum dots and quantum wells), one has to couple the drift-diffusion equations to equations describing the dynamics of the carriers localized in the nanostructures [4,5,7]. A multiscale approach for coupling atomistic with continuum drift-diffusion models is presented in [95].

If heating effects become important, the van Roosbroeck system needs to be extended by an energy transport model for the heat flow in the device. For a thermodynamically consistent extension of the van Roosbroeck system to account for this effect we refer to [14,31].

50.10.3 Methods for Doping Optimization

An important task when designing semiconductor devices is finding a suitable doping profile. This is an analytical and numerical challenge. Electric properties of the device can be improved by optimizing the doping profiles using suitable objective functionals [96–98]. Recently, the approach introduced in [96] has been extented for the optimization of doping profiles in lasers [99].

50.10.4 Alternative Modeling Approaches for Carrier Transport

The van Roosbroeck system uses implicitly the assumption that the carrier ensemble is locally in quasi equilibrium. When this assumption is not met, for example, for hot electrons, one has to consider alternative approaches. The most common ones are hydrodynamic models [100] and approximations of the Boltzmann transport equation which are derived from a spherical harmonics expansion [101]. The book [102] gives a mathematically oriented overview of these topics.

Acknowledgments

The authors are grateful for innumerable discussions with K. Gärtner and H. Gajewski, who helped to shape their still incomplete knowledge on the topic. This work has been supported by ERC-2010-AdG no. 267802 "Analysis of Multiscale Systems Driven by Functionals" (N.R.), by the Deutsche Forschungsgemeinschaft DFG within CRC 787 "Semiconductor Nanophotonics" (T.K., N.R., M.K) and partially funded in the framework of the project "Macroscopic Modeling of Transport and Reaction Processes in Magnesium-Air-Batteries" (grant 03EK3027D) under the research initiative "Energy storage" of the German Federal government.

References

1. W. van Roosbroeck. Theory of the flow of electrons and holes in germanium and other semiconductors. *Bell Syst. Tech. J.*, 29(4):560–607, 1950.
2. S. M. Sze and K. K. Ng. *Physics of Semiconductor Devices*, 3rd edition. Wiley, 2006. ISBN: 978-0-471-14323-9.
3. J. Piprek. *Optoelectronic Devices: Advanced Simulation and Analysis*. Berlin: Springer, 2005.
4. M. Grupen and K. Hess. Simulation of carrier transport and nonlinearities in quantum-well laser diodes. *IEEE J. Quant. Electron.*, 34(1):120–140, 1998.
5. S. Steiger, R. G. Veprek, and B. Witzigmann. Unified simulation of transport and luminescence in optoelectronic nanostructure. *J. Comput. Electron.*, 7(4):509–520, 2008.
6. T. Koprucki, H.-C. Kaiser, and J. Fuhrmann. Electronic states in semiconductor nanostructures and upscaling to semi-classical models. In A. Mielke, editor, *Analysis, Modeling and Simulation of Multiscale Problems*, pp. 365–394. Heidelberg, Berlin: Springer, 2006.
7. T. Koprucki, A. Wilms, A. Knorr, and U. Bandelow. Modeling of quantum dot lasers with microscopic treatment of Coulomb effects. *Opt. Quant. Electron.*, 42:777–783, 2011.
8. R. Coehoorn, W. F. Pasveer, P. A. Bobbert, and M. A. J. Michels. Charge-carrier concentration dependence of the hopping mobility in organic materials with Gaussian disorder. *Phys. Rev. B.*, 72(15):155206, 2005.
9. S. L. M. van Mensfoort and R. Coehoorn. Effect of Gaussian disorder on the voltage dependence of the current density in sandwich-type devices based on organic semiconductors. *Phys. Rev. B.*, 78(8):085207, 2008.
10. G. Paasch and S. Scheinert. Charge carrier density of organics with Gaussian density of states: Analytical approximation for the Gauss-Fermi integral. *J. Appl. Phys.*, 107(10):104501, 2010.

11. D. L. Scharfetter and H. K. Gummel. Large-signal analysis of a silicon Read diode oscillator. *IEEE Trans. Electron. Dev.*, 16(1):64–77, 1969.
12. W. Fichtner, D. J. Rose, and R. E. Bank. Semiconductor device simulation. *SIAM J. Sci. Stat. Comput.*, 4(3):391–415, 1983.
13. S. Selberherr. *Analysis and Simulation of Semiconductor Devices*. Berlin: Springer, 1984.
14. U. Bandelow, H. Gajewski, and R. Hünlich. Fabry–Perot lasers: Thermodynamics-based modeling. In J. Piprek, editor, *Optoelectronic Devices*. Berlin: Springer, 2005.
15. A. Mielke. A gradient structure for reaction-diffusion systems and for energy-drift diffusion systems. *Nonlinearity*, 24(4):1329, 2011.
16. S. R. de Groot and P. Mazur. *Non-Equilibrium Thermodynamics*. North Holland: Dover, 1962.
17. P. Deuflhard *Newton Methods for Nonlinear Problems: Affine Invariance and Adaptive Algorithms*, volume 35. Berlin: Springer, 2011.
18. D. N. Allen and R. V. Southwell. Relaxation methods applied to determine the motion, in two dimensions, of a viscous fluid past a fixed cylinder. *Quart. J. Mech. Appl. Math.*, 8:129–145, 1955.
19. A. M. Il'in. A difference scheme for a differential equation with a small parameter multiplying the second derivative. *Mat. Zametki.*, 6:237–248, 1969.
20. J. S. Chang and G. Cooper. A practical difference scheme for Fokker-Planck equations. *J. Comput. Phys.*, 6(1):1–16, 1970.
21. H.-G. Roos, M. Stynes, and L. Tobiska. *Robust Numerical Methods for Singularly Perturbed Differential Equations*, volume 24. Berlin: Springer, 2008.
22. K. W. Morton. *Numerical Solution of Convection-Diffusion Problems: Applied Mathematics*. London: Taylor & Francis, 1996.
23. D. Schröder. *Modelling of Interface Carrier Transport for Device Simulation*. Wien: Springer, 1994.
24. S. Y. Karpov. *Visible Light-Emitting Diodes*, Chapter 14. In J. Piprek, editor, *Nitride Semiconductor Devices: Principles and Simulation*, pp. 303–325. Weinheim, Germany: Wiley-VCH Verlag GmbH & Co. KGaA, 2007.
25. J. S. Blakemore. The parameters of partially degenerate semiconductors. *Proc. Phys. Soc. London A.*, 65:460–461, 1952.
26. M. Gruber, E. Zojer, F. Schürrer, and K. Zojer. Impact of materials versus geometric parameters on the contact resistance in organic thin-film transistors. *Adv. Funct. Mater.*, 23(23):2941–2952, 2013.
27. T. Koprucki and K. Gärtner. Discretization scheme for drift-diffusion equations with strong diffusion enhancement. *Opt. Quant. Electron.*, 45(7):791–796, 2013.
28. R. Haase. *Thermodynamics of Irreversible Processes*. Berlin: Addison-Wesley, 1968.
29. H. Gajewski and K. Gärtner. On the discretization of van Roosbroeck's equations with magnetic field. *Z. Angew. Math. Mech.*, 76(5):247–264, 1996.
30. H. Gajewski and K. Gröger. Semiconductor equations for variable mobilities based on Boltzmann statistics or Fermi-Dirac statistics. *Math. Nachr.*, 140(1):7–36, 1989.
31. G. Albinus, H. Gajewski, and R. Hünlich. Thermodynamic design of energy models of semiconductor devices. *Nonlinearity*, 15(2):367, 2002.
32. M. S. Mock. On equations describing steady-state carrier distributions in a semiconductor device. *Comm. Pure Appl. Math.*, 25(25):781–792, 1972.
33. H. Gajewski and K. Gröger. On the basic equations for carrier transport in semiconductors. *J. Math. Anal. Appl.*, 113:12–35, 1986.
34. H. Gajewski. On existence, uniqueness and asymptotic behavior of solutions of the basic equations for carrier transport in semiconductors. *Z. Angew. Math. Mech.*, 65(2):101–108, 1985.
35. P. A. Markowich. A nonlinear eigenvalue problem modelling the avalanche effect in semiconductor diodes. *SIAM J. Math. Anal.*, 6:1268–1283, 1985.
36. A. Jüngel. On the existence and uniqueness of transient solutions of a degenerate nonlinear drift-diffusion model for semiconductors. *Math. Models Methods Appl. Sci.*, 04:677, 1994.

37. J. W. Jerome. *Analysis of Charge Transport. A Mathematical Study of Semiconductor Devices*. Berlin Heidelberg: Springer, 1996.
38. H. K. Gummel. A self-consistent iterative scheme for one-dimensional steady state transistor calculations. *IEEE Trans. Electron Dev.*, 11(10):455–465, 1964.
39. P. A. Markowich. *The Stationary Semiconductor Device Equations*. Berlin: Springer-Verlag, 1986.
40. D. Gilbarg and N. S. Trudinger. *Elliptic Partial Differential Equations of Second Order*. Berlin: Springer, 2001. Reprint of the 1998 edition.
41. I. Rubinstein. *Electro-Diffusion of Ions*, volume 11. Philadelphia, PA: SIAM, 1990.
42. R. H. Macneal. An asymmetrical finite difference network. *Quart. Math. Appl.*, 11:295–310, 1953.
43. P. Fleischmann and S. Selberherr. Three-dimensional Delaunay mesh generation using a modified advancing front approach. In *Proceedings of IMR97*, pp. 267–278, 1997.
44. R. Eymard, T. Gallouët, and R. Herbin. Finite volume methods. In *Handbook of Numerical Analysis*, P.G. Ciarlet and J.L. Lions (Eds.), 7:713–1018, 2000.
45. J. W. Barrett and C. M. Elliott. Finite element approximation of the Dirichlet problem using the boundary penalty method. *Numer. Math.*, 49(4):343–366, 1986.
46. R. Eymard, J. Fuhrmann, and K. Gärtner. A finite volume scheme for nonlinear parabolic equations derived from one-dimensional local Dirichlet problems. *Numer. Math.*, 102(3):463–495, 2006.
47. K. Gärtner. Existence of bounded discrete steady state solutions of the van Roosbroeck system with monotone Fermi-Dirac statistic functions. *J. Comput. Electron.*, 14(3):773–787, 2015.
48. M. Bessemoulin-Chatard. A finite volume scheme for convection–diffusion equations with nonlinear diffusion derived from the Scharfetter–Gummel scheme. *Numer. Math.*, 121(4):637–670, 2012.
49. T. Koprucki, N. Rotundo, P. Farrell, D. H. Doan, and J. Fuhrmann. On thermodynamic consistency of a Scharfetter–Gummel scheme based on a modified thermal voltage for drift-diffusion equations with diffusion enhancement. *Opt. Quant. Electron.*, 47(6):1327–1332, 2015.
50. J. Fuhrmann. Comparison and numerical treatment of generalised Nernst-Planck models. *Comput. Phys. Commun.*, 196:166–178, 2015.
51. H. Gajewski. *Analysis und Numerik von Ladungstransport in Halbleitern*. WIAS Report No. 6. Berlin: WIAS, 1993. ISSN 0942-9077.
52. H. Gajewski, M. Liero, R. Nürnberg, and H. Stephan. *WIAS-TeSCA User Manual 1.2*. WIAS Technical Report No. 14. Berlin: WIAS, 2016. ISSN: WIAS 1618-7776.
53. A. Glitzky and K. Gärtner. Energy estimates for continuous and discretized electro-reaction–diffusion systems. *Nonlinear Anal.*, 70(2):788–805, 2009.
54. C. Chainais-Hillairet. Entropy method and asymptotic behaviours of finite volume schemes. In J. Fuhrmann, M. Ohlberger, and C. Rohde, editors, *Finite Volumes for Complex Applications VII – Methods and Theoretical Aspects*, pp. 17–35. Berlin: Springer, 2014.
55. M. Bessemoulin-Chatard and C. Chainais-Hillairet. Exponential decay of a finite volume scheme to the thermal equilibrium for drift–diffusion systems. *J. Numer. Math.*, 2016. (Online first, doi: 10.1515/jnma-2016-0007)

    M. Bessemoulin-Chatard and C. Chainais-Hillairet. Exponential decay of a finite volume scheme to the thermal equilibrium for drift–diffusion systems. *arXiv preprintarXiv:1601.00813*, 2016.
56. K. Gärtner. Existence of bounded discrete steady-state solutions of the van Roosbroeck system on boundary conforming delaunay grids. *SIAM J. Sci. Comput.*, 31(2):1347–1362, 2009.
57. R. D. Lazarov, I. D. Mishev, and P. S. Vassilevski. Finite volume methods for convection diffusion problems. *SIAM J. Numer. Anal.*, 33(1):31–55, 1996.
58. J. J. H. Miller and S. Wang. An analysis of the Scharfetter-Gummel box method for the stationary semiconductor device equations. *RAIRO-Math. Num.*, 28(2):123–140, 1994.
59. J. Xu and L. Zikatanov. A monotone finite element scheme for convection-diffusion equations. *Math. Comput. Am. Math. Soc.*, 68(228):1429–1446, 1999.

60. C. Chainais-Hillairet, J.-G. Liu, and Y.-J. Peng. Finite volume scheme for multidimensional drift-diffusion equations and convergence analysis. *ESAIM: Math. Model. Numer. Anal. Math. Num.*, 37(2):319–338, 2003.
61. R. E. Bank, W. M. Coughran Jr, and L. C. Cowsar. The finite volume Scharfetter-Gummel method for steady convection diffusion equations. *Comput. Vis. Sci.*, 1(3):123–136, 1998.
62. M. Augustin, A. Caiazzo, A. Fiebach, J. Fuhrmann, V. John, A. Linke, and R. Umla. An assessment of discretizations for convection-dominated convection-diffusion equations. *Comp. Meth. Appl. Mech. Eng.*, 200:3395–3409, 2011.
63. A. Kuzmin, M. Luisier, and O. Schenk. Fast methods for computing selected elements of the greens function in massively parallel nanoelectronic device simulations. In F. Wolf, B. Mohr, and D. Mey, editors, *Euro-Par 2013 Parallel Processing*, volume 8097, *Lecture Notes in Computer Science*, pp. 533–544. Berlin: Springer, 2013.
64. O. Schenk, M. Bollhöfer, and R. A. Römer. On large-scale diagonalization techniques for the Anderson model of localization. *SIAM Rev.*, 50(1):91–112, 2008.
65. O. Schenk, A. Wächter, and M. Hagemann. Matching-based preprocessing algorithms to the solution of saddle-point problems in large-scale nonconvex interior-point optimization. *Comput. Optim. Appl.*, 36(2–3):321–341, 2007.
66. O. Schenk. PARDISO version 5.0.0. http://www.pardiso-project.org. Accessed on February 22, 2016.
67. T. A. Davis. *Direct Methods for Sparse Linear Systems (Fundamentals of Algorithms 2)*. Philadelphia, PA: Society for Industrial and Applied Mathematics, 2006.
68. T. Davis. SuiteSparse version 4.5.1. http://faculty.cse.tamu.edu/davis/suitesparse.html. Accessed on February 22, 2016.
69. Y. Saad. *Iterative Methods for Sparse Linear Systems*. Philadelphia, PA: SIAM, 2003.
70. P. Ciarlet. *The Finite Element Method for Elliptic Problems*. Philadelphia, PA: SIAM, 2002.
71. H. Si, K. Gärtner, and J. Fuhrmann. Boundary conforming Delaunay mesh generation. *Comput. Math. Math. Phys.*, 50(1):38–53, 2010.
72. J. Shewchuk. Triangle: A two-dimensional quality mesh generator and Delaunay triangulator. http://www.cs.cmu.edu/~quake/triangle.html. Accessed on December 1, 2015.
73. H. Edelsbrunner. *Algorithms in Combinatorial Geometry*. Heidelberg: Springer-Verlag, 1987.
74. S.-W. Cheng, T. K. Dey, and J. R. Shewchuk. *Delaunay Mesh Generation*. Boca Raton, FL: Chapman & Hall/CRC, 1st edition, 2012.
75. H. Si. TetGen version 1.5. http://tetgen.org/. Accessed on 2015-12-01.
76. J. Schöberl. Netgen an advancing front 2d/3d-mesh generator based on abstract rules. *Comput. Vis. Sci.*, 1(1):41–52, 1997.
77. M. S. Shephard and M. K. Georges. Automatic three-dimensional mesh generation by the finite octree technique. *Int. J. Numer. Meth. Eng.*, 32(4):709–749, 1991.
78. K. Gärtner and R. Richter. DEPFET sensor design using an experimental 3d device simulator. *Nucl. Instrum. Meth. A.*, 568(1):12–17, 2006.
79. B. Schmithüsen, K. Gärtner, and W. Fichtner. Grid adaptation for device simulation according to the dissipation rate. In *Simulation of Semiconductor Processes and Devices 1998*, pp. 197–200. Berlin: Springer, 1998.
80. J. Krause, N. Strecker, and W. Fichtner. Boundary-sensitive mesh generation using an offsetting technique. *Int. J. Numer. Meth. Eng.*, 49(1–2):51–59, 2000.
81. A. Fiebach, A. Glitzky, and A. Linke. Uniform global bounds for solutions of an implicit voronoi finite volume method for reaction–diffusion problems. *Numer. Math.*, 128(1):31–72, 2014.
82. A. Glitzky. Exponential decay of the free energy for discretized electro-reaction–diffusion systems. *Nonlinearity*, 21(9):1989, 2008.
83. E. Hairer, S. P. Norsett, and G. Wanner. *Solving Ordinary Differential Equations: Nonstiff Problems. v. 2: Stiff and Differential-Algebraic Problems*. Berlin: Springer Verlag, 2010.

84. M. S. Mock. A time-dependent numerical model of the insulated-gate field-effect transistor. *Solid-State Electron.*, 24(10):959–966, 1981.
85. P. D. Yoder, K. Gärtner, U. Krumbein, and W. Fichtner. Optimized terminal current calculation for Monte Carlo device simulation. *IEEE Trans. Comput-Aided Design Integr. Circuits Syst.*, 16(10):1082–1087, 1997.
86. M. Auf Der Maur. *A Multiscale Simulation Environment for Electronic and Optoelectronic Devices*. PhD thesis, University of Tor Vergata, 2008.
87. V. John and P. Knobloch. On spurious oscillations at layers diminishing (SOLD) methods for convection–diffusion equations: Part I–A review. *Comput. Methods Appl. Mech. Eng.*, 196(17):2197–2215, 2007.
88. A. Prohl and M. Schmuck. Convergent discretizations for the Nernst-Planck-Poisson system. *Numer. Math.*, 111(4):591–630, 2009.
89. V. Palankovski and R. Quay. *Analysis and Simulation of Heterostructure Devices*. Computational Microelectronics. Vienna: Springer Science & Business Media, 2004.
90. H. Bässler. Charge transport in disordered organic photoconductors a Monte Carlo simulation study. *Phys. Stat. Sol. (B).*, 175(1):15–56, 1993.
91. J. Fabian, S. C. Erwin, and I. Zutic. Bipolar spintronics: Fundamentals and applications. *IBM J. Res. Dev.*, 50(1):121–139, 2006.
92. A. Glitzky and K. Gärtner. Existence of bounded steady state solutions to spin-polarized drift-diffusion systems. *SIAM J. Math. Anal.*, 41(6):2489–2513, 2010.
93. A. Jüngel, C. Negulescu, and P. Shpartko. Bounded weak solutions to a matrix drift diffusion model for spin-coherent electron transport in semiconductors. *Math. Models Methods Appl. Sci.*, 25(5):929–958, 2015.
94. C. Chainais-Hillairet, A. Jüngel, and P. Shpartko. A finite-volume scheme for a spinorial matrix drift-diffusion model for semiconductors. *Numer. Methods for Partial Differ. Equ.*, 32(3):819–846, 2016.
95. M. Auf der Maur, A. Pecchia, G. Penazzi, F. Sacconi, and A. Carlo. Coupling atomistic and continuous media models for electronic device simulation. *J. Comput. Electron.*, 12(4):553–562, 2013.
96. M. Hinze and R. Pinnau. An optimal control approach to semiconductor design. *Math. Models Methods Appl. Sci.*, 12(1):89–107, 2002.
97. M. Hinze and R. Pinnau. Second-order approach to optimal semiconductor design. *J. Optim. Theory. Appl.*, 133(2):179–199, 2007.
98. M. Burger and R. Pinnau. Fast optimal design of semiconductor devices. *SIAM J. Appl. Math.*, 64(1):108–126, 2003.
99. D. Peschka, N. Rotundo, and M. Thomas. Towards doping optimization of semiconductor lasers. To appear in. *J. Comput. Theor. Trans.*, 45(5):410–423, 2016.
100. T. Grasser, T.-W. Tang, H. Kosina, and S. Selberherr. A review of hydrodynamic and energy-transport models for semiconductor device simulation. *Proc. IEEE*, 91(2):251–274, 2003.
101. S.-M. Hong and C. Jungemann. A fully coupled scheme for a Boltzmann-Poisson equation solver based on a spherical harmonics expansion. *J. Comput. Electron.*, 8(3):225–241, 2009.
102. A. Jüngel. *Transport Equations for Semiconductors*, volume 773. Berlin: Springer, 2009.

Taylor & Francis
Taylor & Francis Group
http://taylorandfrancis.com

51

Monte Carlo Device Simulations

Katerina Raleva
Abdul R. Shaik
Raghuraj Hathwar
Akash Laturia
Suleman S. Qazi
Robin Daugherty
Dragica Vasileska
and
Stephen M. Goodnick

51.1 Introduction.. 773
Transistor Scaling • More Moore Technology • Importance of Modeling and Simulation
51.2 Bulk Monte Carlo Method Description 779
51.3 Particle-Based Device Simulation....................................... 791
Calculation of the Current • Ohmic Contacts • Time-Step • PM Coupling
51.4 Representative Simulation Results 796
Bulk Monte Carlo Simulations of Different Materials • Modeling Thermal Effects in SOI Devices Using Particle-Based Device Simulator
51.5 Outlook... 803

51.1 Introduction

The semiconductor device–based electronics industry has been the largest industry in the world with global sales of over a trillion dollars since 1998. The revolution in the semiconductor industry, a large portion of the electronics industry, began in 1947 with the fabrication of bipolar devices on slabs of polycrystalline germanium (Ge). Single-crystalline materials were later proposed and introduced that made possible the fabrication of grown junction transistors. Migration to silicon (Si)–based devices was initially hindered by the stability of the Si/SiO_2 materials system, necessitating a new generation of crystal pullers with improved environmental controls to prevent SiO_2 formation. Later, the stability and low interface-state density of the Si/SiO_2 materials system provided passivation of junctions and eventually, the migration from bipolar devices to field-effect devices in 1960. By 1968, both complementary metal–oxide–semiconductor (CMOS) devices and poly-Si gate technology that allowed self-alignment of the gate to the source/drain of the device had been developed. These innovations permitted a significant reduction in power dissipation and a reduction of the overlap capacitance, improving frequency performance and resulting in the essential components of the modern CMOS device.

Since the invention of the bipolar transistor in 1947, the number and variety of semiconductor devices have increased tremendously as advanced technology, new materials, and broadened comprehension have been applied to the creation and innovation of new devices. Dr. William Shockley suggested in the fifties the use of semiconductors of different bandgap for the fabrication of heterostructure devices. Professor Herbert Kroemer's contributions to heterostructures—from heterostructures bipolar transistors to lasers—culminated in a Nobel Prize in Physics in 2000 and have paved the way for novel heterostructure devices including those in silicon. The unique properties of semiconductor materials have enabled the development of a wide variety of ingenious devices that have literally changed our world. To date, there are about

TABLE 51.1 Major Semiconductor Device Discoveries

- 1874: Metal-semiconductor contact
- 1907: Light-emitting diode
- 1947: Bipolar junction transistor (BJT)
- 1954: Solar cell
- 1957: Heterojunction bipolar transistor (HBT)
- 1958: Tunnel diode
- 1959: Integrated circuits
- 1960: Field-effect transistors (FETs)
- 1962: Semiconductor lasers
- 1966: Metal-semiconductor FET (MESFET)
- 1967: Nonvolatile semiconductor memory
- 1974: Resonant tunneling diode (RTD)
- 1980: Modulation (MOD) FET
- 1994: Room-temperature single-electron memory cell (SEMC)
- 2001: 15 nm MOSFET
- 2006: 3 nm FinFET
- 2012: Single atom transistor in Si technology

60 major devices, with over 100 device variations related to them. A list of most of the basic semiconductor devices discovered and used over the past century with the date of their introduction is shown in Table 51.1.

The metal–oxide–semiconductor field-effect transistor (MOSFET) and related integrated circuits (ICs) now constitute about 90% of the semiconductor device market. Combining Si with the elegance of the field-effect transistor (FET) structure has allowed simultaneously making devices smaller, faster, and cheaper—the mantra that has driven the modern semiconductor microelectronics industry. Indeed, the single factor driving continuous device improvement has been the semiconductor industry's relentless effort to reduce the cost per function on a chip. The way this is done is to put more devices on a chip while either reducing manufacturing costs or holding them constant. This leads to three methods of reducing the cost per function. The first is *transistor scaling*, which involves reducing the transistor size in accordance with some goal, i.e., keeping the electric field constant from one generation to the next. With smaller transistors, more can fit into a given area than in previous generations. The second method is *circuit cleverness*, which is associated with the physical layout of the transistors with respect to each other. If the transistors can be packed into a tighter space, then more devices can fit into a given area than before. The third method is to make the die larger. More devices can be fabricated on a *larger die*. All the while, the semiconductor industry is constantly looking for technological breakthroughs to decrease the manufacturing cost. All of this effort serves to reduce the cost per function on a chip.

51.1.1 Transistor Scaling

Device engineers are most concerned with the method of scaling introduced in Section 51.1. The semiconductor industry has been so successful in providing continued system performance improvement year after year that the Semiconductor Industry Association (SIA) has been publishing roadmaps for semiconductor technology since 1992. These roadmaps represent a consensus outlook of industry trends, taking history as a guide. SIA roadmaps [1] incorporate participation from the global semiconductor industry, including the United States, Europe, Japan, Korea, and Taiwan. They basically affirm the desire of the industry to continue with *Moore's law* [2], which is often stated as the doubling of transistor performance and quadrupling of the number of devices on a chip every 3 years. The phenomenal progress signified by Moore's law has been achieved through scaling of the MOSFET from larger to smaller physical dimensions [3].

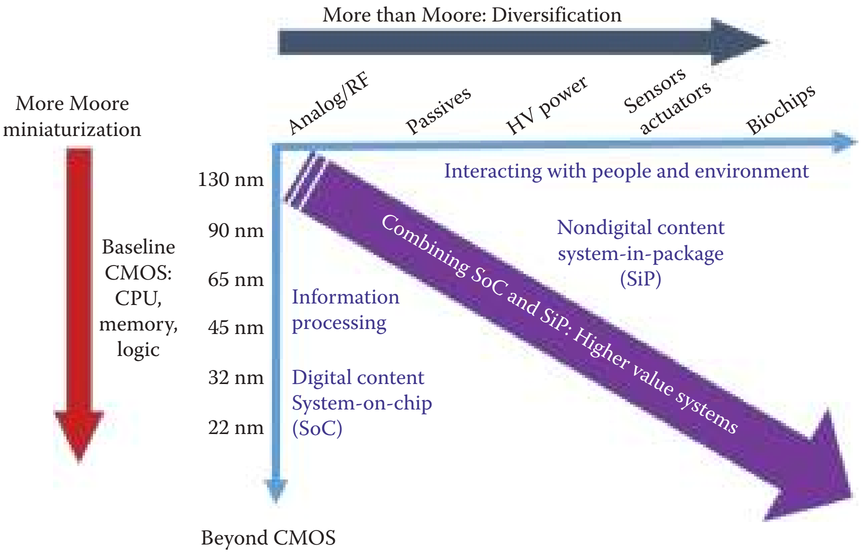

More than Moore: Diversification
More Moore miniaturization
Analog/RF
Passives
HV power
Sensors actuators
Biochips
Interacting with people and environment
130 nm
90 nm
65 nm
45 nm
32 nm
22 nm
Baseline CMOS: CPU, memory, logic
Combining SoC and SiP: Higher value systems
Nondigital content system-in-package (SiP)
Information processing
Digital content System-on-chip (SoC)
Beyond CMOS

FIGURE 51.1 Schematic description of More Moore, More than Moore, and beyond complementary metal–oxide–semiconductor (CMOS). (RF, radio frequency; HV, high-voltage. [Reprinted from ITRS Roadmap for Semiconductors. Available at: www.itrs.org. With permission.])

The *More Moore* domain is internationally defined as an attempt to further develop advanced CMOS technologies and reduce the associated cost per function by introducing strain and silicon-on-insulator (SOI) technology (see Figure 51.1). Almost 70% of the total semiconductor components market is directly impacted by advanced CMOS miniaturization achieved in the More Moore domain. This 70% comprises three component groups of similar size, namely microprocessors, mass memories, and digital logic. As an example, in recent years, technology has advanced to fabricate ICs at 14-nm gate length commercially [4]. Fab industry giants such as Intel, TSMC, Samsung, and Global Foundries have plans to fabricate ICs at 10 nm by 2017. Samsung has already fabricated and tested 128 Mb SRAM in 10 nm [5] and is hoping to commercialize the process by the end of 2016. This aggressive scaling of technology is possible because of the advent of FinFETs and fully-depleted silicon-on-insulator (FD-SOI) device technology [6].

More than Moore (MtM) refers to a set of technologies that enable nondigital micro-/ nanoelectronic functions. MtM radio frequency (RF), high-voltage power, solid-state lighting (SSL), medical ultrasound, biochips and microfluidics, energy scavenging, electronic imaging, sensors, and actuators on CMOS platforms are some of the applications. They are based on, or derived from, Si technology but do not necessarily scale with Moore's Law. MtM devices typically provide conversion of nondigital as well as nonelectronic information, such as mechanical, thermal, acoustic, chemical, optical, and biomedical functions to digital data, and vice versa.

51.1.2 More Moore Technology

For digital circuits, a figure of merit for MOSFETs for unloaded circuits is CV/I, where C is the gate capacitance, V the voltage swing, and I the current drive of the MOSFET. For loaded circuits, the current drive of the MOSFET is of paramount importance. Keeping in mind both the CV/I metric and the benefits of a large current drive, one notes that device performance may be improved [7] by (1) inducing a larger charge density for a given gate voltage drive; (2) enhancing the carrier transport by improving the mobility, saturation velocity, or ballistic transport; (3) ensuring device scalability to achieve a shorter channel length; and (4) reducing parasitic capacitances and resistances. For capitalizing these opportunities, the proposed technology options generally fall into two categories: new materials and new device structures. In many cases, the introduction of a new material requires the use of a new device structure, or vice versa. To fabricate devices beyond current scaling limits, IC companies have simultaneously pushed the planar, bulk Si CMOS design while incorporating alternative gate stack materials (high-k dielectric [8] and metal gates), band engineering methods (using strained Si [9] or SiGe [10]), and alternative transistor structures. The concept of a band-engineered transistor is to enhance the mobility of electrons and/or holes in the channel by modifying the band structure of Si in the channel in a way such that the physical structure of the transistor remains substantially unchanged. This enhanced mobility increases the transistor transconductance (g_{m}) and on-drive current (I_{on}). A SiGe layer or a strained Si on a relaxed SiGe layer is used as an enhanced-mobility channel layer. It has already been demonstrated experimentally that at $T = 300$ K (room temperature), effective hole enhancement of about 50% can be achieved using the SiGe technology [10]. Intel has adopted strained Si technology for the first time for its 90-nm process [11]. This resulted in nearly 20% performance improvement per technology generation, with only a few additional process steps. The challenge in identifying suitable high-k dielectrics and metal gates for both conventional p-channel MOS (PMOS) and n-channel MOS (NMOS) transistors has led to early adoption of alternative transistor designs. These include primarily partially-depleted (PD) SOI and FD-SOI devices. Today, there is extensive research in double-gate (DG) structures and FinFET transistors [12], which have better electrostatic integrity and, theoretically, have better transport properties than single-gated FETs. A FinFET is a form of a DG transistor having surface conduction channels on two opposite vertical surfaces and having current flow in the horizontal direction. The channel length is given by the horizontal separation between source and drain and is usually determined by a lithographic step combined with a side-wall spacer etch process. Many innovative structures, involving structural challenges such as fabrication on nanometer-scale fins and nanometer-scale planarization over an entire wafer are currently being implemented. Some new and revolutionary technology such as carbon nanotubes or molecular transistors might be on the horizon, but it is not clear, in view of the predicted future capabilities of CMOS, whether they will be competitive. In conclusion, the semiconductor industry is approaching the end of an era of scaling gains by rote shrinkage of device dimensions and entering a postscaling era, a new phase of CMOS evolution in which innovation is demanded simply to compete. The trends in benefits to density, performance, and power will be continued through such innovations. Rather than coming to a close, a new era of CMOS technology is just beginning.

51.1.3 Importance of Modeling and Simulation

As semiconductor devices are scaled into nanoscale regimes, first, velocity saturation starts to limit the carrier mobility due to pronounced intervalley scattering; when the device dimensions are scaled below 200 nm, velocity overshoot (velocity of the carriers in the channel is larger than the steady-state saturation velocity) starts to dominate the device behavior leading to larger on-state currents. Alongside with the developments in the semiconductor nanotechnology, in recent years, there has been significant progress in physically based modeling of semiconductor devices. There are two issues that make simulation important. Product cycles are getting shorter with each generation, and the demand for production wafers shadows development efforts in the factory. Consider the product cycle issue first. In order for companies to maintain their competitive edge, products have to be taken from design to production in less than 18 months. As a result, the development phase of the cycle is getting shorter. Contrast this requirement with the fact that it takes 2–3 months to run a wafer lot through a factory, depending on its complexity. The specifications for experiments run through the factory must be near the final solution. While simulations may not be completely predictive, they provide a good initial guess. This can ultimately reduce the number of iterations during the device development phase. The second issue that reinforces the need for simulation is the production pressures that factories face. To meet customer demand, development factories are making way for production space. It is also expensive to run experiments through a production facility. Such resources could have otherwise been used to produce sellable products. Again, device simulation can be used to decrease the number of experiments run through a factory. Device simulation can be used as a tool to guide manufacturing down the right path, thereby decreasing the development time and costs. Besides offering the possibility to test hypothetical devices that have not (or could not) yet been manufactured, device simulation offers unique insight into device behavior by allowing the observation of phenomena that cannot be measured on real devices. It is related to, but usually separate from, process simulation, which deals with various physical processes such as material growth, oxidation, impurity diffusion, etching, and metal deposition inherent in device fabrication leading to ICs. Device simulation is distinct from another important aspect of computer-aided design (CAD), device modeling, which deals with compact behavioral models for devices and subcircuits relevant for circuit simulation in commercial packages such as simulation program with integrated circuit emphasis (SPICE). Device simulation can provide the parameters that are used to generate the compact behavioral models, and when coupled with process simulation and circuit simulation, it provides a hierarchical approach to technology computer-aided design (TCAD).

The main components of semiconductor device simulation at any level of approximation are illustrated in Figure 51.2 [13]. There are two main kernels, which must be solved self-consistently with one another, the *transport equations* governing charge flow, and the *fields* driving charge flow. Both are coupled strongly to one another, and hence must be solved simultaneously. The fields arise from external sources, as well as the charge and current densities which act as sources for the time-varying electric and magnetic fields obtained from the solution of Maxwell's equations. Under appropriate conditions, only the quasi-static electric fields arising from the solution of Poisson's equation are necessary. The fields, in turn, are driving forces for charge transport as illustrated in Figure 51.3 for the various levels of approximation within a hierarchical structure ranging from compact modeling at the top to an exact quantum-mechanical description at the bottom. For devices for which gradual channel approximation cannot be used due to the two-dimensional (2D) nature of the electrostatic potential and the electric fields driving the carriers from source to drain, drift-diffusion models have been exploited. These models are valid, in general, for large devices in which the fields are not that high so that there is no degradation of the mobility due to the electric field. The validity of the drift-diffusion models can be extended to take into account the velocity saturation effect with the introduction of field-dependent mobility and diffusion coefficients. When velocity overshoot becomes important (for length scales less than 200 nm—see Figure 51.4), the drift-diffusion model is no longer valid and the hydrodynamic model must be used.

The hydrodynamic model has been the workhorse for technology development and several high-end commercial device simulators have appeared including Silvaco, Synopsys, Crosslight, etc. The advantages of the hydrodynamic model are that it allows quick simulation runs, but the problem is that the amount of the velocity overshoot depends upon the choice of the energy relaxation time. The smaller the device,

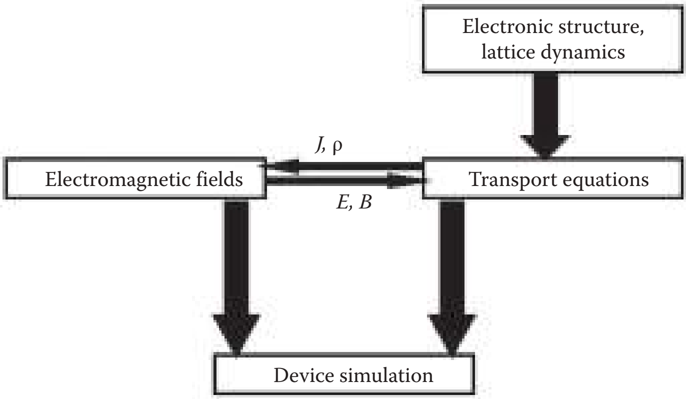
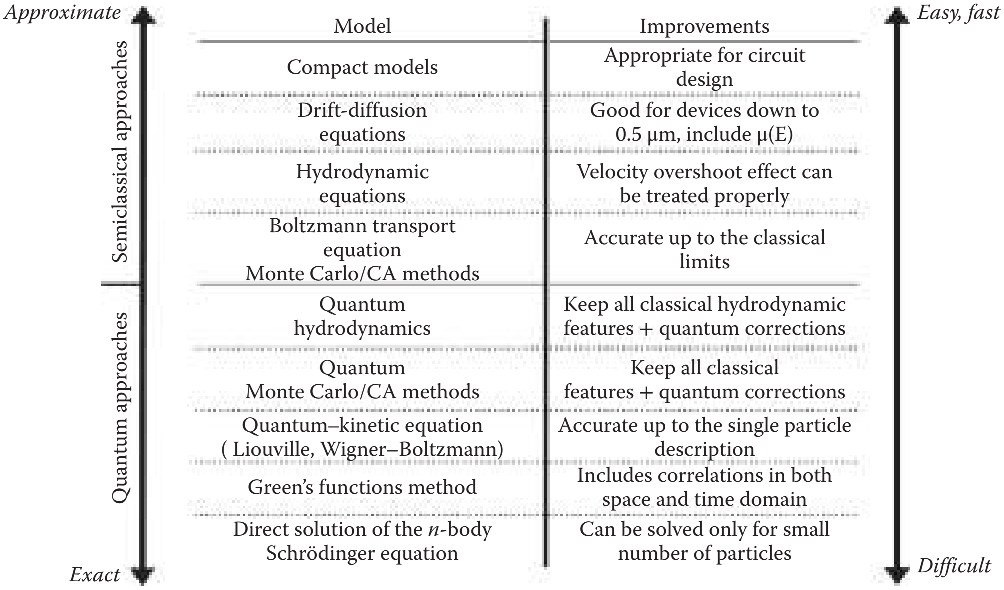


FIGURE 51.2 A schematic description of the building blocks of device simulator.


Approximate
Exact
Semiclassical approaches
Quantum approaches
Easy, fast
Difficult

Model	Improvements
Compact models	Appropriate for circuit design
Drift-diffusion equations	Good for devices down to 0.5 μm, include μ(E)
Hydrodynamic equations	Velocity overshoot effect can be treated properly
Boltzmann transport equation Monte Carlo/CA methods	Accurate up to the classical limits
Quantum hydrodynamics	Keep all classical hydrodynamic features + quantum corrections
Quantum Monte Carlo/CA methods	Keep all classical features + quantum corrections
Quantum–kinetic equation (Liouville, Wigner–Boltzmann)	Accurate up to the single particle description
Green's functions method	Includes correlations in both space and time domain
Direct solution of the *n*-body Schrödinger equation	Can be solved only for small number of particles

FIGURE 51.3 Illustration of the hierarchy of transport models.

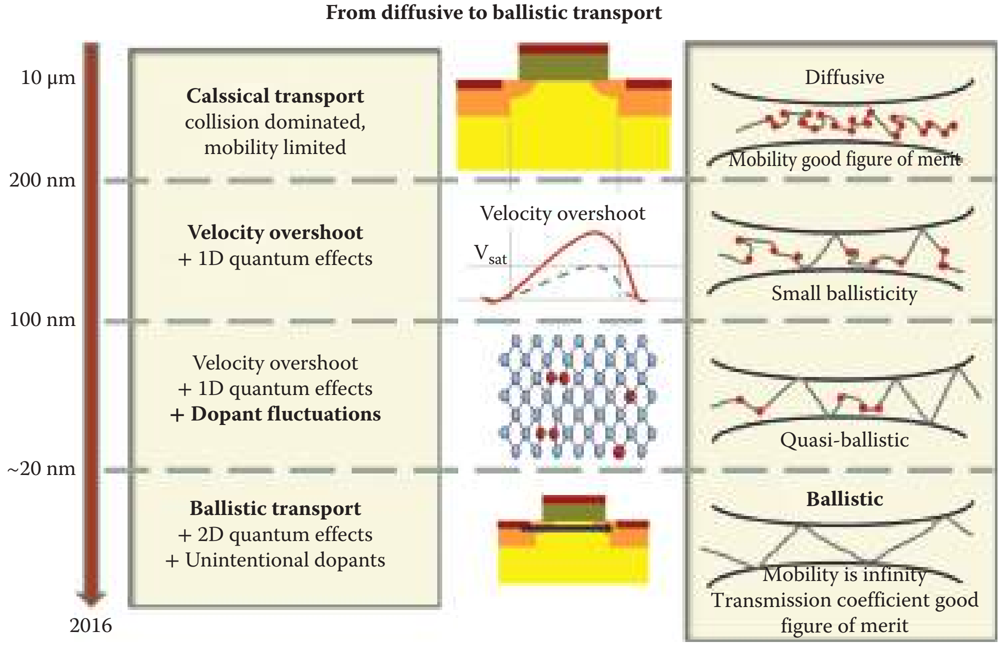

From diffusive to ballistic transport
10 μm
200 nm
100 nm
~20 nm
2016
Calssical transport
collision dominated,
mobility limited
Velocity overshoot
+ 1D quantum effects
Velocity overshoot
+ 1D quantum effects
+ Dopant fluctuations
Ballistic transport
+ 2D quantum effects
+ Unintentional dopants
Velocity overshoot
V_sat
Diffusive
Mobility good figure of merit
Small ballisticity
Quasi-ballistic
Ballistic
Mobility is infinity
Transmission coefficient good figure of merit

FIGURE 51.4 From diffusive to ballistic transport when scaling semiconductor devices into nanometer scale.

the larger the deviation when using the same set of energy relaxation times. A standard way in calculating the energy relaxation times is to use bulk Monte Carlo simulations. However, the energy relaxation times are material-, device geometry-, and doping-dependent parameters, so their determination ahead of time is not possible.

To avoid the problem of the proper choice of the energy relaxation times in hydrodynamic models, a direct solution of the Boltzmann transport equation (BTE)

$$\frac{\partial f}{\partial t} + \frac{1}{\hbar}\nabla_k E(k) \cdot \nabla_r f + \left(-\frac{e\mathbf{E}}{\hbar}\right) \cdot \nabla_k f = \left(\frac{\partial f}{\partial t}\right)_{\mathbf{coll}} \tag{51.1}$$

using the Monte Carlo method is the best method of choice for nanoscale devices in which position and momentum can still be treated as independent variables. In Equation 51.1, $f(r, k, t)$ is the semiclassical distribution function, $E(k)$ the dispersion relation for the electrons, and $\mathbf{E}$ the electric field. When, for example, collisional broadening of the states is important, quantum transport approaches must be pursued.

This chapter focuses on Monte Carlo methods for the solution of the BTE and is organized as follows. Monte Carlo solution of the BTE for homogeneous systems is given in Section 51.2. Particle-based device simulator is described in Section 51.3. Representative simulation results for bulk material and for semiconductor devices are given in Section 51.4. Conclusions and future directions of research are presented in Section 51.5.

51.2 Bulk Monte Carlo Method Description

The ensemble Monte Carlo (EMC) techniques have been used for well over 40 years as a numerical method to simulate nonequilibrium transport in semiconductor materials and devices and have been the subject of numerous books and reviews [14–16]. In the applications to transport problem, a random walk is generated using the random number–generating algorithms common to modern computers to simulate the stochastic motion of particles subject to collision processes. This process of random walk generation is part of a very general technique used to evaluate integral equations and is connected to the general random sampling technique used in the evaluation of multidimensional integrals [17].

The basic technique as applied to transport problems is to simulate the free particle motion (referred to as the *free flight*) terminated by instantaneous random *scattering events*. The Monte Carlo algorithm consists of generating random free-flight times for each particle, choosing the type of scattering occurring at the end of the free flight, changing the final energy and momentum of the particle after scattering, and then repeating the procedure for the next free flight. Sampling the particle motion at various times throughout the simulation allows for the statistical estimation of physically interesting quantities such as the single particle distribution function, the average drift velocity in the presence of an applied electric field, the average energy of the particles, etc. By simulating an *ensemble* of particles, representative of the physical system of interest, the nonstationary time-dependent evolution of the electron and hole distributions under the influence of a time-dependent driving force may be simulated. This particle-based picture, in which the particle motion is decomposed into free flights terminated by instantaneous collisions, is basically the same approximate picture underlying the derivation of the semiclassical BTE. In fact, it may be shown that the one-particle distribution function obtained from the random walk Monte Carlo technique satisfies the BTE for a homogeneous system in the long-time limit [18]. This semiclassical picture breaks down when quantum-mechanical effects become pronounced, and one cannot unambiguously describe the instantaneous position and momentum of a particle. In the following, we describe first the standard Monte Carlo algorithm used to simulate charge transport in semiconductors. We then discuss how this basic model for charge transport within the BTE is self-consistently solved with the appropriate field equations to perform particle-based device simulation (Section 51.3).

In the bulk Monte Carlo method [19], particle motion is assumed to consist of free flights terminated by instantaneous scattering events, which change the momentum and energy of the particle after scattering. So, the first task is to generate free flights of random time duration for each particle. To simulate this process, the probability density, $P(t)$, is required, in which $P(t)\mathrm{d}t$ is the joint probability that a particle will arrive at time t without scattering after a previous collision occurring at time $t = 0$, and then suffer a collision in a time interval dt around time t. The probability of scattering in the time interval dt around t may be written as $\Gamma[\mathbf{k}(t)]\mathrm{d}t$, where $\Gamma[\mathbf{k}(t)]$ is the scattering rate of an electron or hole of wavevector **k**. The scattering rate, $\Gamma[\mathbf{k}(t)]$, represents the sum of the contributions from each individual scattering mechanism, which are usually calculated quantum mechanically using perturbation theory [20]. The implicit dependence of $\Gamma[\mathbf{k}(t)]$ on time reflects the change in **k** due to acceleration by internal and external fields. For electrons subject to time-independent electric and magnetic fields, the time evolution of **k** between collisions is represented as

$$\mathbf{k}(t) = \mathbf{k}(0) - \frac{e(\mathbf{E} + \mathbf{v} \times \mathbf{B})t}{\hbar}, \tag{51.2}$$

where **E** is the electric field, **v** the electron velocity, and **B** the magnetic flux density. In terms of the scattering rate, $\Gamma[\mathbf{k}(t)]$, the probability that a particle has not suffered a collision after a time t is given by $\exp\left(-\int_0^t \Gamma\left[k(t')\right]\mathrm{d}t'\right)$. Thus, the probability of scattering in the time interval dt after a free flight of time $t$may be written as the joint probability:

$$P(t)\mathrm{d}t = \Gamma[k(t)]\exp\left[-\int_0^t \Gamma\left[k\left(t'\right)\right]\mathrm{d}t'\right]\mathrm{d}t. \tag{51.3}$$

Random flight times may be generated according to the probability density $P(t)$ above using, for example, the pseudo-random number generator implicit on most modern computers, which generate uniformly distributed random numbers in the range [0, 1]. Using a direct method, random flight times sampled from $P(t)$ may be generated according to

$$r = \int_0^{t_r} P(t)\mathrm{d}t, \tag{51.4}$$

where r is a uniformly distributed random number and t_r is the desired free-flight time. Integrating Equation 51.4 with $P(t)$ given by Equation 51.3 above yields

$$r = 1 - \exp\left[-\int_0^{t_r} \Gamma\left[k(t')\right]dt'\right]. \tag{51.5}$$

Since $1 - r$ is statistically the same as r, Equation 51.5 may be simplified to

$$-\ln r = \int_0^{t_r} \Gamma\left[\mathbf{k}\left(t'\right)\right]\mathrm{d}t'. \tag{51.6}$$

Equation 51.6 is the fundamental equation used to generate the random free-flight time after each scattering event, resulting in a random walk process related to the underlying particle distribution function. If there is no external driving field leading to a change of **k** between scattering events (e.g., in ultrafast photoexcitation experiments with no applied bias), the time dependence vanishes, and the integral is trivially evaluated. In the general case, where this simplification is not possible, it is expedient to introduce the so-called self-scattering method [21] in which one introduces a fictitious scattering mechanism whose scattering rate always adjusts itself in such a way that the total (self-scattering plus real scattering) rate is energy-independent:

$$\Gamma = \Gamma\left[\mathbf{k}(t')\right] + \Gamma_{\text{self}}\left[\mathbf{k}(t')\right], \tag{51.7}$$

where $\Gamma_{\text{self}}[\mathbf{k}(t')]$ is the self-scattering rate. The self-scattering mechanism itself is defined such that the final state before and after scattering is identical. Hence, it has no effect on the free-flight trajectory of a particle when selected as the terminating scattering mechanism, yet results in the simplification of Equation 51.6 such that the free flight is given by

$$t_r = -\frac{1}{\Gamma}\ln r. \tag{51.8}$$

The constant total rate (including self-scattering) Γ must be chosen at the start of the simulation interval (there may be multiple such intervals throughout an entire simulation) so that it is larger than the maximum scattering encountered during the same time interval. In the simplest case, a single value is chosen at the beginning of the entire simulation (constant gamma method), checking to ensure that the real rate never exceeds this value during the simulation. Other schemes may be chosen that are more computationally efficient, and which modify the choice of Γ at fixed time increments [22].

The algorithm described above determines the random free-flight times during which the particle dynamics is treated semiclassically. For the scattering process itself, we need the type of scattering (i.e., impurity, acoustic phonon, photon emission, etc.) which terminates the free flight, and the final energy and momentum of the particle(s) after scattering. The type of scattering that terminates the free flight is chosen using a uniform random number between 0 and Γ, and using this pointer to select among the relative total scattering rates of all processes including self-scattering at the final energy and momentum of the particle:

$$\Gamma = \Gamma_{\text{self}}[n,\mathbf{k}] + \Gamma_1[n,\mathbf{k}] + \Gamma_2[n,\mathbf{k}] + \cdots \Gamma_N[n,\mathbf{k}], \tag{51.9}$$

with n the band index of the particle (or sub-band in the case of reduced-dimensionality systems), and $\mathbf{k}$ the wavevector at the end of the free flight. This process is illustrated schematically in Figure 51.5.

Once the type of scattering terminating the free flight is selected, the final energy and momentum (as well as band or subband) of the particle due to this type of scattering must be selected. For elastic scattering processes such as ionized impurity scattering, the energy before and after scattering is the same. For the interaction between electrons and the vibrational modes of the lattice described as quasi-particles known as phonons, electrons exchange finite amounts of energy with the lattice in terms of emission and absorption of phonons. For determining the final momentum after scattering, the scattering rate, $\Gamma_j[n,\mathbf{k};m,\mathbf{k}']$ of the jth scattering mechanism is needed, where n and m are the initial and final band indices, and $\mathbf{k}$ and $\mathbf{k}'$ are the particle wavevectors before and after scattering. Defining a spherical coordinate system

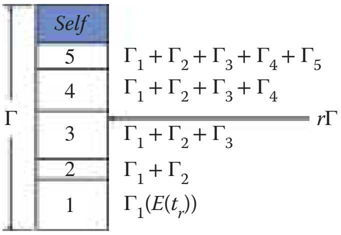

Self
5
4
3
2
1
Γ
$\Gamma_1+\Gamma_2+\Gamma_3+\Gamma_4+\Gamma_5$
$\Gamma_1+\Gamma_2+\Gamma_3+\Gamma_4$
$r\Gamma$
$\Gamma_1+\Gamma_2+\Gamma_3$
$\Gamma_1+\Gamma_2$
$\Gamma_1(E(t_r))$

FIGURE 51.5 Selection of the type of scattering terminating a free flight in the Monte Carlo algorithm.

around the initial wavevector **k**, the final wavevector **k**′ is specified by |**k**′| (which depends on conservation of energy) as well as the azimuthal and polar angles, ϕ and θ around **k**. Typically, the scattering rate, $\Gamma_j[n,\mathbf{k};m,\mathbf{k}']$, only depends on the angle θ between **k** and **k**′. Therefore, ϕ may be chosen using a uniform random number between 0 and 2π (i.e., $2\pi r$), while θ is chosen according to the angular dependence for scattering arising from $\Gamma_j[n,\mathbf{k};m,\mathbf{k}']$. If the probability for scattering into a certain angle $P(\theta)d\theta$ is integrable, then random angles satisfying this probability density may be generated from a uniform distribution between 0 and 1 through a prescribed recipe. Otherwise, a rejection technique may be used to select random angles according to $P(\theta)$. Common scattering mechanisms that contribute to transport are summarized in Figure 51.6. The corresponding scattering rate expressions for general nonparabolic bands are summarized in Table 51.2 [23–25].

A general Monte Carlo code is then developed as follows: First, a subroutine is typically called that contains all material and scattering rates parameters for the scattering mechanisms included in the theoretical model. After the material and run parameters are read, in the first step of the Monte Carlo simulation procedure, it is necessary to construct scattering tables for the Γ, L, and X valleys (for GaAs as a prototypical example) that initialize a series of events that are summarized in Figure 51.7. Then, for each energy, the cumulative scattering rates for each valley are stored in separate look-up tables and renormalized according to the maximum scattering rate (including self-scattering) that occurs over the range of energies stored.

Having constructed the scattering table and after renormalizing the table, examples of which are given for GaAs in Figures 51.8 and 51.9 for the Γ, L, and X valleys, the next step is to initialize the carrier's wavevector and energy and the initial free-flight time. This is accomplished by calling the initialization subroutine. Energy and wavevector histograms of the initial carrier energy and the components of the wavevector along the x-, y-, and z-axes are shown in Figure 51.10. For good statistics, the number of particles simulated is typically taken to be 10,000, and one can see the statistical fluctuation of these average quantities associated with the finite number of particles. Notice that the variance is inversely proportional to the square-root of the number of particles. Also note that the initial y-component for the wavevector is symmetric around the y-axis, which means that the average wavevector along the y-axis is zero, which should be expected since the electric field along the y-component is zero at $t = 0$. Identical distributions have been obtained for the x- and for the z-components of the wavevector. Also note that the energy distribution has the Maxwell–Boltzmann form, as it should be expected. One can also estimate from this graph that the average energy of the carriers is on the order of $(3/2)k_BT$ (k_B is the Boltzmann constant and T the temperature).

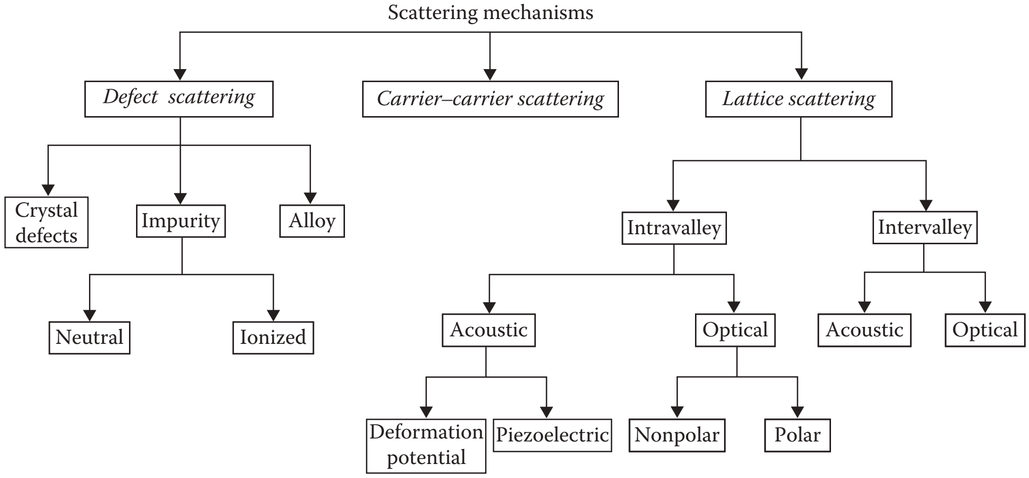


FIGURE 51.6 Scattering mechanisms in a typical semiconductor.

TABLE 51.2 Scattering Rate Expressions $\Gamma(k) = W(k) = W(E)$ for Electrons for General Nonparabolic Energy Bands

---

1. *Ionized impurity scattering:* Scattering from ionized Coulomb charges in the Brooks–Herring approach.

$$W(E) = \left(\frac{\sqrt{2}e^4 N_I m_{\mathrm{d}}^{\frac{3}{2}}}{\pi\varepsilon_{\mathrm{s}}^2 ?^4}\right) \times \left(\sqrt{E(1+\alpha E)} \times (1+2\alpha E)\right) \times \left(\frac{1}{q_{\mathrm{D}}^2\left(q_{\mathrm{D}}^2 + \frac{8m_{\mathrm{d}}E(1+\alpha E)}{\hbar^2}\right)}\right) \quad q_{\mathrm{D}} = \sqrt{\frac{e^2(n+p)}{\varepsilon_{\mathrm{s}}K_{\mathrm{B}}T_{\mathrm{L}}}},$$

where n and p are the electron and hole concentrations, respectively, k_{B} is the Boltzmann constant, m_{d} is the density of states mass, T_{L} is the lattice temperature, ε_{s} is the permittivity of the material, E is the energy, and α is the nonparabolicity factor.

2. *Alloy disorder scattering* (e.g., $\mathrm{Al}_x\mathrm{Ga}_{1\text{-}x}\mathrm{As}$): Scattering due to the presence of a random alloy.

$$W(E) = \left(\frac{x(1-x)a^3}{\pi}\right) \times \left(\frac{D_{\mathrm{alloy}}^2 d}{\hbar^4}\right) \times m_{\mathrm{d}}\sqrt{2m_{\mathrm{d}}E(1+\alpha E)} \times (1+2\alpha E),$$

where x is the mole fraction, d the lattice disorder ($0 \le d \le 1$), a the lattice constant, and D_{alloy} the alloy disorder scattering potential.

3. *Dislocation scattering* (e.g., GaN): Scattering due to crystal defects/dislocations (Coulomb-type interaction).

$$W(E) = \left(\frac{N_{\mathrm{dis}}m_{\mathrm{d}}e^4}{4\hbar^3\varepsilon^2c^2}\right) \times \left(\frac{\lambda^4}{\left(1+\frac{8\lambda^2 m_{\mathrm{d}}E(1+\alpha E)}{\hbar^2}\right)^{\frac{3}{2}}}\right) \times \left(1 + \frac{4\lambda^2 m_{\mathrm{d}}E(1+\alpha E)}{\hbar^2}\right) \times (1+2\alpha E) \quad \lambda = \sqrt{\frac{\varepsilon K_B T_L}{e^2 n'}},$$

where n' is the effective screening concentration and N_{dis} the line dislocation density.

4. *Acoustic deformation potential scattering:* Describes fluctuations in the band structure due to deformation of the unit cell because of the strain due to the different amounts of displacement of the atoms (in the same direction).

$$W(E) = \left(\frac{2\pi D_{\mathrm{ac}}^2 K_{\mathrm{B}}T_{\mathrm{L}}}{\hbar C_{\mathrm{l}}}\right) \times \left(\frac{(2m_{\mathrm{d}})^{\frac{3}{2}}\sqrt{E(1+\alpha E)}}{4\pi^2\hbar^3}\right) \times (1+2\alpha E),$$

where D_{ac} is the acoustic deformation potential constant and C_{l} is the sound velocity.

5. *Piezoelectric scattering:* Dipole-type interaction in zinc-blende materials in which there is charge transfer between the two atoms in the unit cell. The optical modes of vibration (vibration of the atoms in opposite direction) directly affect the magnitude of the dipole moment, which in turn leads to scattering potential.

$$W(\mathrm{E}) = \left(\frac{m_{\mathrm{d}}^{\frac{1}{2}}\mathrm{K_B T_L}}{4\sqrt{2\pi\rho\mathrm{V}_{\mathrm{s}}^2}\hbar^2}\right) \times \left(\frac{1+2\alpha E}{\sqrt{E(1+\alpha E)}}\right) \times \left(\frac{ee_{\mathrm{pz}}}{\varepsilon_\infty}\right)^2 \times \ln\left(1 + \frac{8m_{\mathrm{d}}E(1+\alpha E)}{\hbar^2 q_{\mathrm{D}}^2}\right) \quad q_{\mathrm{D}} = \sqrt{\frac{e^2(n+p)}{\varepsilon_{\mathrm{s}}K_{\mathrm{B}}T_{\mathrm{L}}}},$$

where ρ is the crystal density, v_{s} sound velocity, e_{pz} the piezoelectric constant, and ε_∞ the high-frequency dielectric constant.

6. *Polar optical phonon (POP) scattering:* In polar materials (where there is transfer of charge from one atom to the other, which in turn, leads to a dipole moment), optical modes of vibration cause direct modification of the dipole moment, which then leads to polar optical phonon scattering which is a Coulomb-type interaction.

$$W(E) = \left(\frac{\sqrt{m_{\mathrm{d}}}e^2 W_{\mathrm{LO}}}{4\sqrt{2}\pi\hbar\varepsilon_{\mathrm{p}}}\right) \times \left(N_0 + \tfrac{1}{2} \mp \tfrac{1}{2}\right) \times \left(\frac{1+2\alpha E_k'}{\gamma_k}\right) \times F\left(E_k, E_k'\right)$$

$$N_0 = \frac{1}{\frac{{}_{\hbar}\mathrm{W}_{\mathrm{LO}}}{e^{K_B T_{L-1}}}} \qquad \varepsilon_{\mathrm{p}} = \frac{1}{\frac{1}{\varepsilon_{\mathrm{high}}} - \frac{1}{\varepsilon_{\mathrm{low}}}} \qquad F(E_k,\ E_k') = \ln\left(\mathrm{mod}\left(\frac{\sqrt{\gamma_k}+\sqrt{\gamma_{k\prime}}}{\sqrt{\gamma_k}-\sqrt{\gamma_{k\prime}}}\right)\right)$$

$$\gamma_K = E_k(1+\alpha)E_k \qquad E_k = E_k \pm \hbar W_{\mathrm{LO}},$$

where W_{LO} is the optical phonon energy and the indices high/low stand for high frequency/static dielectric constants.

7. *Intervalley phonon scattering:* Scattering between equivalent or nonequivalent valleys.

$$W(E) = \left(\frac{\pi D_{ij}^2 Z_j}{\rho W_{ij}}\right) \times \left(n(W_{ij})\tfrac{1}{2} \mp \tfrac{1}{2}\right) \times \left(\frac{(2m_{\mathrm{d}})^{\frac{3}{2}}\sqrt{E_f(1+\alpha E_f)}}{4\pi^2\hbar^3}\right) \times (1+2\gamma E_f) \qquad E_f = F \pm \hbar w_{ij} - \Delta E_{ij},$$

where D_{ij} is the coupling constant, Z_j the number of final valleys, ρ the crystal density, and w_{ij} the energy of the phonon involved in the process.

Sources: M. Lundstrom, Fundamentals of Carrier Transport, Cambridge University Press, London, 2000; B. K. Ridley, *Quantum Processes in Semiconductors, Oxford University Press, 1982; D. Vasileska nanoHUB.org page: nanohub.org/groups/dragica_vasileska/semiclassical* For more details on the derivation of these formulas, please refer to Refs. [26–28].

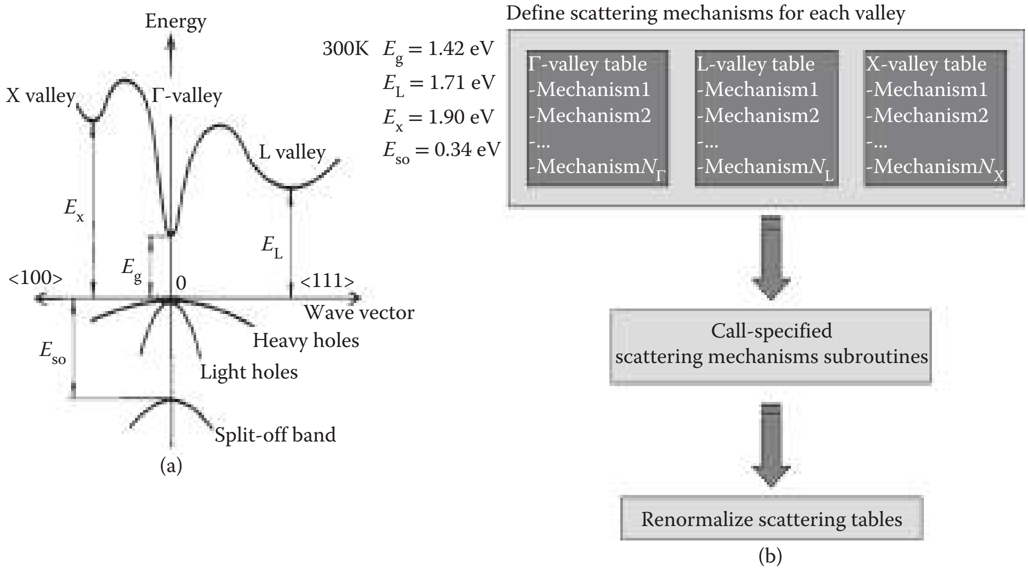

Energy
X valley
Γ-valley
L valley
E_x
E_L
E_g
0
<100>
<111>
Wave vector
Heavy holes
Light holes
E_{so}
Split-off band
(a)
300K $E_g = 1.42$ eV
$E_L = 1.71$ eV
$E_x = 1.90$ eV
$E_{so} = 0.34$ eV
Define scattering mechanisms for each valley
Γ-valley table
-Mechanism1
-Mechanism2
-...
-Mechanism$N_Γ$
L-valley table
-Mechanism1
-Mechanism2
-...
-MechanismN_L
X-valley table
-Mechanism1
-Mechanism2
-...
-MechanismN_X
Call-specified scattering mechanisms subroutines
Renormalize scattering tables
(b)

FIGURE 51.7 (a) Schematics of a bandstructure for GaAs. (b) Procedure for the creation of the scattering tables for electron transport.

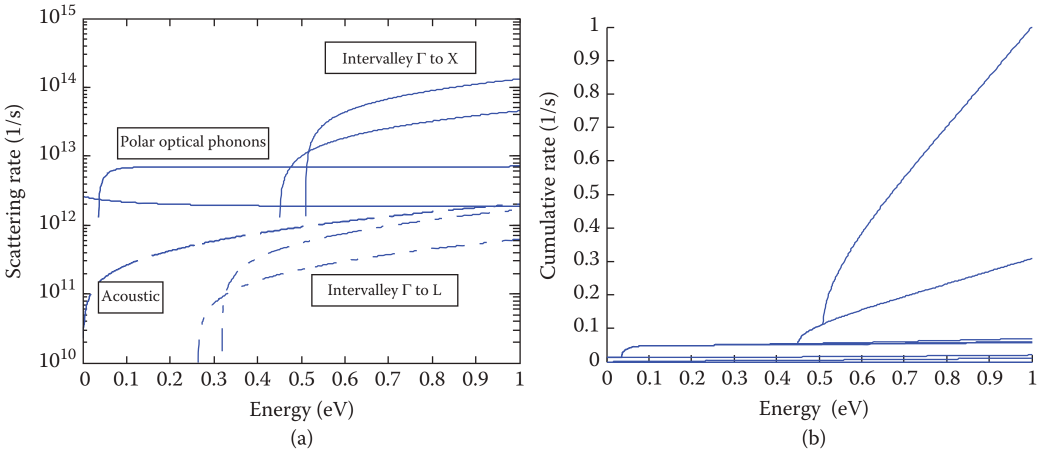


FIGURE 51.8 (a) Scattering rates for the Γ valley. For simplicity, we have omitted Coulomb scattering in these calculations. In (a) the dashed line corresponds to the acoustic phonon scattering rate, solid lines correspond to polar optical phonon scattering (absorption and emission), and the dashed-dotted line corresponds to intervalley scattering from Γ to L valleys. Since the L valley is along the [111] direction, there are eight equivalent directions, and since these valleys are shared, there are a total of four equivalent L valleys. The dotted line corresponds to scattering from Γ to X valleys. The X valleys are in the [100] direction and since there are six equivalent [100] directions and the valleys are shared between Brillouin zones, there are three equivalent X valleys. (b) Normalized cumulative scattering table for the Γ valley. Everything above the top line up to Γ = 1 is self-scattering so it is advisable when checking the scattering mechanisms to first check whether the scattering mechanism chosen is self-scattering or not. This is in particular important for energies below 0.5 eV for this particular scattering table when the Γ to X intervalley scattering (absorption and emission) takes over.

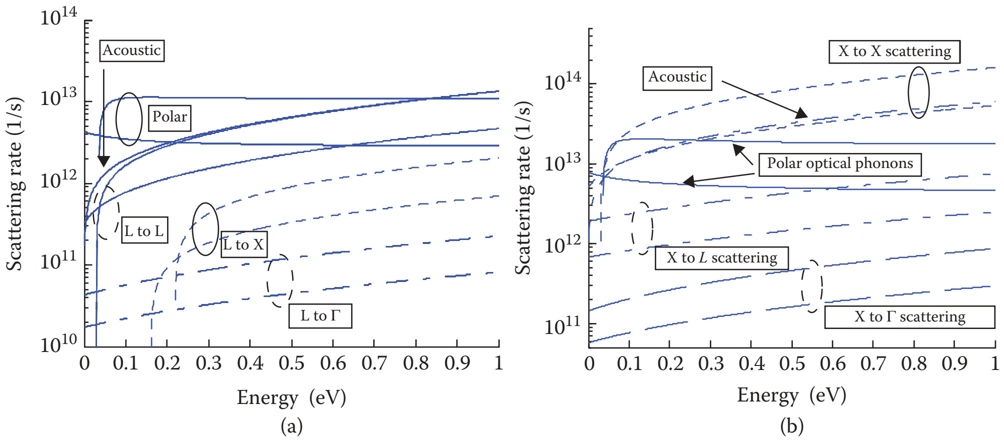


FIGURE 51.9 Scattering rates for the L (a) and X (b) valleys used to create the corresponding normalized scattering tables (not shown here).

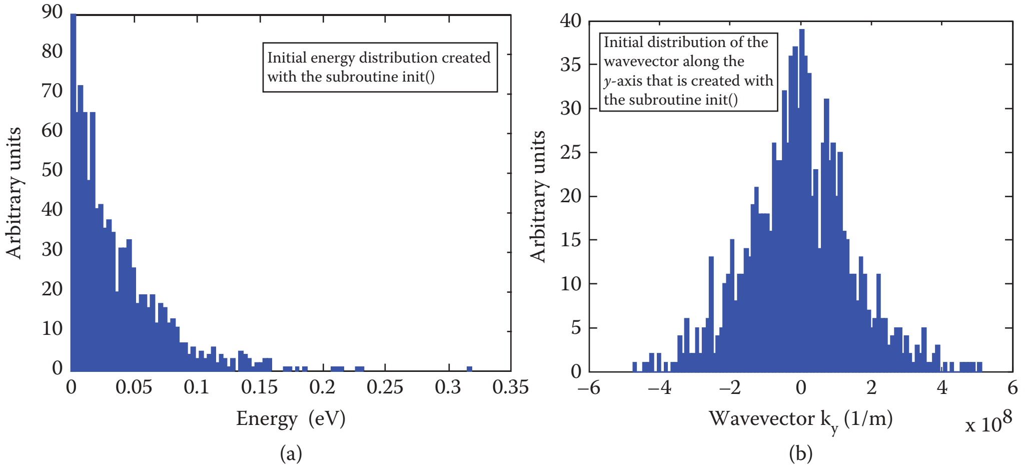


FIGURE 51.10 Initial carrier distribution for an ensemble of 10,000 particles. (a) Energy distribution. (b) Distribution of wavevector k_y.

When the initialization process is finished, the main free-flight-scatter procedure takes place until the completion of the simulation time. There are two components in this process: first, the carriers accelerate freely due to the electric field, accomplished by calling the **drift()** subroutine (which implements Equation 51.2), and then their free-flights are interrupted by random scattering events that are managed by the **scatter_carrier()** subroutine. The flowchart for performing the free-flight-scatter process within one time-step Δt is shown diagrammatically in Figure 51.11.

In the **scatter_carrier()** subroutine from Figure 51.11, first the scattering mechanism terminating the free flight is chosen, to which certain attributes are associated such as the change in energy after scattering. For inelastic scattering processes, we have the change in energy due to emission or absorption of phonons, for example. Also, the nature of the scattering process is identified as isotropic or anisotropic. Note that when performing acoustic phonon and intervalley scattering for GaAs, both of which are isotropic scattering processes, no coordinate system transformation is needed to determine the final wavevector after scattering. Because polar optical phonon and Coulomb scattering mechanisms are anisotropic, it is necessary to do a rotation of the coordinate system, scatter the carrier in the rotated system, and then perform

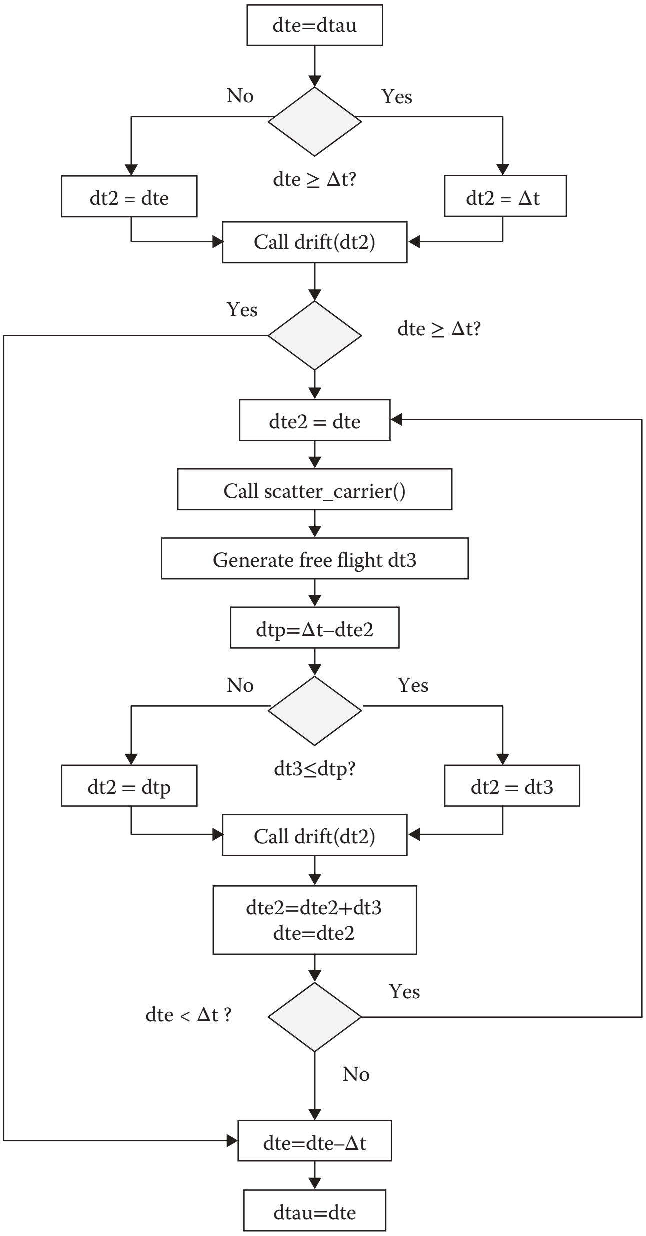

dte=dtau
No
Yes
dte ≥ Δt?
dt2 = dte
dt2 = Δt
Call drift(dt2)
Yes
dte ≥ Δt?
dte2 = dte
Call scatter_carrier()
Generate free flight dt3
dtp=Δt–dte2
No
Yes
dt3≤dtp?
dt2 = dtp
dt2 = dt3
Call drift(dt2)
dte2=dte2+dt3
dte=dte2
Yes
dte < Δt ?
No
dte=dte–Δt
dtau=dte

FIGURE 51.11 Free-flight-scatter procedure within one time-step. Time variable dtau is the remaining free-flight time from previous time-step Δt.

inverse coordinate transformation. This procedure is needed because it is much easier to determine final carrier momentum in the rotated coordinate system in which the initial wavevector k is aligned with the z-axis. For this case, one can calculate that the final polar angle of scattering with polar optical phonons for parabolic bands in the rotated coordinate system as

$$\cos\theta = \frac{(1+\xi)-(1+2\xi)^r}{\xi}, \quad \xi = \frac{2\sqrt{E_k\left(E_k \pm \hbar\omega_0\right)}}{\left(\sqrt{E_k}-\sqrt{E_k \pm \hbar\omega_0}\right)^2}, \tag{51.10}$$

where E_k is the carrier energy, $\hbar\omega_0$ the polar optical phonon energy, and r a random number uniformly distributed between 0 and 1. The final angle for scattering with ionized impurities (Coulomb scattering) and for parabolic bands is

$$\cos\theta = 1 - \frac{2r}{1 + 4k^2L_{\mathrm{D}}^2(1-r)}, \tag{51.11}$$

where k is the carrier wavevector and L_{D} is the Debye screening length. The azimuthal angle for both scattering processes is simply calculated using $\varphi = 2\pi r$. The importance of properly calculating the polar angle θ after scattering to describe small angles of deflection in the case of Coulomb or polar optical phonon scattering is illustrated in Figure 51.13. The histogram of the polar angle (from 0 to $\pi = 3.141592654$) after scattering for electron–polar optical phonon scattering is presented in Figure 51.12, from where we can clearly see the preference for small angle deflections that are characteristic for any Coulomb-type interaction (polar optical phonon is in fact electron–dipole interaction). Graphical representation of the determination of the final angle after scattering for both isotropic and anisotropic scattering processes is shown in Figure 51.13.

The direct technique described above can be applied when the integrals describing $\cos\theta$ can be analytically calculated. For most cases of interest, the integral cannot be easily inverted. In these cases, a rejection technique may be employed. The procedure for the rejection technique goes as follows:

- Choose a maximum value C, such that $C > f(x)$ for all x in the interval (a, b).
- Choose pairs of random numbers, one between a and b $\left(x_1 = a + r_1(b-a)\right)$ and another one $f_1 = r_1'C$ between 0 and C, where r_1 and r_1' are random numbers uniformly distributed between 0 and 1.
- If $f_1 \le f(x_1)$, then the number x_1 is accepted as a suitable value, otherwise it is rejected.

The three steps described above are schematically shown in Figure 51.14. For $x = x_1$, r_1C is larger than $f(x_1)$ and in this case, if this represents the final polar angle for scattering, this angle is rejected and a new sequence of two random numbers is generated to determine x_2 and r_2C. In this second case, $f(x_2) > r_2C$ and the polar angle $\theta = x_2$ is selected (for polar angle selection $a = 0$ and $b = \pi$).

After the simulation is completed, typical results to check are the velocity–time, the energy–time, and the valley occupation versus time characteristics, such as those shown in Figure 51.15, where the velocity–time characteristics for applied electric fields ranging from 0.5 to 7 kV/cm, with an electric field increment

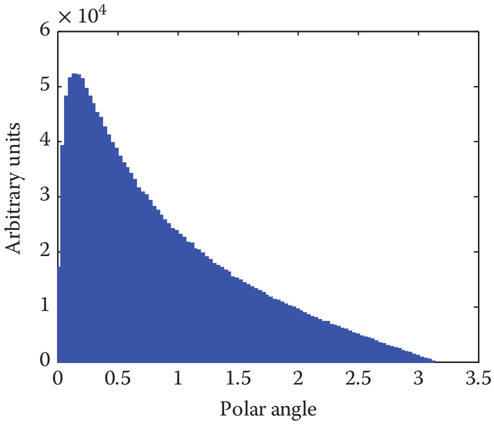
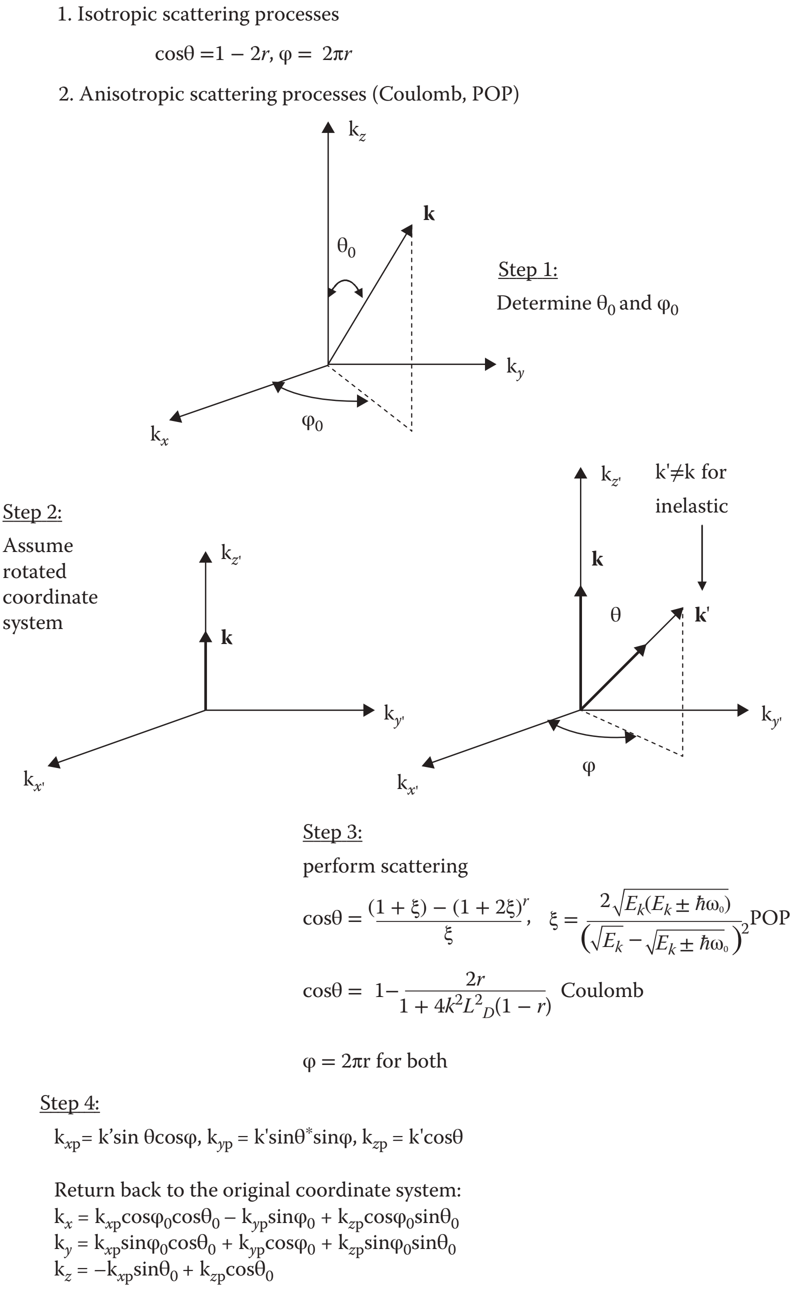


FIGURE 51.12 Histogram of the polar angle for electron–polar optical phonon scattering.



1. Isotropic scattering processes

$$\cos\theta = 1 - 2r,\ \varphi = 2\pi r$$

2. Anisotropic scattering processes (Coulomb, POP)

Step 1:
Determine θ_0 and φ_0

Step 2:
Assume rotated coordinate system

Step 3:
perform scattering

$$\cos\theta = \frac{(1+\xi) - (1+2\xi)^r}{\xi},\quad \xi = \frac{2\sqrt{E_k(E_k \pm \hbar\omega_0)}}{\left(\sqrt{E_k} - \sqrt{E_k \pm \hbar\omega_0}\right)^2}\ \text{POP}$$

$$\cos\theta = 1 - \frac{2r}{1 + 4k^2L^2{}_D(1-r)}\quad \text{Coulomb}$$

$\varphi = 2\pi r$ for both

Step 4:
$k_{xp} = k'\sin\theta\cos\varphi,\ k_{yp} = k'\sin\theta^{*}\sin\varphi,\ k_{zp} = k'\cos\theta$

Return back to the original coordinate system:
$k_x = k_{xp}\cos\varphi_0\cos\theta_0 - k_{yp}\sin\varphi_0 + k_{zp}\cos\varphi_0\sin\theta_0$
$k_y = k_{xp}\sin\varphi_0\cos\theta_0 + k_{yp}\cos\varphi_0 + k_{zp}\sin\varphi_0\sin\theta_0$
$k_z = -k_{xp}\sin\theta_0 + k_{zp}\cos\theta_0$

FIGURE 51.13 Description of final angle selection for isotropic and anisotropic scattering processes using the direct technique.

of 0.5 kV/cm, are shown. These clearly demonstrate that after a transient phase, the system reaches a stationary steady state, after which time we can start taking averages for calculating steady-state quantities.

From the results shown in Figure 51.15, one can see that steady state is achieved for larger time intervals when the electric field value is increased and the carriers are still sitting in the Γ valley. Afterward, time needed to get to steady state decreases. This trend is related to the valley repopulation and movement of the

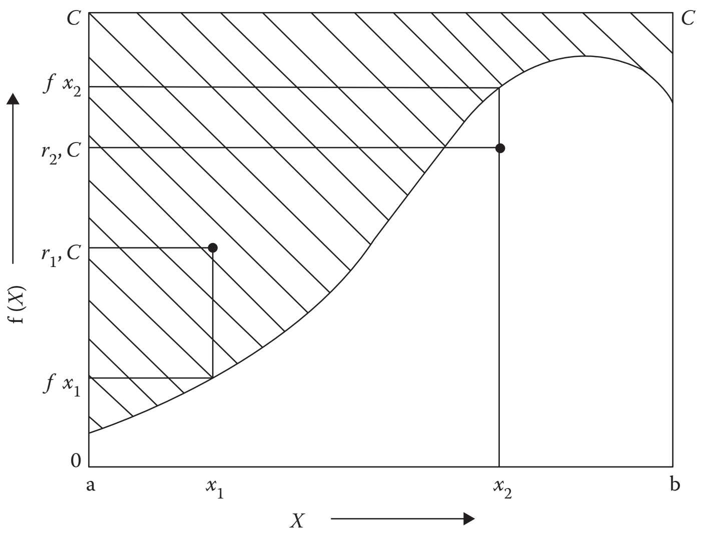


FIGURE 51.14 Schematic description of the rejection technique.

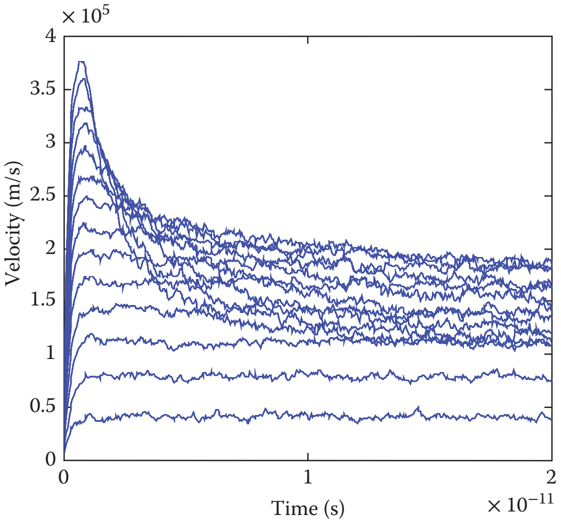


FIGURE 51.15 Time evolution of the drift velocity for electric field strengths ranging between 0.5 and 7 kV/cm, in 0.5 kV/cm increments.

carriers from Γ, into the X and finally into the L valley. The steady-state velocity-field and valley population versus electric field characteristics are shown in Figures 51.16 and 51.17, respectively. One can clearly see on the velocity–field characteristics that a low field mobility of about 8000 cm^2/V-s is correctly reproduced for GaAs without the use of any adjustable parameters.

At this point, it is advisable to check the energy and wavevector histograms (Figure 51.18) to ensure that the energy range chosen in the scattering tables is correct or not for the particular maximum electric field strength being considered, which gives the worst case scenario. Since, as already noted, we apply the electric field in the y-direction, for comparative purposes, we plot the histograms of the x-component of the wavevector, y-component of the wavevector, and the histogram of the final carrier energy distribution

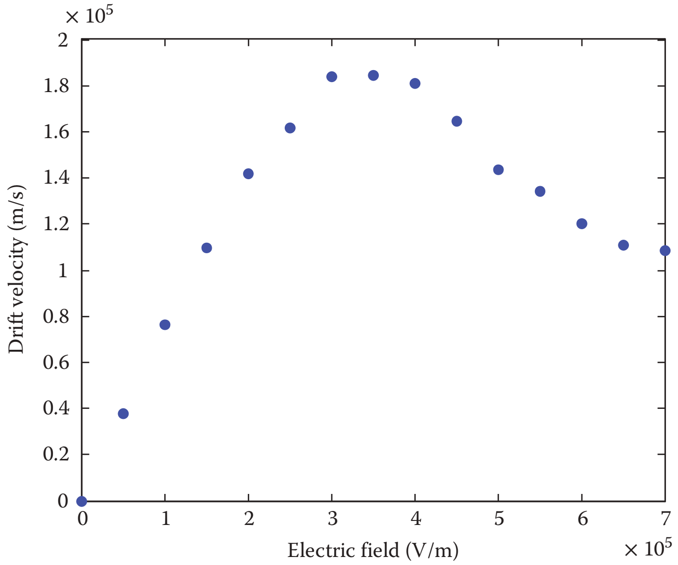


FIGURE 51.16 Steady-state drift velocity versus electric field.

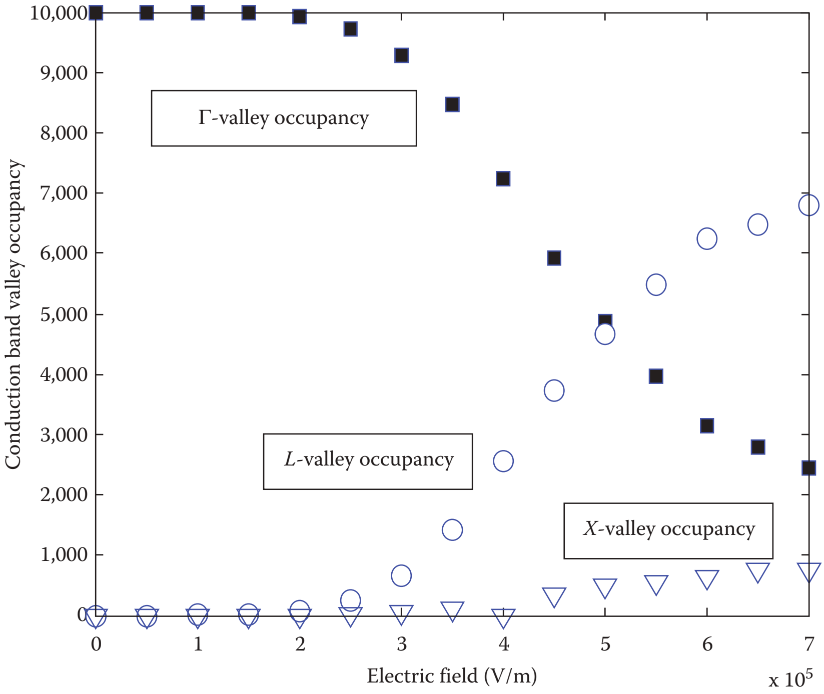


FIGURE 51.17 Different valley occupancy versus electric field.

for which a drifted Maxwellian form is evident. Since there is no field applied in the x-direction, we see that the average wavevector in the x-direction is 0. Due to the application of the field in the y-direction, there is a finite positive shift in the y-component of the velocity, which is yet another signature for the displaced Maxwellian form of the energy distribution in the bottom histogram.

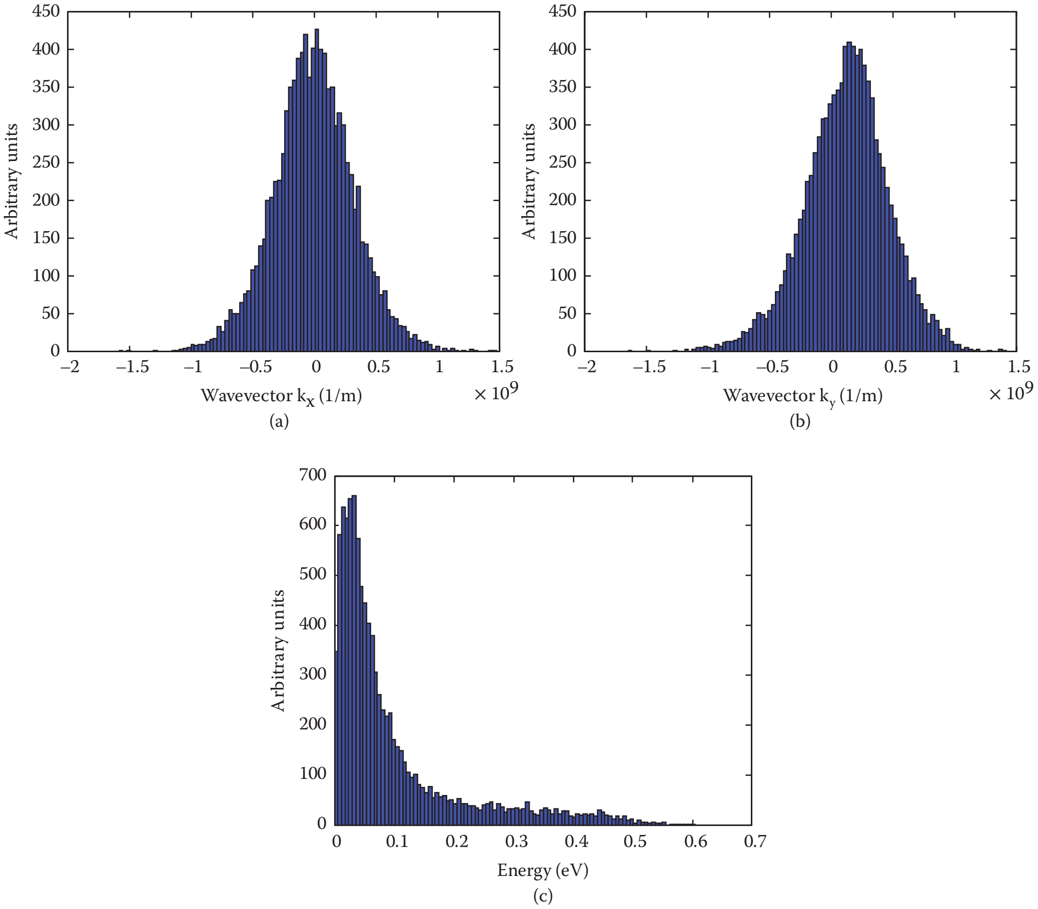


FIGURE 51.18 (a) Histogram of the *x*-component of the wavevector. (b) Histogram of the *y*-component of the wavevector. (c) Histogram of the carrier energy. Applied electric field is 7 kV/cm.

51.3 Particle-Based Device Simulation

In Section 51.2, we introduced the numerical solution of the BTE using Monte Carlo method. Within a device, both the transport kernel and the field solver are coupled to each other. In Poisson's equation, the field associated with the potential is the driving force accelerating particles in the Monte Carlo phase, for example, while the distribution of mobile (both electrons and holes) and fixed charges (e.g., donors and acceptors) provides the source of the electric field. Below, we give a brief description of the Monte Carlo particle-based device simulator development.

Within the particle-based EMC method, with its time-marching algorithm, Poisson's equation may be decoupled from the BTE over a suitably small time step (typically less than the inverse plasma frequency corresponding to the highest carrier density in the device). Over this time interval, carriers accelerate according to the frozen field profile from the previous time-step solution of Poisson's equation, and then Poisson's equation is solved at the end of the time interval with the frozen configuration of charges arising from the Monte Carlo phase [6,26]. Note that Poisson's equation is solved on a mesh, whereas the solution of charge motion using EMC occurs over a continuous range of coordinate space in terms of the particle position. Therefore, a particle-mesh (PM) coupling is needed for both the charge assignment and the force interpolation. The PM coupling is broken into four steps: (1) assign particle charge to the mesh, (2) solve the Poisson equation on the mesh, (3) calculate the mesh-defined forces, and (4) interpolate to find forces on the particle.

The motion in real space of particles under the influence of electric fields is somewhat more complicated due to the band structure. The velocity of a particle in real space is related to the E-$\mathbf{k}$ dispersion relation defining the bandstructure as

$$\begin{aligned}\mathbf{v}(t) &= \frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \frac{1}{\hbar}\nabla_k E(\mathbf{k}(t)) \\ \frac{\mathrm{d}\mathbf{k}}{\mathrm{d}t} &= \frac{q\mathbf{E}(\mathbf{r})}{\hbar}\end{aligned}, \tag{51.12}$$

where the rate of change of the crystal momentum is related to the local electric field acting on the particle through the acceleration theorem expressed by the second equation. In turn, the change in crystal momentum, $\mathbf{k}(t)$, is related to the velocity through the gradient of E with respect to $\mathbf{k}$. If one has to use the full bandstructure of the semiconductor, then integration of these equations to find $\mathbf{r}(t)$ is only possible numerically, using, for example, a Runge–Kutta algorithm [27].

To simulate the steady-state behavior of a device, the system must be initialized in some initial condition, with the desired potentials applied to the contacts, and then the simulation proceeds in a time stepping manner until steady state is reached. This process may take several picoseconds of simulation time, and consequently several thousand time-steps based on the usual time increments required for stability. Clearly, the closer the initial state of the system is to the steady-state solution, the quicker the convergence. If one is, for example, simulating the first bias point for a transistor simulation, and has no *a priori* knowledge of the solution, a common starting point for the initial guess is to start out with charge neutrality, i.e., to assign particles randomly according to the doping profile in the device and based on the super-particle charge assignment of the particles, so that initially the system is charge neutral on the average. (For 2D device simulation, one should keep in mind that each particle actually represents a rod of charge into the third dimension.) Subsequent simulations at the same device at different bias conditions can use the steady-state solution at the previous bias point as a good initial guess. After assigning charges randomly in the device structure, charge is then assigned to each mesh point using the nearest-grid-point (NGP), cloud-in-cell (CIC), or nearest-element-cell (NEC) PM coupling methods, and Poisson's equation is solved. The forces are then interpolated on the grid, and particles are accelerated over the next time-step. A flowchart of a typical Monte Carlo device simulation is shown in Figure 51.19.

As the simulation evolves, charge will flow in and out of the contacts, and depletion regions internal to the device will form until steady state is reached. The charge passing through the contacts at each time step can be tabulated, and a plot of the cumulative charge as a function of time gives the steady-state current. Figure 51.20 shows the particle distribution in three dimension of a metal-semiconductor field-effect transistor (MESFET), where the dots indicate the individual simulated particles for two different gate biases. Here, the heavily doped MESFET region (shown by the inner box) is surrounded by semi-insulating GaAs forming the rest of the simulation domain. Figure 51.20a corresponds to no net gate bias (i.e., the gate is positively biased to overcome the built-in potential of the Schottky contact), while Figure 51.20b corresponds to a net negative bias applied to the gate, such that the channel is close to pinch-off. One can see the evident depletion of carriers under the gate under the latter conditions.

51.3.1 Calculation of the Current

The device output current can be determined using two different yet consistent methods. First, by keeping track of the charges entering and exiting each terminal/contact, the net number of charges over a period of the simulation can be used to calculate the terminal current. The net charge crossing a terminal boundary

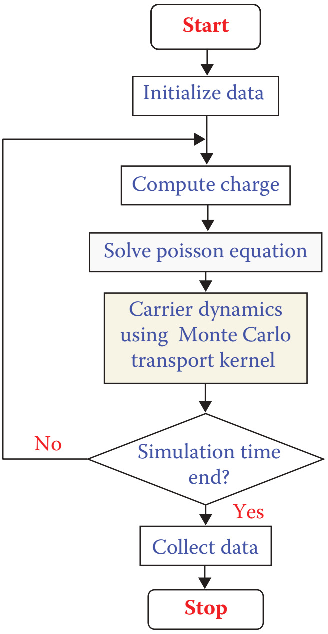
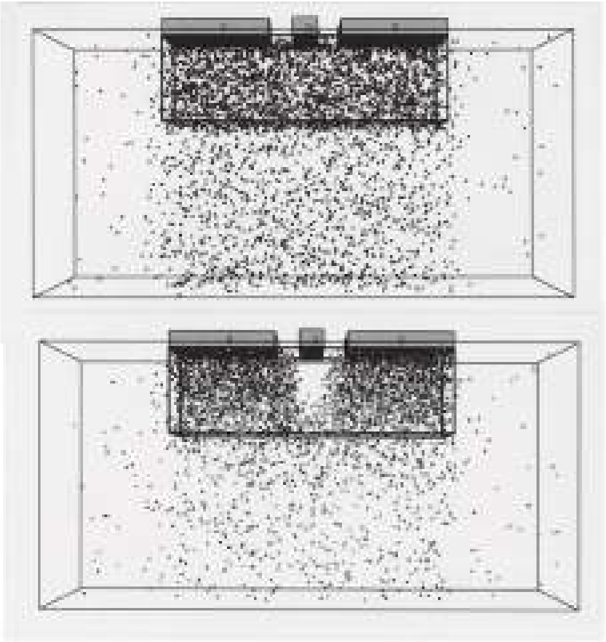

Start
Initialize data
Compute charge
Solve poisson equation
Carrier dynamics using Monte Carlo transport kernel
Simulation time end?
No
Yes
Collect data
Stop

FIGURE 51.19 Flowchart of a typical particle-based device simulation.

(a)



(b)

FIGURE 51.20 Example of the particle distribution in an MESFET structure simulated in three dimension using an EMC approach. (a) The device with zero gate voltage applied and (b) the device with a negative gate voltage applied, close to pinch-off.

is determined by

$$Q(t) = e(n_{\text{abs}}(t) - n_{\text{injec}}(t)) + \varepsilon \int E_y(x, t)\mathrm{d}y, \tag{51.13}$$

where n_{abs} is the number of particles that are absorbed by the contact (exit), n_{injec} the number of particles that have been injected at the contact, and E_y the vertical field at the contact. The second term in Equation 51.13 on the right-hand side is used to account for the displacement current due to the changing field at the contact. Equation 51.13 assumes the contact is at the top of the device and that the fields in the x- and z- directions are negligible. The charge e in Equation 51.13 should be multiplied by the particle charge if it is not unity. The slope of $Q(t)$versus time gives a measure of the terminal current. In steady state, the current can be found by

$$I = \frac{\mathrm{d}Q(t)}{\mathrm{d}t} = \frac{e(n_{\text{net}})}{\Delta t}, \tag{51.14}$$

where $n_{\text{net}} = n_{\text{abs}} - n_{\text{injec}}$ is the net number of particles exiting the contact over a fixed period of time Δt.

In a second method, the sum of the electron velocities in a portion of the channel region of the device is used to calculate the current. The electron current density through a cross-section of the device is given by

$$J = env_{\text{d}}, \tag{51.15}$$

where v_{d} is the average electron drift velocity and n the carrier concentration. If there are a total of N particles in a differential volume, $\mathrm{d}V = \mathrm{d}L \cdot \mathrm{d}A$,the current found by integrating Equation 51.15 over the cross-sectional area, $\mathrm{d}A$, is

$$I = \frac{eNv_{\text{d}}}{\mathrm{d}L}, \text{ or } I = \frac{e}{\mathrm{d}L}\sum_{i=1}^{N} v_x(i), \tag{51.16}$$

where $v_x(i)$ is the velocity along the channel of the ith electron. The device is divided into several sections along the x-axis, and the number of electrons and their corresponding velocity is added for each section after each free flight. The total x-velocity in each section is then averaged over several time-steps to determine the current for that section. Total device current can be determined from the average of several sections, which gives a much smoother result compared to counting terminal charges. By breaking the device into sections, individual currents can be compared to verify that there is conservation of particles (constant current) throughout the device. In addition, sections near the source and drain regions may have a high y-component in their velocity and should be excluded from the current calculations. Finally, by using several sections in the channel, the average energy and velocity of electrons along the channel can be observed to ensure the proper physical characteristics. The two methods for the calculation of the current are performed, as illustrated in Figure 51.17, on a 50-nm channel length MOSFET device.

Extrapolating the slope of the curve shown in Figure 51.21a, that represents the cumulative electron charge that enters/exits the source/drain contact, leads to source/drain current of 0.5205/0.5193 mA/μm. When compared with the results shown in Figure 51.21b, it is evident that both current measurement techniques discussed in this section give current values with relative error less than 1%.

51.3.2 Ohmic Contacts

Another issue that has to be addressed in particle-based simulations is the real space boundary conditions for the particle part of the simulation. Reflecting boundary conditions are usually imposed at the artificial

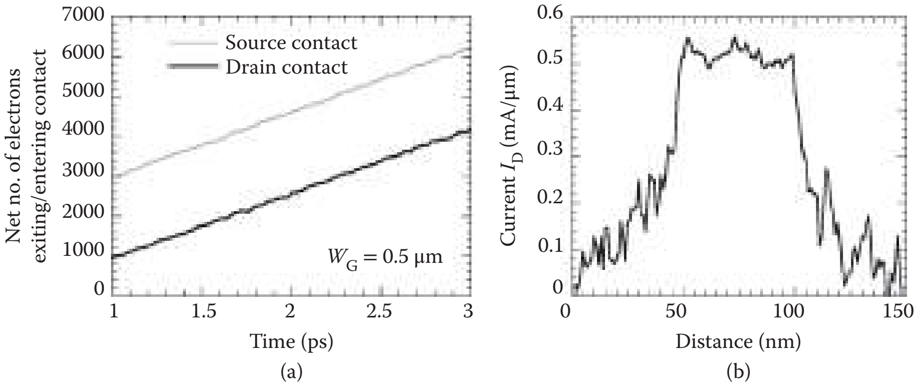


FIGURE 51.21 (a) Net charge entering/exiting the source/drain contact. (b) Average current along the channel. The gate length of the device being modeled equals 50 nm. $V_G = 1.4$ V and $V_D = 1$ V in these simulations. Width of the gate is $W = 0.5$ μm.

boundaries. As far as the ohmic contacts are concerned, they require more careful consideration because electrons crossing the source and drain contact regions contribute to the corresponding terminal current. Commonly employed models for the contacts include the following [28]:

- Electrons are injected at the opposite contact with the same energy and wavevector **k**. If the source and drain contacts are in the same plane, as in the case of MOSFET simulations, the sign of **k**, normal to the contact will change. This is an unphysical model, however [29].
- Electrons are injected at the opposite contact with a wavevector randomly selected based upon a thermal distribution. This is also an unphysical model.
- Contact regions are considered to be in thermal equilibrium. The total number of electrons in a small region near the contact are kept constant, with the number of electrons equal to the number of dopant ions in the region. This is a very good model most commonly employed in actual device simulations.
- Another method uses "reservoirs" of electrons adjacent to the contacts. Electrons naturally diffuse into the contacts from the reservoirs, which are not treated as part of the device during the solution of Poisson's equation. This approach gives results similar to the velocity weighted Maxwellian but at the expense of increased computational time due to the extra electrons simulated. It is an excellent model employed in few most sophisticated particle-based simulators.

There are also several possibilities for the choice of the distribution function—Maxwellian, displaced Maxwellian, and velocity-weighted Maxwellian [10].

51.3.3 Time-Step

As in the case of solving the drift-diffusion, hydrodynamic, or full Maxwell's equations, for a stable Monte Carlo device simulation, one has to choose the appropriate time step, Δt, and the spatial mesh size (Δx, Δy, and/or Δz). The time-step and the mesh size may correlate to each other in connection with the numerical stability. For example, as discussed in the context of solving drift-diffusion simulations, the time-step Δt must be related to the plasma frequency:

$$\omega_p = \frac{e^2 n}{\varepsilon_s m^*}, \tag{51.17}$$

where n is the carrier density. From the viewpoint of the stability criterion, Δt must be much smaller than the inverse plasma frequency. The highest carrier density specified in the device model is used to estimate

Δt. If the material is a multivalley semiconductor, the smallest effective mass to be experienced by the carriers must be used in Equation 51.17 as well. In the case of GaAs, with the doping of 5×10^{17} cm^{-3}, $\omega_p \cong 5 \times 10^{13}$; hence, Δt must be smaller than 0.02 ps.

The mesh size for the spatial resolution of the potential is dictated by the charge variations. Hence, one has to choose the mesh size to be smaller than the smallest wavelength of the charge variations. The smallest distance is approximately equal to the Debye length, given as

$$\lambda_D = \sqrt{\frac{\varepsilon_s k_B T}{e^2 n}}. \tag{51.18}$$

The highest carrier density specified in the model should be used to estimate λ_D from the stability criterion. The mesh size must be chosen to be smaller than the value given by Equation 51.18. In the case of GaAs, with the doping density of 5×10^{17} cm^{-3}, $\lambda_D \cong 6$ nm.

Based on the discussion above, the time step (Δt) and the mesh size (Δx, Δy, and/or Δz) can be specified separately. However, the Δt chosen must be checked again by calculating the distance l_{max}, defined as

$$l_{max} = \mathbf{v}_{max} \times \Delta t, \tag{51.19}$$

where $\mathbf{v}_{max}$ is the maximum carrier velocity that can be approximated by the maximum group velocity of the electrons in the semiconductor (on the order of 10^8 cm/s). Therefore, the distance l_{max} is regarded as the maximum distance the carriers can propagate during Δt. The time step chosen must be small enough so that l_{max} is smaller than the spatial mesh size chosen using Equation 51.19. This is because large Δt chosen may cause substantial change in the charge distribution, while the field distribution in the simulation is only updated every Δt.

51.3.4 PM Coupling

As mentioned earlier, the position of charge as described by the EMC algorithm is continuous, whereas Poisson's equation is solved on a mesh, hence the charge associated with the individual particles must be mapped onto the field mesh in some manner. The charge assignment and force interpolation schemes usually employed in self-consistent Monte Carlo device simulations are the NGP, CIC, and NEC schemes [30]. In the NGP scheme, the particle position is mapped into the charge density at the closest grid point to a given particle. This has the advantage of simplicity, but leads to a noisy charge distribution, which may exacerbate numerical instability. Alternately, within the CIC scheme, a finite volume is associated with each particle spanning several cells in the mesh, and a fractional portion of the charge per particle is assigned to grid points according to the relative volume of the "cloud" occupying the cell corresponding to the grid point. This method has the advantage of smoothing the charge distribution due to the discrete charges of the particle-based method, but may result in an artificial "self-force" acting on the particle, particularly if an inhomogeneous mesh is used. For the case of inhomogeneous mesh or spatially varying dielectric permittivity, the NEC scheme is the best choice. Within the NEC scheme, the charge is moved at the center of the cell and then CIC-like charge assignment to the node points is performed. The PM coupling sequence is presented in Figure 51.22.

51.4 Representative Simulation Results

51.4.1 Bulk Monte Carlo Simulations of Different Materials

In most semiconductors, in order to properly simulate high field transport, it is necessary to consider more than one conduction band valley. To calculate the drift velocity along any direction, the effective mass along

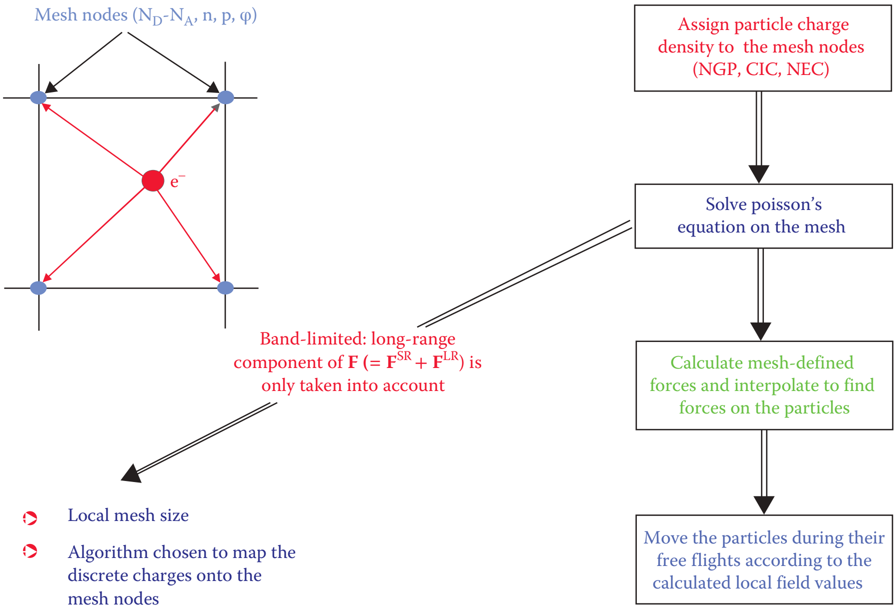

Mesh nodes (N_D-N_A, n, p, φ)
e^-
Band-limited: long-range component of **F** (= $\mathbf{F}^{SR} + \mathbf{F}^{LR}$) is only taken into account
Local mesh size
Algorithm chosen to map the discrete charges onto the mesh nodes
Assign particle charge density to the mesh nodes (NGP, CIC, NEC)
Solve poisson's equation on the mesh
Calculate mesh-defined forces and interpolate to find forces on the particles
Move the particles during their free flights according to the calculated local field values

FIGURE 51.22 Particle-mesh coupling sequence. (SR, short range; LR, long range.)

that particular direction is required. In GaAs, for example, a subvalley of the L valley lies along the [111] direction. We know the effective masses along the transverse and longitudinal directions of the subvalley but we do not know the effective mass along the [100] direction. This makes it hard to calculate the drift velocity along the [100] direction.

Assume the Monte Carlo simulation is run on the x, y, z- coordinate system where the x-direction is [100], y-direction is [010], and z-direction is [001]. Let the three perpendicular directions that describe the subvalley be $[a_1, b_1, c_1]$, $[a_2, b_2, c_2]$, and $[a_3, b_3, c_3]$. The electrons in the Monte Carlo simulation will be drifted according to the x, y, z coordinate system so there will be k_x the wavevector along [100], k_y the wavevector along [010], and k_z the wavevector along [001]. The drift velocity of the electron is first calculated along the three directions that describe the subvalley and along which we know the effective masses.

$$v_{\mathrm{d}}, [[a]_1 b_1 c_1] = \frac{\hbar k_{[[a]_1 b_1 c_1]}}{m_1\,(1+2\alpha E)} v_d, [[a]_1 b_1 c_1] = \frac{\hbar k_{[[a]_1 b_1 c_1]}}{m_1\,(1+2\alpha E)} \tag{51.20}$$

Here m_1 is the effective mass of the electron along $[a_1, b_1, c_1]$. The final expression of drift velocity along that direction is then calculated by calculating $k_{[[a]_1 b_1 c_1]} k_{[[a]_1 b_1 c_1]}$ using a simple transformation of coordinates to give [31]

$$k_{[[a]_1 b_1 c_1]} = \frac{k_x a_1}{\sqrt{\left(a_1^2+b_1^2+c_1^2\right)}} + \frac{k_y b_1}{\sqrt{\left(a_1^2+b_1^2+c_1^2\right)}} + \frac{k_z a_1}{\sqrt{\left(a_1^2+b_1^2+c_1^2\right)}}. \tag{51.21}$$

Using similar methods we get $v_{\mathrm{d}}, \left[[a]_2 b_2 c_2\right]$ and $v_{\mathrm{d}}, \left[[a]_3 b_3 c_3\right]$. The coordinates system is then transformed once again back to the x,y,z coordinate system to get the drift velocities along x,y, and z.

$$v_x = \frac{a_1 v_{\mathrm{d},[[a]_1 b_1 c_1]}}{\sqrt{\left(a_1^2 + b_1^2 + c_1^2\right)}} + \frac{a_2 v_{\mathrm{d},[[a]_2 b_2 c_2]}}{\sqrt{\left(a_2^2 + b_2^2 + c_2^2\right)}} + \frac{a_3 v_{\mathrm{d},[[a]_3 b_3 c_3]}}{\sqrt{\left(a_3^2 + b_3^2 + c_3^2\right)}}$$

$$v_y = \frac{b_1 v_{\mathrm{d},[[a]_1 b_1 c_1]}}{\sqrt{\left(a_1^2 + b_1^2 + c_1^2\right)}} + \frac{b_2 v_{\mathrm{d},[[a]_2 b_2 c_2]}}{\sqrt{\left(a_2^2 + b_2^2 + c_2^2\right)}} + \frac{b_3 v_{\mathrm{d},[[a]_3 b_3 c_3]}}{\sqrt{\left(a_3^2 + b_3^2 + c_3^2\right)}} \qquad (51.22)$$

$$v_x = \frac{c_1 v_{\mathrm{d},[[a]_1 b_1 c_1]}}{\sqrt{\left(a_1^2 + b_1^2 + c_1^2\right)}} + \frac{c_2 v_{\mathrm{d},[[a]_2 b_2 c_2]}}{\sqrt{\left(a_2^2 + b_2^2 + c_2^2\right)}} + \frac{c_3 v_{\mathrm{d},[[a]_3 b_3 c_3]}}{\sqrt{\left(a_3^2 + b_3^2 + c_3^2\right)}}$$

Also, as the three directions are mutually perpendicular we have

$$\begin{aligned} a_1 a_2 + b_1 b_2 + c_1 c_2 &= 0 \\ a_1 a_3 + b_1 b_3 + c_1 c_3 &= 0 \\ a_3 a_2 + b_3 b_2 + c_3 c_2 &= 0. \end{aligned} \qquad (51.23)$$

For N electrons in the simulation, an average drift velocity is then calculated as

$$\langle v_x \rangle = \frac{1}{N}\sum_{i=1}^{N} v_{x,i}, \quad \left\langle v_y \right\rangle = \frac{1}{N}\sum_{i=1}^{N} v_{y,i}, \quad \langle v_z \rangle = \frac{1}{N}\sum_{i=1}^{N} v_{z,i}, \qquad (51.24)$$

where each drift velocity depends on the subvalley the electron is currently in. The mobility is calculated as the slope of the velocity versus field curve for low fields. The range of "low fields" varies from material to material, but in all cases, it is the range of fields for which the velocity linearly varies with the field.

The results presented in Figure 51.23 shows the difference in velocity of electrons for different materials. For GaN, the peak velocity of 3×10^7 cm/s occurs at around 150 kV/cm so that in the electric field range shown in Figure 51.19 it is clear that for electric fields below 30 kV/cm, GaN carriers in the bulk GaN materials are still in the low-field regime.

As the drift velocities saturate at much smaller electric fields in Si and Ge, they are plotted on a separate graph (see Figure 51.24) [32].

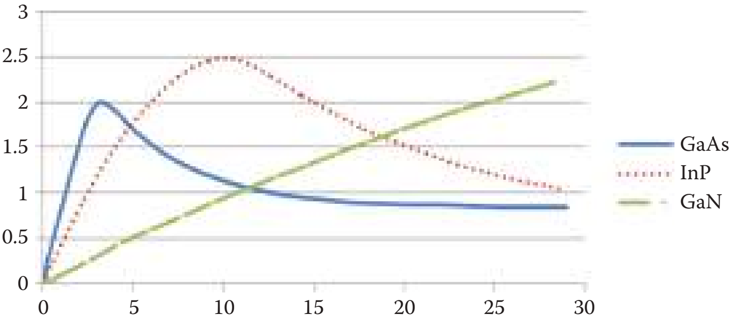


FIGURE 51.23 Drift velocity (10^7 cm/s) versus electric field (kV/cm).

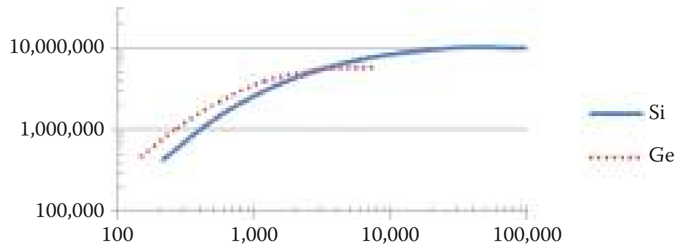


FIGURE 51.24 Electron drift velocity (cm/s) versus electric field (V/cm) in Si and Ge along the principal crystallographic directions.

51.4.2 Modeling Thermal Effects in SOI Devices Using Particle-Based Device Simulator

There are many applications of the Monte Carlo approach to device simulations that can be found in the literature. Here, we show the example of Monte Carlo device simulation to self-heating which requires coupling of the Monte Carlo device solver self-consistently with a solver for the phonon bath. Self-heating (despite the fact that supply voltages are reduced when reducing the size of the channel of nanoscale transistors) is still a problem, since there are sufficiently large electric fields in the smallest devices that can accelerate the carriers to average energies of 0.5–0.8 eV. These hot carriers then interact with the lattice and transfer their energy to the phonon bath (both acoustic and optical), thus heating the lattice and increasing the lattice vibrations (scattering). This leads to a negative feedback on the carrier drift velocity or mobility (if transport is scattering dominated). Thus, we may conclude that when transistors are in the ON-state, there is a strong interaction of the electrons with the lattice on a very short time scale (in particular with the optical phonons). The optical phonons in a longer time scale decay into acoustic phonons through a harmonic multiple phonon processes, thus removing the heat from the hot spot. Therefore, to properly treat the operation of nanoscale devices it is necessary to, at least, consider the electrons within the BTE picture and, in the lowest approximation, via energy balance equations including the acoustic and the optical phonon baths separately. This is what the Arizona State University (ASU) group has implemented couple of years ago [33]. Our ultimate goal is to have coupled the developed phonon BTE solver [34] with the electron BTE solver in a self-consistent manner.

51.4.2.1 Theoretical Model and Computation Details

A schematic description of our 2D/3D electrothermal simulator is shown in Figure 51.25. It is important to note that the EMC transport kernel for the electrons is in the loop for the solution of the energy balance equations for the acoustic and optical phonon temperatures. The criterion for convergence of the whole self-consistent loop (measured in terms of Gummel cycles) is convergence in the current to the third significant digit.

Between iterations, there are variables that are being exchanged between the electron and phonon solvers. Namely, after running the EMC for the electrons in the device for about 5–10 ps for 2D device analysis and 10 ps for the 3D nanowire simulations, the electron transport solver passes to the phonon energy balance solver the information about the spatial variation of the electron density, spatial variation of the average drift velocity, and spatial variation of the average electron temperature, where an assumption is made that the drift energy is much smaller than the thermal energy. These variables are then used in the solution of the energy balance equations for the acoustic and optical phonons to get updated values for the spatial variation of the optical phonon and acoustic phonon temperatures, which are then used as inputs, in the choice of the scattering table for the electrons (proper scattering table is chosen with the acoustic and optical phonon temperatures at the NGP). Then, when the proper table is selected, based on the energy of the electron and a selection of a random number, the scattering mechanism is selected.

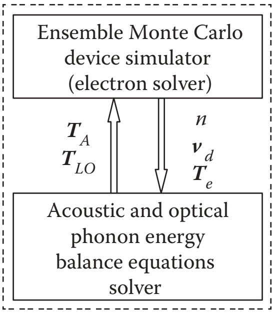

Ensemble Monte Carlo device simulator (electron solver)
T_A
T_{LO}
n
v_d
T_e
Acoustic and optical phonon energy balance equations solver

FIGURE 51.25 Exchange of variables between the two kernels. The electron density (n), electron drift velocity (v_d), and electron temperature (T_e), obtained from the electron solver, are input variables for the phonon kernel. The output variables of the phonon solver are acoustic and optical phonon temperature profiles (T_A and T_{LO}, respectively). They enter in the beginning of the free-flight-scattering phase through phonon temperature-dependent scattering tables.

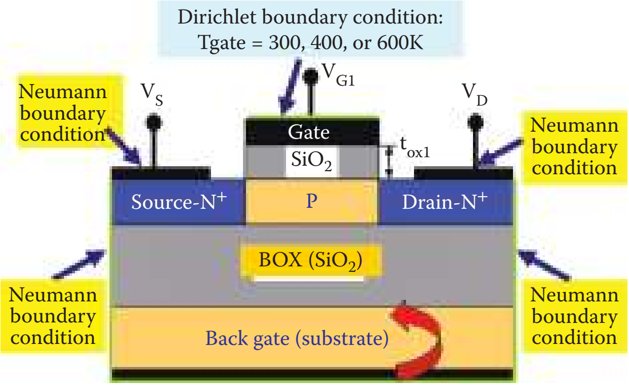

Dirichlet boundary condition:
Tgate = 300, 400, or 600K
V_{G1}
Neumann boundary condition
V_S
V_D
Gate
SiO_2
t_{ox1}
Neumann boundary condition
Source-N+
P
Drain-N+
BOX (SiO_2)
Neumann boundary condition
Neumann boundary condition
Back gate (substrate)

FIGURE 51.26 Schematic of the simulated device structure which illustrates the boundary conditions used in some of the simulations presented. Note also that the boundary conditions at the bottom of the substrate can be mapped as a boundary condition at the bottom of the BOX.

The question of the proper boundary conditions for the electronic part of the problem is rather clear and has been discussed in many papers and texts in the literature [35]. The problem in properly specifying the phonon boundary conditions is the selection of acoustic and optical phonon temperatures either at the artificial boundaries or at the contacts. Typically, acoustic phonon temperature is equated with the lattice temperature. To better understand the choice of boundary conditions we have considered in various works [15], we want to point out that lattice temperature is analogous to electrostatic potential and heat flux is analogous to current. As it is well known, when considering electrical behavior of the device, at least one node within the structure has to have Dirichlet boundary conditions specified (specify the potential for Poisson). Analogously, for the lattice temperature, we need at least one node that is a thermal contact and whose temperature is set to 300 K. As illustrated in Figure 51.26 for the case of FD-SOI devices, the bottom of the BOX is assumed to be at thermal equilibrium. Also, the gate contact is assumed to be at thermal equilibrium, but not necessarily at 300 K. For example, in one set of simulations, we vary the temperature on the gate to be 300, 400, and 600 K. Now comes the question: What happens to the source and drain contacts and the side artificial boundaries? Specifying Dirichlet boundary conditions at the ohmic contacts is not accurate from a standpoint that there is current flowing through the contacts, and since the contacts have finite resistance, there will be Joule heating (so the problem becomes unconstrained). The best solution to this, as we have done in several studies, is to extend the metal contact to become part of the simulation

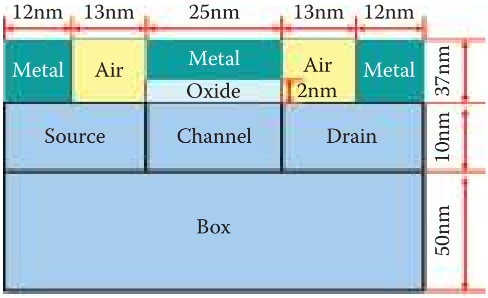


FIGURE 51.27 Device structure with extended domain. The geometrical dimensions are for the simulated 25-nm channel length FD-SOI nMOSFET.

domain and to apply at the very ends of the contact isothermal boundary conditions (see Figure 51.27). With the top and bottom specified, we have vertical transport of heat through the structure.

The next question is, is there lateral transfer of heat across the artificial boundaries? Here we can consider two cases: case when the neighboring device is ON, in which case Neumann boundary conditions are appropriate (derivative of the temperature), and case when the neighboring device is OFF, in which case Dirichlet boundary conditions are appropriate. In the case when Neumann boundary conditions are applied across the artificial boundaries, the heat transport remains vertical, but for the case when Dirichlet boundary conditions are applied across the boundary, the heat transport has both vertical and horizontal component. The case of Neumann boundary conditions can happen in analog circuits such as current mirror in which both transistors are simultaneously ON, whereas the case of Dirichlet boundary conditions occurs in digital circuits in which most of the time the transistors are OFF.

Yet another issue that deserves attention in getting physically correct results is the proper choice of the thermal conductivity for thin Si slabs and for nanowires. Asheghi and coworkers [36] and Li Shi and coworkers [37] have demonstrated via experimental measurements that thermal conductivities of thin Si films and Si nanowires, respectively, strongly depend of the geometry as for smaller geometries phonon boundary scattering can reduce the thermal conductivity of the Si film or nanowire by a factor of 10 or more from its bulk value. Moreover, thermal conductivity is temperature-dependent quantity. We made extensive efforts, using the theory of Sondhaimer for conductivity of metals [38], to arrive at an empirical formula that simultaneously described the thickness and temperature dependence of the thermal conductivity [39]. Our empirical expression perfectly matched the experimental data of Asheghi and coworkers [16]. In the model, it was assumed that phonon boundary scattering was perfectly diffusive.

51.4.2.2 Simulation Results

To validate our premise that self-heating effects become smaller as we scale devices into the nanometer regime (due to more pronounced velocity overshoot effects and thinner BOX layer that leads to smaller thermal resistance), in Table 51.3 we present the parameters of the devices being simulated. Simulation results for the lattice temperature profile in the Si film for devices with channel lengths from 25 to 180 nm are shown in Figure 51.28. We clearly see that the hot spot is in the channel for 180-nm channel length device and moves into the drain contact for 25-nm channel length device. The current degradation for isothermal temperatures on the gate of 300, 400, and 600 K are shown in Figure 51.29. These are the worst case scenario results as we use Neumann boundary conditions at the side boundaries.

In obtaining the results presented in Figures 51.28 and 51.29, constant thermal conductivity value of 13 W/m-K was used in the simulations [40]. The use of constant thermal conductivity model with value of 13 W/m-K leads to underestimation (25 nm channel length) or overestimation (180 nm channel length) of the average maximum temperature of the lattice in the channel region of the device.

TABLE 51.3 Parameters for Various Simulated Device Technology Nodes (Constant Field Scaling)

L(nm)	t_{ox} (nm)	t_{si} (nm)	t_{box} (nm)	N_{ch} (cm^{-3})	$V_{GS} = V_{DS}$ (V)	I_D (mA/ μm)
25	2	10	50	1×10^{18}	1.2	1.82
45	2	18	60	1×10^{18}	1.2	1.41
60	2	24	80	1×10^{18}	1.2	1.14
80	2	32	100	1×10^{17}	1.5	1.78
90	2	36	120	1×10^{17}	1.5	1.67
100	2	40	140	1×10^{17}	1.5	1.57
120	3	48	160	1×10^{17}	1.8	1.37
140	3	56	180	1×10^{17}	1.8	1.23
180	3	72	200	1×10^{17}	1.8	1.03

Note: *Parameters of the simulated structure given in Figure 51.26 are L, gate length; t_{ox}, gate oxide thickness; t_{Si}, active Si-layer thickness; t_{box}, BOX thickness; N_{ch}, channel doping concentration; I_D, isothermal current value (300 K).*

51.4.2.3 Summary of Results

In this section, we have presented modeling of self-heating effects in FD-SOI devices. Larger velocity overshoot and smaller BOX thickness lead to smaller degradation due to self-heating effects in 25-nm channel length FD-SOI devices. The amount of self-heating significantly depends upon the magnitude of the thermal conductivity used. Bulk values are inadequate and proper thickness and temperature dependence of

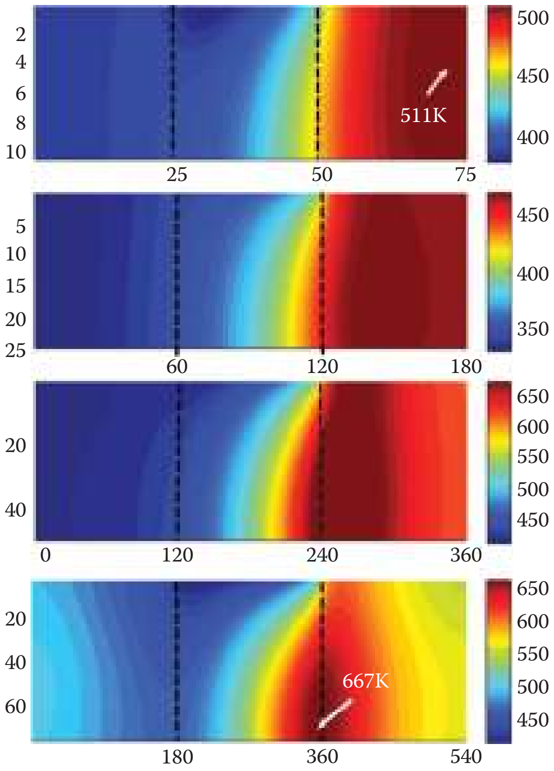


FIGURE 51.28 Lattice temperature profile in the active Si-layer for the FD-SOI devices from Table 51.2 ranging from 25 nm (top) to 180 nm (bottom) channel length. Shown here is only the thin Si film of the SOI device, namely the active area of the device.

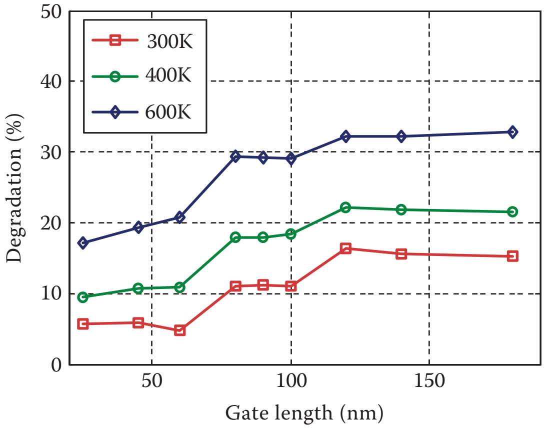


FIGURE 51.29 Current degradation (abs($I_{th}-I_o$)/I_o, where I_{th} is the on current when the self-heating model is incorporated and I_o is current in the isothermal case) versus technology generation for the FD-SOI devices from Table 51.2 ranging from 25- to 180-nm channel length. Different curves correspond to different temperatures on the gate electrode (bottom set of results: T_{gate} = 300 K, middle set of results: T_{gate} = 400 K and top set of results: T_{gate} = 600 K).

the thermal conductivity must be taken into consideration. The choice of the BOX material makes significant impact on the amount of observed self-heating. Both diamond and Si are good materials to be used as BOX. Diamond, when compared to aluminum nitride (AlN) is a better heat spreader.

51.5 Outlook

The drift-diffusion solvers are applicable in situations in which the bias conditions and the device geometry are such that electric fields are relatively low and velocity saturation model is applicable [41]. Situations in which drift-diffusion models are applicable are the Si-based power MOSFET devices, bipolar junction transistors, light-emitting diodes (that are used more and more in solid-state lightning), and crystalline solar cells, just to name a few. In some of these devices, such as power transistors and light-emitting diodes (LEDs) used for solid-state lighting, it is of paramount importance to incorporate self-heating models within the drift-diffusion framework [42]. The accuracy of simple heating models in conjunction with drift-diffusion models is to some degree questionable so that it is in many circumstances justifiable to use particle-based device simulators with more exact self-heating models [43].

On the other hand, the hydrodynamic models do not suffer from the limitations of the drift-diffusion approaches and the incorporation of the additional energy balance equation allows one to include velocity overshoot in the model. Velocity overshoot and nonstationary transport are key features of conventional MOSFET devices with gate lengths of 200 nm and below [44]. However, the magnitude of the velocity overshoot observed via simulations depends strongly upon the choice of the energy relaxation time, mostly in sub-100-nm channel length devices. This, in turn, affects the magnitude of the drain current. The reason for such drastic differences in the results when different energy relaxation times are used is due to the fact that the energy relaxation time is a material-dependent as well as device geometry-dependent parameter. So, to calculate better estimates for the velocity overshoot, higher moments of the BTE are needed [45]. These, in turn involve parameters that are more and more ambiguous on the expense of increased computational cost and when the computational cost of hydrodynamic models exceeds one of particle-based device simulators, there is no point in using moment methods. In these circumstances, the direct solution of the BTE via the Monte Carlo method becomes a method of choice. Thus, we might conclude that it is advisable to use particle-based device simulators when nanoscale devices are concerned [33].

But how far down in the scaling can we go? Particle-based device simulators capture on an equal footing both ballistic and diffusive transport, so if the ballisticity factor in the device increases [46], there is no problem that ballistic transport is effectively captured with particle-based device simulators [47]. Quantum-mechanical size quantization effects can also be captured by solving in slices the corresponding one-dimensional (1D) or 2D Schrödinger equation if one is concerned with conventional or FD-SOI MOSFETs, or nanowire transistors, respectively. What cannot be captured with particle-based device simulators is if there are local strains and stresses in the ultra-nanoscale devices, but that can also be cured via coupling of Monte Carlo device simulators with atomistic models [48,49] for bandstructure calculation.

In summary, Monte Carlo device simulators are a powerful tool for modeling devices ranging from the nanoscale regime to the microscale regime. What cannot be modeled with particle-based device simulators are resonant tunneling diodes in which quantum interference effects dominate the device behavior [50]. Efforts have been made along this direction as well [51], but the inclusion of the quantum-mechanical phase alongside with well-defined particle trajectory still remains open field of research.

References

1. ITRS International Technology Roadmap for Semiconductors. Available at: http://www.itrs.org.
2. G. Moore. Progress in digital integrated electronics. *IEDM Tech. Digest*, 11–13, 1975.
3. R. Dennard, F. H. Gaensslen, H. N. Yu, L. Rideout, E. Bassous, and A. R. LeBlanc. Design of Ion-Implanted MOSFET's with very small physical dimensions. *IEEE J. Solid State Circ.*, Vol. 9 (5), 256–268, 1974.
4. E. Fayneh, M. Yuffe, E. Knoll, M. Zelikson, M. Abozaed, Y. Talker, Z. Shmuely, and S. A. Rahme. 14 nm 6th-generation core processor SoC with low power consumption and improved performance. *IEEE International Solid-State Circuits Conference*, San Francisco, pp. 72–74, 2016.
5. T. Song, W. Rim, S. Park, Y. Kim, J. Jung, G. Yang, S. Baek, J. Choi, B. Kwon, Y. Lee, S. Kim, G. Kim, H. S. Won, J. H. Ku, S. S. Paak, E. Jung, S. S. Park, and K. Kim. A 10 nm FinFET 128 Mb SRAM with assist adjustment system for power, performance, and area optimization. *IEEE International Solid-State Circuits Conference*, San Francisco, pp. 306–309, 2016.
6. H.-Y. Chen, F.-C. Chen, Y.-L. Chan, K.-N. Yang, F.-L. Yang, and C. Hu. Method of fabricating a necked FINFET device. US 7122412 B2, 2006.
7. L. Geppert. A quantum leap for photonics. *IEEE Spectr.*, 28–33, April 9, 2004.
8. W. Zhu, J. P. Han, and T. P. Ma. Mobility measurement and degradation mechanisms of MOSFETs made with ultrathin high-k dielectrics. *IEEE Trans. Electron Dev.*, Vol. 51, 98–105, 2004.
9. J. Welser, J. L. Hoyt, and J. F. Gibbons. Temperature and scaling behavior of strained-Si N-MOSFET's. *IEDM Tech. Dig.,* 1000–1002, 1992.
10. R. Oberhuber, G. Zandler, and P. Vogl. Subband structure and mobility of two-dimensional holes in strained Si/SiGe MOSFET's. *Phys. Rev. B.*, Vol. 58, 9941–9948, 1998.
11. R. Chau, J. Kavalieros, B. Doyle, A. Murthy, N. Paulsen, D. Lionberger, D. Barlage, R. Arghavani, B. Roberts, and M. Doczy. A 50 nm depleted-substrate CMOS transistor (DST). *IEDM Techn. Dig.*, 29.1.1–29.1.4, 2001.
12. Y.-K. Choi, K. Asano, N. Lindert, V. Subramanian, T.-J King., J. Bokor, and C. Hu. Ultrathin-body SOI MOSFET for deep-sub-tenth micron era. *IEEE Electron Dev. Lett.*, Vol. 21 (5), 254–255, 2000.
13. D. Vasileska and S. M. Goodnick. Computational electronics. *Mater. Sci. Eng. R Rev. J.*, Vol. R38, 181–236, 2002.
14. C. Jacoboni and L. Reggiani. The Monte Carlo method for the solution of charge transport in semiconductors with applications to covalent materials. *Rev. Mod. Phys.*, Vol. 55, 645–705, 1983.
15. C. Jacoboni and P. Lugli. *The Monte Carlo Method for Semiconductor Device Simulation*. Vienna, VA: Springer-Verlag, 1989.
16. K. Hess. *Monte Carlo Device Simulation: Full Band and Beyond*. Boston, MA: Kluwer Academic Publishing, 1991.
17. M. H. Kalos and P. A. Whitlock. *Monte Carlo Methods*. New York, NY: Wiley, 1986.

18. D. K. Ferry. *Semiconductors*. New York, NY: Macmillan, 1991.
19. D. Vasileska, S. M. Goodnick, and G. Klimeck. *Computational Electronics: Semiclassical and Quantum Transport Modeling*. New York, NY: Taylor & Francis, 2010.
20. D. K. Ferry. *Quantum Mechanics: An Introduction for Device Physicists and Electrical Engineers*. New York, NY: Taylor & Francis, 2001.
21. H. D. Rees. Calculation of distribution functions by exploiting the stability of the steady state. *J. Phys. Chem. Solids*, Vol. 30, 643, 1969.
22. R. M. Yorston. Free-flight time generation in the Monte Carlo simulation of carrier transport in semiconductors. *J. Comp. Phys.*, Vol. 64, 177, 1986.
23. M. Lundstrom. *Fundamentals of Carrier Transport*. London: Cambridge University Press, 2000.
24. B. K. Ridley. *Quantum Processes in Semiconductors*. Oxford: Oxford University Press, 1982.
25. D. Vasileska. https://nanohub.org/groups/dragica_vasileska/semiclassical.
26. K. Tomizawa. *Numerical Simulation of Submicron Semiconductor Devices*. Boston, MA: Artech House, 1993.
27. M. V. Fischetti and S. E. Laux. Band structure, deformation potentials and carrier mobility in strained Si, Ge and SiGe alloys. *J. Appl. Phys.*, Vol. 80, 2234–2252, 1996.
28. T. Gonzalez and D. Pardo. Physical models of ohmic contact for Monte Carlo device simulation. *Solid State Electron.*, Vol. 39, 555–562, 1996.
29. P. A. Blakey, S. S. Cherensky, and P. Sumer. *Physics of Submicron Structures*. New York, NY: Plenum Press, 1984.
30. S. E. Laux. On particle-mesh coupling in Monte Carlo device simulation. *IEEE Trans. Comp. Aided Des. Int. Circ. Sys.*, Vol. 15, 1266, 1996.
31. R. Hathwar. Generalized Monte Carlo tool for investigating low-field and high field properties of materials using non-parabolic band structure model. MS Thesis, Arizona State University, June 2011.
32. S. S. Qazi. Electrical and thermal transport in alternative device technologies. MS Thesis, Arizona State University, November 2013.
33. K. Raleva, D. Vasileska, S. M. Goodnick, and M. Nedjalkov. Modeling thermal effects in nanodevices. *IEEE Trans. Electron Dev.*, Vol. 55 (6), 1306, 2008.
34. A. R. Shaik. Multi-scale study of heat transfer using Monte Carlo technique for phonon transport. MS Thesis, Arizona State University, April 2016.
35. D. Vasileska, K. Raleva, and S. M. Goodnick. Self-heating effects in nanoscale FD SOI devices: The role of the substrate, boundary conditions at various interfaces, and the dielectric material type for the box. *IEEE Trans. Electron Dev.*, Vol. 56, 3064, 2009.
36. W. Liu and M. Asheghi. Phonon-boundary scattering in ultrathin single-crystal silicon layers. *Appl. Phys. Lett.*, Vol. 84, 3819, 2004.
37. D. Li, Y. Wu, P. Kim, L. Shi, P. Yang, and A. Majumdar. Thermal conductivity of individual silicon nanowires. *Appl. Phys. Lett.*, Vol. 83, 2934, 2003.
38. E. H. Sondheimer. The mean free path of electrons in metals. *Adv. Phys.*, Vol. 1, 1, 1952, reprinted in *Adv. Phys.*, 50, 499, 2001.
39. D. Vasileska, K. Raleva, and S. M. Goodnick. Electrothermal studies of FD SOI devices that utilize a new theoretical model for the temperature and thickness dependence of the thermal conductivity. *IEEE Transactions on Electron Devices*, Vol. 57, 726–728, 2010.
40. K. Raleva and D. Vasileska. The importance of thermal conductivity modeling for simulations of self-heating effects in FD SOI devices. *J. Comput. Elec.*, Vol. 12 (4), 601–610, 2013.
41. D. Vasileska and S. M. Goodnick, Editors. *Nano-Electronic Devices: Semiclassical and Quantum Transport Modeling.* Vienna, VA: Springer Verlag, June 2011.
42. *www.silvaco.com*.
43. D. Vasileska. Modeling thermal effects in nano-devices. *Microelectron. Eng.*, Vol. 109 (9), 163–167 2013.

44. T. Grasser, T. W. Tang, H. Kosina, and S. Selberherr. A review of hydrodynamic and energy-transport models for semiconductor device simulation. *Proc. IEEE*, Vol. 91 (2), 251–274, 2003.
45. T. Grasser, H. Kosina, C. Heitzinger, and S. Selberherr. Characterization of the hot electron distribution function using six moments. *J. Appl. Phys,* Vol. 91, 3869, 2002.
46. M. Lundstrom. Elementary scattering theory of the Si MOSFET. *IEEE Electron Dev. Lett.,* Vol. 18, 361–363, 1997.
47. M. V. Fischetti and S. E. Laux. Monte Carlo analysis of electron transport in small semiconductor devices including band-structure and space-charge effects. *Physical Review B*, Vol. 38(14), 9721, 1988.
48. T. Frauenheim, G. Seifert, M. Elstner, T. Niehaus, C. Köhler, M. Amkreutz et al. Atomistic simulations of complex materials: Ground-state and excited-state properties. *J. Phys. Condens. Matter*, Vol. 14 (11), 3015, 2002.
49. G. Klimeck, S. S. Ahmed, H. Bae, N. Kharche, S. Clark, B. Haley, S. Lee et al. Atomistic simulation of realistically sized nanodevices using NEMO 3-D—Part 1: models and benchmarks. *IEEE Trans. Electron Dev.,* Vol. 54 (9), 2079–2089, 2007.
50. R. Lake, G. Klimeck, R. C. Bowen, and D. Jovanovic. Single and multiband modeling of quantum electron transport through layered semiconductor devices. *J. Appl. Phys.*, Vol. 81 (12), 7845–7869, 1997.
51. L. Shifren and D. K. Ferry. Wigner function quantum Monte Carlo. *Physica B: Condens. Matter*, Vol. 314 (1), 72–75, 2002.

52

Photonics

52.1 Landscape of Numerical Methods 808
Time-Harmonic versus Time-Dependent Modeling • Exact versus Approximate Models • Order of Approximation and Flexibility of Meshes • Computing Aspects: CPU-Time versus Memory • Special Adaptation to Photonics
52.2 Maxwell's Equations in Nondispersive, Chargeless Media 809
Scattering, Mode Propagation, and Resonance Problems • Simplification: 2D Geometries and 2D Fields • Transformation Rules • Perfectly Matched Layers
52.3 FDTD Method 817
Historical Notes • Concept: Discrete Differential Operators on Staggered Grids • Spatial Discretization of Maxwell's Equations • Time Evolution • Dispersion, Stability, and Courant–Friedrichs–Lewy Condition • Discrete Div-Conditions • Convergence
52.4 Fourier Modal Method 823
Historical Notes • Concept: Segmentwise Waveguide Treatment • Transfer-Matrix Approach • Scattering-Matrix Approach • Layered Media Stack as Model Case • FMM in 2D • Convergence
52.5 Finite Element Method 833
Historical Notes • Concept: Patchwise Tangential Continuous, Polynomial Approximation on Unstructured Grids • 1D: Weak Form and Variational Formulation • 1D: Linear Finite Elements, Local Matrices, and Assembly Process • 2D and 3D: Weak Form and Variational Formulation • 2D and 3D: Linear Finite Elements, Local Matrices, and Assembly Process • Convergence
52.6 DG Method.. 844
Historical Notes • Concept: Combine Adaptive Spatial FEM Discretization with the Time Treatment in Finite Volume Methods • Scalar Advection Equation as Model Equation: The Finite Volume Method in 1D • Maxwell's Equations in Conservation Form • Concept of DG-Algorithms for 3D Maxwell's Equations • Convergence

Frank Schmidt

This chapter deals with numerical simulation in photonics. It provides an overview of the most widely used methods, explains their concepts, and gives some details, as long as it is necessary for a qualitative understanding. It is far from being complete but tries to describe the landscape of simulation tools. Parts of this review are closely related to the book *Numerical Methods in Photonics* [1], which itself is a comprehensive representation of the field covered by many very detailed monographs. The methods we describe in the following are the *finite-difference time-domain FDTD method*, closely related to the *finite integration technique* (FIT), the *Fourier modal method* (FMM), the *finite element method* (FEM), and the *discontinuous galerkin* (DG) method. We omit other very useful methods often specialized to certain applications, in particular:

- Beam propagation method (BPM), for simulation of paraxial beam propagation [2]
- Integral methods such as boundary integral methods, for simulation problems in grating design [3]
- Plane wave expansion (PWE) for simulations of band structures [4]

and refer the reader to the given monographs.

52.1 Landscape of Numerical Methods

There are many aspects to consider if one has to choose a numerical method to solve a particular problem in photonics. We now discuss a few.

52.1.1 Time-Harmonic versus Time-Dependent Modeling

In the most general case, one wants to solve time-dependent Maxwell's equations with general, i.e., nonlocal, nonlinear, dispersive media. Both FDTD and DG methods in time provide such general solutions. On the other hand, especially in the case of high accuracy requirements, strict conditions on the maximal stepsizes in time may cause a high amount of central processing unit (CPU)-time consumption. Therefore, in case of linear and nondispersive problems, it is often useful to consider time-harmonic problems, and, if needed, to combine time-harmonic solutions via Fourier transform with different frequencies to obtain a time-dependent behavior.

52.1.2 Exact versus Approximate Models

It is common to approximate Maxwell's equations by structurally simpler equations. Especially, the simplification in terms of polarizations have a long tradition, where the full vectorial equations are replaced by semivectorial or scalar approximations. The scalar Helmholtz equation, e.g., is well investigated and a number of efficient numerical approaches exists, hence at first glance it seems reasonable to try Helmholtz solvers instead of Maxwell solvers. But the full vectorial Maxwell solvers available today do not have significant disadvantages with respect to speed or memory consumption, thus it is obsolete in many cases to use such semivectorial or scalar approximations. The case is different for special situations such as paraxial field propagations along distinguished propagation directions. As long as reflections can be neglected, in the case of waveguide tapers for example, so-called BPM are useful. They map a time-harmonic three-dimensional (3D) scattering problem approximately to a time-like evolution problem with 2D cross sections, which results in a substantial reduction of the computational effort.

52.1.3 Order of Approximation and Flexibility of Meshes

All methods described in the following offer, at least in principle, high-order approximations in space and time, thus resulting in high-order convergence. Nevertheless, this is not always used in practice. FDTD methods, e. g., are typically linked to a second-order discretization in time and space on staggered grids. Therefore, once geometrical flexibility is needed, one would typically, instead of enhancing FDTD, directly apply DG. Implementation of DG is more involved than of FDTD, but offers the flexibility of FEM meshes by design and yields the time evolution with costs comparable to FDTD.

52.1.4 Computing Aspects: CPU-Time versus Memory

The time-dependent solvers FDTD and DG can be treated by explicit time integrators. Thus, the evolution procedure results in much less memory consumption than that is typical for FMM or FEM, where either full or large sparse matrices have to be inverted. Hence, on memory-restricted systems, FDTD or DG are advantageous. The price to pay is longer computing times.

52.1.5 Special Adaptation to Photonics

FDTD, DG, and FEM are general methods for solving electromagnetic problems; they are not specially tailored to photonics. Since mode propagation is often the central physical effect, numerical methods based on the interaction of waveguide or cavity modes are often very effective. The dominant class of methods here are the FMMs. Subdividing the structure into waveguide sections and relating waveguide modes to each other result in algorithms which often offer surprisingly fast convergence in the range of lower, but in practice sufficient accuracy.

52.2 Maxwell's Equations in Nondispersive, Chargeless Media

In the most general case, the numerical methods of photonics solve time-dependent Maxwell's equations with general media, i.e., with nonlocal, nonlinear, dispersive media. For a concise and clear presentation of the different numerical methods, we use only a subset of the general equations. Our discussions are based on

$$\begin{aligned}\nabla \times \mathbf{H} &= \mathbf{J} + \frac{\partial \mathbf{D}}{\partial t}\\ \nabla \times \mathbf{E} &= -\frac{\partial \mathbf{B}}{\partial t}\\ \nabla \cdot \mathbf{D} &= 0\\ \nabla \cdot \mathbf{B} &= 0\end{aligned} \tag{52.1}$$

completed by the corresponding constitutive equations

$$\mathbf{D} = \varepsilon \mathbf{E} \qquad \mathbf{B} = \mu \mathbf{H} \qquad \mathbf{J} = \sigma \mathbf{E}. \tag{52.2}$$

These equations describe the fields in linear, nondispersive, and chargeless media, where the magnetic field **H**, the electric field **E**, the electric displacement **D**, and the magnetic flux density **B** are vectorial quantities defined for the entire space and positive times, $\mathbf{H}, \mathbf{E}, \mathbf{D}, \mathbf{B} : \mathbb{R}^3 \times \mathbb{R}^+ \longrightarrow \mathbb{R}^3$, and the permittivity ε, the permeability μ, and the conductivity σ are scalar, piecewise continuous functions in space, $\varepsilon, \mu, \sigma : \mathbb{R}^3 \longrightarrow \mathbb{R}$.

The field equations combined with the constitutive equations yield the system, on which our representation of the FDTD method and the DG method are based:

$$\begin{pmatrix} -\sigma & \nabla\times \\ \nabla\times & \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = \frac{\partial}{\partial t} \left[\begin{pmatrix} \epsilon & \\ & -\mu \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} \right]. \tag{52.3}$$

The other methods, FEM, FMM, etc., are based on the time-harmonic form of Equation 52.3. The concept is to use a Fourier transform of the fields in Equation 52.3 for a single frequency ω. Assuming the single frequency dependence to be $\exp(-i\omega t)$, the Fourier transformed fields become complex but live in space only, $\mathbf{H}, \mathbf{E} : \mathbb{R}^3 \longrightarrow \mathbb{C}^3$. We do not introduce a special notation for the time-harmonic fields, since the meaning becomes clear from the context. In the frequency domain, Equation 52.3 reads

$$\begin{pmatrix} & \nabla\times \\ \nabla\times & \end{pmatrix}\begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = -i\omega\left[\begin{pmatrix} \epsilon - \frac{\sigma}{i\omega} & \\ & -\mu \end{pmatrix}\begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix}\right]. \tag{52.4}$$

52.2.1 Scattering, Mode Propagation, and Resonance Problems

Maxwell's equations take different forms depending on the different modeling situations. In the case of time-dependent problems, these are the different types of linear and nonlinear scattering problems. In the case of time-harmonic problems, where mostly but not exclusively, linear problems are considered, we have scattering, mode propagation, and resonance problems. Time-dependent solutions to Equation 52.3 including also nonlocal, dispersive, or nonlinear materials are the most general classical solutions to Maxwell's equations—the corresponding computational effort is usually relatively large. Solutions to the time-harmonic system (52.4) are often much cheaper.

52.2.1.1 Scattering on Unbounded Domain

Figure 52.1 shows the situation of a time-harmonic scattering problem, the situation for time-dependent scattering is very similar. There is a source field $\mathbf{E}_{\text{source}}$ traveling through the unbounded space, being itself a solution to Maxwell's equations (Equation 52.4). If no scatterer would be present, this would be the only field we have. But if it hits an object, scattered fields $\mathbf{E}_{\text{scattered}}$ (or in short $\mathbf{E}_{\text{s}}$) are generated. We need to compute the field inside a computational domain Ω, which we also call the interior domain, in contrast to the exterior domain surrounding Ω. Further, we denote the boundary of the computational domain with $\partial\Omega$.

Let the source field $\mathbf{E}_{\text{source}}(\mathbf{r})$, with $\mathbf{r}$ in the exterior domain, $\mathbf{r} \in \mathbb{R}^3 \setminus \bar{\Omega}$ be given. Then, the general scattering problem in 3D is defined by the following conditions:

1. Interior problem. The field on the computational domain obeys Equation 52.4.
2. Exterior problem. The fields in the exterior domain are a superposition of the source and the scattered fields $\mathbf{E}_{\text{source}}$ and $\mathbf{E}_{\text{s}}$. The source field and the scattered field must also obey Maxwell's equations.
3. Continuity of the tangential data and their normal derivative along the boundary of the computational domain.
4. In the case of time-harmonic fields, the Silver–Müller radiation condition for the scattered field

$$\lim_{r\to\infty} r\left(\left(\nabla \times \mathbf{E}_{\text{s}}\right) \times \mathbf{r}^0 - i\,|\mathbf{k}|\,\mathbf{E}_{\text{s}}\right) = 0 \tag{52.5}$$

holds true uniformly in all directions $\mathbf{r}$. Here, $\mathbf{r}^0$ denotes the radial vector of unit length.

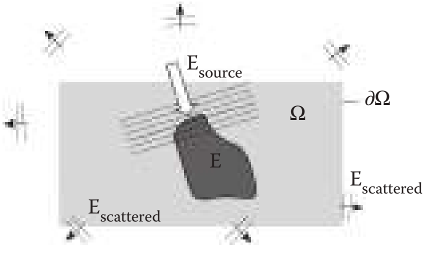


FIGURE 52.1 Scattering problem.

Typically, the radiation condition is not directly applied. But the methods that treat the exterior part of the problem ensure that the scattered fields satisfy the condition. The most prominent and successful method here is the perfectly matched layer (PML) method, see the Section 52.2.4.

52.2.1.2 Resonance Problems

In resonance problems, we consider Equation 52.4 as an eigenvalue problem with the unknown eigenvalue ω and the unknown resonance mode, the eigenvector, $(\mathbf{E}, \mathbf{H})^T$. The special challenge lies in the treatment of the exterior domain. Either the boundary conditions on the boundary of the computational domain are explicitly known, e. g., as a magnetic or electric wall, or methods from scattering problems are applied.

52.2.1.3 Mode Propagation Problems

Again, we consider Equation 52.4 as an eigenvalue problem. This time, we specialize the problem further to z-invariant problems: $\varepsilon = \varepsilon(x, y)$. We take a fixed frequency ω and look for solutions of the type $\mathbf{E}(x, y, z) = \mathbf{E}(x, y)e^{ik_z z}$ with a (unknown) phase velocity k_z. Two common forms of the eigenvalue problem for k_z are the following: Since any z-dependence is given by factors $\exp(ik_z z)$, one can introduce a specialized nabla-operator $\nabla_{k_z} = \mathbf{e}_x\partial_x + \mathbf{e}_y\partial_y + ik_z\mathbf{e}_z$ and Equation 52.4 reads

$$\begin{pmatrix} & \nabla_{k_z}\times \\ \nabla_{k_z}\times & \end{pmatrix}\begin{pmatrix}\mathbf{E}\\ \mathbf{H}\end{pmatrix} = -i\omega\left[\begin{pmatrix} \epsilon - \frac{\sigma}{i\omega} & \\ & -\mu\end{pmatrix}\begin{pmatrix}\mathbf{E}\\ \mathbf{H}\end{pmatrix}\right].$$

Here, the eigenvector consists of both $\mathbf{E}$ and $\mathbf{H}$, the eigenvalue k_z appears in linear form. Alternately, eliminating, e. g., the magnetic field, and using the decompositions $\mathbf{E} = \mathbf{E}_\perp + \mathbf{e}_z E_z$ and $\nabla = \nabla_\perp + i\mathbf{e}_z k_z$, one derives from Equation 52.4 the modal eigenvalue problem

$$\begin{bmatrix} \frac{1}{\varepsilon}\nabla_\perp \times \frac{1}{\mu}\nabla_\perp \times \quad +\frac{k_z^2}{\varepsilon\mu} & \frac{ik_z}{\varepsilon\mu}\nabla_\perp \\ i\frac{k_z}{\varepsilon}\nabla_\perp\frac{1}{\mu}\cdot & -\frac{1}{\varepsilon}\nabla_\perp\cdot\frac{1}{\mu}\nabla_\perp \end{bmatrix}\begin{bmatrix}\mathbf{E}_\perp \\ E_z\end{bmatrix} = \omega^2\begin{bmatrix}\mathbf{E}_\perp \\ E_z\end{bmatrix} \tag{52.6}$$

The eigenvector consists just of the $\mathbf{E}$-field, but this is a quadratic eigenvalue problem in k_z. The same procedure for the magnetic field yields

$$\begin{bmatrix} \frac{1}{\mu}\nabla_\perp \times \frac{1}{\varepsilon}\nabla_\perp \times \quad +\frac{k_z^2}{\varepsilon\mu} & \frac{ik_z}{\varepsilon\mu}\nabla_\perp \\ i\frac{k_z}{\mu}\nabla_\perp\frac{1}{\varepsilon}\cdot & -\frac{1}{\mu}\nabla_\perp\cdot\frac{1}{\varepsilon}\nabla_\perp \end{bmatrix}\begin{bmatrix}\mathbf{H}_\perp \\ H_z\end{bmatrix} = \omega^2\begin{bmatrix}\mathbf{H}_\perp \\ H_z\end{bmatrix}. \tag{52.7}$$

52.2.2 Simplification: 2D Geometries and 2D Fields

In a 2D geometry, the permittivity ε depends only on two coordinates, say the x- and y-coordinates: $\varepsilon = \varepsilon(x, y)$, just as in the mode propagation case in Section 52.2.1. In the mode propagation case, all fields depend on a common phase factor $\exp(ik_z z)$. If we set $k_z = 0$, i. e., we consider only fields with transversal position dependence $\mathbf{E} = \mathbf{E}(x, y)$, we obtain from (Equations 52.6 and 52.7) the decoupled 2D equations

$$\begin{aligned} \nabla_\perp \times \nabla_\perp \times \mathbf{E}_\perp(x, y) &= \varepsilon(x, y)\mu\omega^2\mathbf{E}_\perp(x, y) \\ -\nabla_\perp \cdot \nabla_\perp E_z(x, y) &= \varepsilon(x, y)\mu\omega^2 E_z(x, y). \end{aligned} \tag{52.8}$$

If we additionally assume uniformity in y-direction, we find the eigenvalue problem for waveguides with a 1D cross section

$$\partial_x^2 E_z + \varepsilon(x)\mu\omega^2 E_z = k_y^2 E_z. \tag{52.9}$$

The same applies to the magnetic field:

$$\nabla_\perp \times \frac{1}{\varepsilon(x,y)} \nabla_\perp \times \mathbf{H}_\perp = \mu\omega^2 \mathbf{H}_\perp(x,y)$$

$$\varepsilon(x)\partial_x \frac{1}{\varepsilon(x)} \partial_x H_z + \varepsilon(x)\mu\omega^2 H_z = k_y^2 H_z. \tag{52.10}$$

These equations are simpler than the original ones and they build the basis of many (approximate) solvers. They are simpler in the operator structure and they are linear in the eigenvalue k_y^2.

Note that, conventionally, the propagation direction of waveguide modes along the axis of waveguides is chosen to be the z-axis. To meet this convention, one has to interchange the y- and z-indices in Equations 52.9) and 52.10.

As a consequence of decoupling, we can split the entire field globally into a transversal electric (TE or s-polarization) and a transversal magnetic (TM or p-polarization) field:

$$\begin{pmatrix} \mathbf{E}(x,y) \\ \mathbf{H}(x,y) \end{pmatrix} = \underbrace{\begin{pmatrix} E_z(x,y)\mathbf{e}_z \\ H_x(x,y)\mathbf{e}_x + H_y(x,y)\mathbf{e}_y \end{pmatrix}}_{\text{s-polarization (TE)}} + \underbrace{\begin{pmatrix} E_x(x,y)\mathbf{e}_x + E_y(x,y)\mathbf{e}_y \\ H_z(x,y)\mathbf{e}_z \end{pmatrix}}_{\text{p-polarization (TM)}}. \tag{52.11}$$

52.2.3 Transformation Rules

Transformation techniques often yield considerable simplifications, e.g.,

- Exploiting symmetries by adapting coordinate systems (cylindrical, spherical coordinate systems, etc.)
- Using precomputations on reference elements as in FEM, cf. Section 52.5
- Handling of PML, cf. Section 52.2.4, by a complex coordinate stretching

The transformation rules describe the change of the equations when we change a local coordinate system. The PML, seen as a complex scaling of coordinates, can be treated as change of a coordinate system. The computations in FEM simplify if they are done in the reference coordinate systems and then mapped to the physical systems.

Geometric quantities and field quantities transform differently. For an easy *algebraic* notation, it is useful to write geometric quantities in row vectors and field quantities in column vectors of their coefficients. Expressions like $\mathbf{dl} \cdot \mathbf{E}$ with $\mathbf{dl} = (\mathrm{d}x, \mathrm{d}y, \mathrm{d}z)$ (a row vector of path components) and $\mathbf{E} = (E_x, E_y, E_z)^T$ (a column vector of field components) have then the right algebraic form and can be programed directly as they appear in the formula.

52.2.3.1 Mapping of Geometric Quantities

Let an arbitrary domain mapping $Q : \hat{K} \to K$ be given by

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = Q\left(\hat{x}, \hat{y}, \hat{z}\right), \tag{52.12}$$

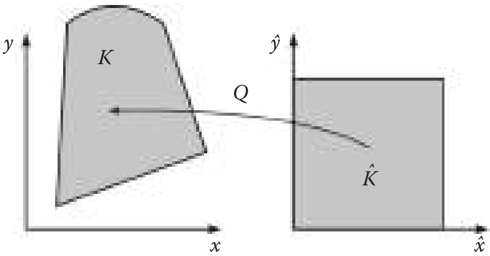


FIGURE 52.2 Example of a domain mapping. A square is mapped to a curvilinear quadrilateral. Quantities belonging to the reference domain are marked with a hat.

see Figure 52.2. This way we map *points* from the reference domain to *points* in the original domain, where the points are defined by the coefficients $\widehat{x}, \widehat{y}, \widehat{z}$ in the reference system and by the coefficients x, y, z in the original system.

We study how infinitesimally small path, face, and volume elements **dl**, d**A**, $\mathrm{d}V$, transform. In 3D, we describe **dl** and d**A** by three-coefficient row vectors. A linearization of the mapping at point $\left(\widehat{x}, \widehat{y}, \widehat{z}\right)$ yields

$$\begin{pmatrix} \mathrm{d}x \\ \mathrm{d}y \\ \mathrm{d}z \end{pmatrix} = J\left(\widehat{x}, \widehat{y}, \widehat{z}\right) \begin{pmatrix} \mathrm{d}\widehat{x} \\ \mathrm{d}\widehat{y} \\ \mathrm{d}\widehat{z} \end{pmatrix}, \qquad J\left(\widehat{x}, \widehat{y}, \widehat{z}\right) := \begin{pmatrix} \partial_{\widehat{x}} Q_1 & \partial_{\widehat{y}} Q_1 & \partial_{\widehat{z}} Q_1 \\ \partial_{\widehat{x}} Q_2 & \partial_{\widehat{y}} Q_2 & \partial_{\widehat{z}} Q_2 \\ \partial_{\widehat{x}} Q_3 & \partial_{\widehat{y}} Q_3 & \partial_{\widehat{z}} Q_3 \end{pmatrix}. \tag{52.13}$$

Defining the path element as the row vector $\mathbf{dl} = \left(\mathrm{d}x, \mathrm{d}y, \mathrm{d}z\right)$, we obtain the path element transformation

$$\mathbf{dl} = \mathrm{d}\widehat{\mathbf{l}} J^T. \tag{52.14}$$

A similar computation for a face element $\mathrm{d}\mathbf{A} = \left(\mathrm{d}A_x, \mathrm{d}A_y, \mathrm{d}A_z\right)$ spanned by two path elements gives, with $|J|$ the determinant of J, the transformation rule

$$\mathrm{d}\mathbf{A} = \mathrm{d}\widehat{\mathbf{A}} J^{-1} \, |J| \tag{52.15}$$

and for a volume element $\mathrm{d}V$ spanned by three path elements we obtain

$$\mathrm{d}V = \mathrm{d}\widehat{V} \, |J| \, . \tag{52.16}$$

52.2.3.2 Transformation of Grad, Curl, Div

We use the nabla notation describing a column vector

$$\mathbf{grad}\, v(x, y, z) = \nabla_{xyz} v(x, y, z) = \begin{pmatrix} \partial_x v(x, y, z) \\ \partial_y v(x, y, z) \\ \partial_z v(x, y, z) \end{pmatrix}.$$

Let an arbitrary function scalar $v(x, y, z)$ be given which lives on K and allows for the computation of the gradient. We compute the scalar quantity $\mathbf{dl} \cdot \nabla_{xyz} v$. This is a physical quantity with the meaning of a voltage (e.g., $\mathbf{E} \cdot \mathrm{d}\mathbf{s}$) or a current ($\mathbf{H} \cdot \mathrm{d}\mathbf{s}$). This quantity must not change if we change the coordinate system, hence, we require

$$\mathbf{dl} \cdot \nabla_{xyz} v(x, y, z) = \mathrm{d}\widehat{\mathbf{l}} \cdot \nabla_{\widehat{x}\widehat{y}\widehat{z}} \widehat{v} \left(\widehat{x}, \widehat{y}, \widehat{z}\right),$$

TABLE 52.1 Transformation of Geometric and Field Quantities

	Geometric Quantity	Field Quantity	Nabla Operation
Segment	$\mathbf{dl} = \mathrm{d}\widehat{\mathbf{l}}J^T$	$\mathbf{E} = J^{-T}\widehat{\mathbf{E}}$ $\mathbf{H} = J^{-T}\widehat{\mathbf{H}}$	$\nabla_{xyz}v = J^{-T}\nabla_{\widehat{x}\widehat{y}\widehat{z}}\widehat{v}$
Face	$\mathrm{d}\mathbf{A} = \mathrm{d}\widehat{\mathbf{A}}\,\lvert J\rvert\, J^{-1}$	$\mathbf{D} = \frac{J}{\lvert J\rvert}\widehat{\mathbf{D}}$ $\mathbf{B} = \frac{J}{\lvert J\rvert}\widehat{\mathbf{B}}$	$\nabla_{xyz} \times \mathbf{v} = \frac{J}{\lvert J\rvert}\nabla_{\widehat{x}\widehat{y}\widehat{z}} \times \widehat{\mathbf{v}}$
Volume	$\mathrm{d}V = \mathrm{d}\widehat{V}\,\lvert J\rvert$	$\rho = \frac{1}{\lvert J\rvert}\widehat{\rho}$	$\nabla_{xyz} \cdot \mathbf{v} = \frac{1}{\lvert J\rvert}\nabla_{\widehat{x}\widehat{y}\widehat{z}} \cdot \widehat{\mathbf{v}}$

where $\widehat{v}\left(\widehat{x},\widehat{y},\widehat{z}\right)$ in the reference domain follows from $v\left(\widehat{x},\widehat{y},\widehat{z}\right)$ in the original domain by $v(x,y,z) = v\left(Q\left(\widehat{x},\widehat{y},\widehat{z}\right)\right) = (v \circ Q)\left(\widehat{x},\widehat{y},\widehat{z}\right)$, hence $\widehat{v} = v \circ Q$, based on the coordinate transform Equation 52.12. It follows from Equation 52.14 that

$$\mathrm{d}\widehat{\mathbf{l}}J^T \cdot \nabla_{xyz}v(x,y,z) = \mathrm{d}\widehat{\mathbf{l}} \cdot \nabla_{\widehat{x}\widehat{y}\widehat{z}}\widehat{v}\left(\widehat{x},\widehat{y},\widehat{z}\right).$$

Since this should hold true for all functions v, we find the mapping of the gradient

$$\nabla_{xyz} = J^{-T}\nabla_{\widehat{x}\widehat{y}\widehat{z}}. \tag{52.17}$$

The coefficient of the nabla operator transform like the coefficients of the field strength. The same way one computes the transformations rules for **curl** and **div**, as collected in Table 52.1.

52.2.3.3 Transformation of μ and ε

Rewriting Maxwell's **curl** equations from the physical to the reference domain we get

$$\frac{J}{\lvert J\rvert}\nabla_{\widehat{x}\widehat{y}\widehat{z}} \times \widehat{\mathbf{H}} = i\omega\varepsilon J^{-T}\widehat{\mathbf{E}}$$

$$\frac{J}{\lvert J\rvert}\nabla_{\widehat{x}\widehat{y}\widehat{z}} \times \widehat{\mathbf{E}} = -i\omega\mu J^{-T}\widehat{\mathbf{H}},$$

which shows that the entire transformation can be traced back to a mapping of the material quantities

$$\widehat{\varepsilon} := \lvert J\rvert\, J^{-1}\varepsilon J^{-T} \tag{52.18}$$

$$\widehat{\mu} := \lvert J\rvert\, J^{-1}\mu J^{-T}. \tag{52.19}$$

52.2.4 Perfectly Matched Layers

A standard problem in the simulation of scattering problems is the treatment of scattered fields outside the domain of interest Ω_{int}, cf. Figure 52.3. The predominantly used method is the PML technique introduced by Berenger [5]. The PML treatment of Berenger relies on a split field formulation of Maxwell's equations in the PML region. This has been shown in [6] to be only weakly stable. In contrast, the uniaxial PML (UPML) [7] does not require such a splitting of Maxwell's fields and maintains the strong well-posedness of the original Maxwell system. Both formulations are equivalent to an elegant and general approach: the stretched-coordinate PML [8,9]. We describe briefly only the latter one. It has among others two remarkable properties

1. It can be used in a similar way in higher space dimensions.
2. It can be plugged into existing codes provided that the codes allow for sufficient general materials and field representations.

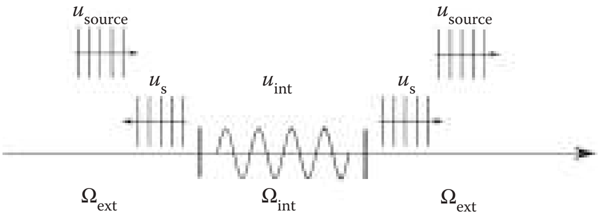


FIGURE 52.3 Sketch of a 1D time-harmonic scattering problem. The source fields are given, the interior field u_{int} and the scattered field u_S has to be computed.

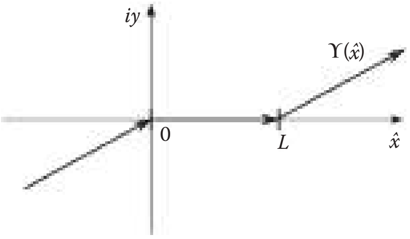


FIGURE 52.4 Complex PML path. Within the interior domain $(0, L)$ the x-coordinate is real but becomes complex outside.

52.2.4.1 PML in 1D

Let us consider the right exterior domain $x > L$, where L marks the right boundary point of the situation depicted in Figure 52.3. The general solution of the scattered field in the scalar 1D version of Equation 52.8 is

$$u_s = c_- e^{-ikx} + c_+ e^{ikx}, \quad x > L.$$

We have to discriminate between the incoming version of the scattered field $c_- \exp(-ikx)$, which we need to drop and the outgoing scattered field $c_+ \exp(ikx)$. Here, x is the real space coordinate. Now, we allow x to be a path in the complex plane, not just on the real axis. We parameterize path x with the real parameter $\hat{x}$ to derive a path in the complex plane; now $\hat{x}$ takes the role of the reference coordinate as in Section 52.2.3:

$$\begin{aligned} &x(\hat{x}) = L + (1 + i\sigma)\hat{x} \\ &\quad \sigma > 0, \text{ real, fixed} \\ &\quad \hat{x} > 0, \text{ real reference coordinate;} \end{aligned} \tag{52.20}$$

compare Figure 52.4, which yields the following mapping:

$$\begin{aligned} c_- \exp(-ikx) &\mapsto c_- \exp(-ikL) \exp(-ik\hat{x}) \exp(\sigma k\hat{x}) \\ c_+ \exp(ikx) &\mapsto c_+ \exp(ikL) \exp(ik\hat{x}) \exp(-\sigma k\hat{x}). \end{aligned}$$

The first expression, the incoming field, blows up for $\hat{x} \to \infty$, due to the factor $\exp(\sigma k\hat{x})$, whereas the second one decreases exponentially with $\exp(-\sigma k\hat{x})$. So for $\hat{x}$ large enough, we have a clear discrimination:

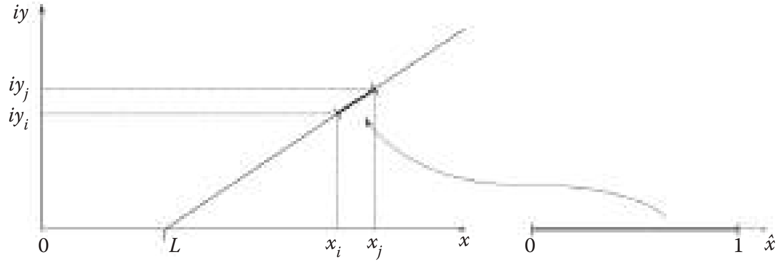


FIGURE 52.5 Transformation from the unit interval to the complex-valued PML interval.

nearly the entire field value comes from the left traveling scattered field. Hence, a proper formulation of Sommerfeld's radiation condition in terms of the complexified path $x(\hat{x})$ is

$$u\left(x\left(\hat{x}\right)\right) \longrightarrow 0 \ \text{ as } \hat{x} \longrightarrow \infty.$$

Let us stay in the right exterior domain and introduce the complex distance variable x:

$$x(\hat{x}) = \hat{x} + i\sigma\hat{x}.$$

We consider the complexification of the spatial variable within the Helmholtz equation:

$$\partial_x^2 u + k^2 u = 0 \qquad \Rightarrow \qquad \frac{1}{(1+i\sigma)^2}\frac{\partial^2 u}{\partial \hat{x}^2} + k^2 u = 0.$$

This step is decisive. If we are allowed to generalize the real derivative to a complex derivative, i. e., if the derivative is independent of the direction in the complex plane, we can use the linear path transformation, see Figure 52.5, (or others) to obtain the last equation.

This transformation is what we in fact exploit within the algorithmic realization. We define the complexification of the real x-coordinate to a complex coordinate x by means of the real reference coordinate $\hat{x}$

$$x = \begin{cases} \hat{x}\,(1+i\sigma)\,, & \hat{x} < 0 \\ \hat{x}, & 0 \le \hat{x} < L \\ L + \left(\hat{x} - L\right)(1+i\sigma)\,, & \hat{x} \ge L. \end{cases} \tag{52.21}$$

52.2.4.2 PML in 2D and 3D

The (tensor product type) generalization of Equation 52.20 to 2D is obviously the following (σ_x, σ_y are real, positive constants):

$$x = \begin{cases} \hat{x}\left(1+i\sigma_x\right), & \hat{x} < 0 \\ \hat{x}, & 0 \le \hat{x} < L\,, \\ L + \left(\hat{x} - L\right)\left(1+i\sigma_x\right), & \hat{x} \ge L \end{cases} \quad y = \begin{cases} \hat{y}\left(1+i\sigma_y\right), & \hat{y} < 0 \\ \hat{y}, & 0 \le \hat{y} < L \\ L + \left(\hat{y} - L\right)\left(1+i\sigma_y\right), & \hat{y} \ge L. \end{cases} \tag{52.22}$$

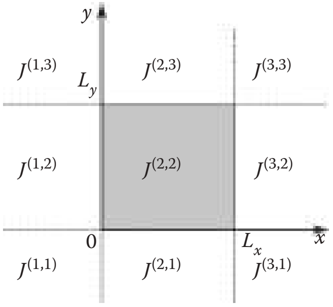


FIGURE 52.6 PML-regions in 2D lexicographically enumerated.

The Jacobian of the mapping $(\hat{x}, \hat{y}) \mapsto (x, y)$ depends on the position, see Figure 52.6, and evaluates to, e.g.,

$$J^{(2,2)} = \begin{pmatrix} 1 & \\ & 1 \end{pmatrix} \qquad J^{(3,2)} = \begin{pmatrix} 1 + i\sigma_x & \\ & 1 \end{pmatrix} \qquad J^{(3,3)} = \begin{pmatrix} 1 + i\sigma_x & \\ & 1 + i\sigma_y \end{pmatrix}.$$

The realization of the PML now consists in using the transformed permittivity (Equation 52.18) and permeability (Equation 52.19), where the Jacobians are computed from the complex stretching of the space variables (Equation 52.22). We proceed now as follows. We consider a subdomain of the exterior, say $\Omega^{(2,1)} \subset \mathbb{R}^2$ from Figure 52.6 and the complex solution $u^{(2,1)} : \Omega^{(2,1)} \to \mathbb{C}$ living there. Then, we use the mapping from the real to the complex coordinates to define a complex-valued spatial domain: $\Omega^{(2,1)} \to \Omega^{c,(2,1)} \subset \mathbb{C}^2$. On the complexified domain $\Omega^{c,(2,1)}$, we have a complex extension of $u^{(2,1)}$ which we denote by $u^{c,(2,1)}$. The solution $u^{c,(2,1)}$ is different to the one on the real space $\Omega^{(2,1)}$, it is the complex continuation of the original function. The key point is that our original function $u^{(2,1)}(x, y)$ has a complex extension by its nature, and we point to its values by introducing complex instead of only real spatial coordinates.

52.3 FDTD Method

The FDTD method is the most widely used method for the simulation of electromagnetic field propagation. It combines algorithmic simplicity with a deep universality. Nonlinear, nonlocal, and dispersive material responses can be treated. The reason for this generality lies in the fact that the structure of the FDTD scheme is very close to the structure of Maxwell's equations and has several algorithmic features:

- It is simple to implement and fast.
- High-order discretizations are feasible.
- It is explicit in time.
- There is a comprehensive theory.

Besides these pros, there are also severe drawbacks:

- The simple local approximation based on Cartesian staggered grids restricts geometric flexibility unless the grid is prohibitively fine (staircasing problem).
- The inherent second-order accuracy in time and space of the original method limits long-time computations on large domains.

52.3.1 Historical Notes

The initial and most important step was done by Yee in 1966 [10]. He used a so-called staggered mesh discretization in space and time, see the following subsections, and replaced the differential operators **grad**, **curl**, and **div** by finite-difference approximations to obtain a discrete approximation of second-order accuracy. Since then, a huge number of contributions improved the method and extended the fields of applications. An interesting view on the derivation of the FDTD equations has been given by Weiland in 1977 [11] starting from an integral formulation of Maxwell's equations. This approach is called FIT and yields the same set of equations. We use this approach to motivate the finite difference equations.

A number of efforts have been aimed at addressing the above-mentioned problems. The staircasing problem has been treated in many papers, e.g., [12]. High-order methods have been proposed, e.g., in [13, 14].The *generalized finite difference method* introduced by Bossavit [15] in 2001 takes ideas from differential geometry and topology and stimulates new interpretations of classical finite element schemes in terms of finite differences. Clemens and Weiland [16] follow a similar spirit in the framework of FIT to extend classical schemes.

A comprehensive overview of FDTD methods and further improvements along with detailed applications is given in [17] which is the standard reference book on FDTD methods. A thoroughly different approach to time-dependent problems is given by the class of DG methods which by its inherent construction offer high-order approximations in space and time as well as flexible meshes, cf. Section 52.6.

52.3.2 Concept: Discrete Differential Operators on Staggered Grids

Figure 52.7 explains the basic construction principle. We consider two entangled loops. One, say the light gray loop, is associated with the electric field, the other one with the magnetic field (left part of Figure 52.7). These loops will be used later for line integrals. The straight edges of the loops themselves are the edges of a cube, see the middle part of Figure 52.7. Then, by adding cubes related to the electric and magnetic fields, the complete spatial Yee mesh is generated. It can be considered as composed from two independent Cartesian meshes where the "electric" mesh and the "magnetic" mesh are shifted compared to each other by half of an elementary edge length of a cube.

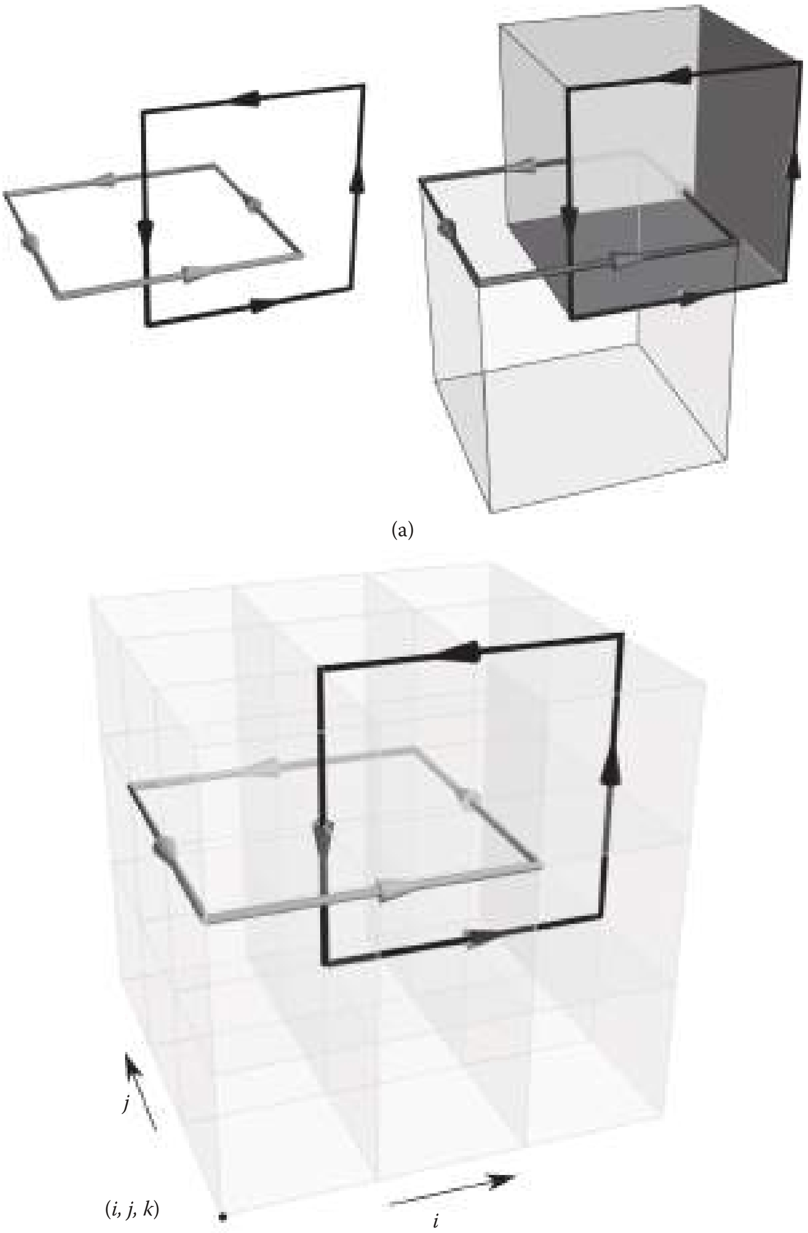
52.3.2.1 Degrees of Freedom

A central question in any numerical scheme is the proper definition of the degrees of freedom (DoF). Typically, the stability of a method essentially relies on this. In FDTD methods, it is common to project the fields to the edges and to take the values at the midpoints of the edges as DoF. For example, we project $\mathbf{E}$ to the ith edge $\mathbf{e}_i$, $\mathbf{E} \cdot \mathbf{e}_i$, take the value at the midpoint, and denote it as E_i. We motivate the DoFs in the following by approximations of integral expressions (e.g., by the midpoint rule) as it is done in the FIT. This offers additional insight into the structure of the equations. The same results are obtained if the differential operators are approximated by finite difference stencils.

52.3.2.2 Discrete Gradient

For a direct discretization of Maxwell's equations, it is not necessary to consider the discrete approximation of the gradient but it completes the insight into the structure of the subsequent discretization process. Consider an edge $\mathbf{e}$ of length h connecting two neighboring vertices v_1, v_2 of the Yee mesh. The fundamental theorem of calculus

$$\int_{v_1}^{v_2} \mathbf{grad}\, \phi \; d\mathbf{e} = \phi(v_2) - \phi(v_1),$$



(a)

(b)

FIGURE 52.7 Schematic interaction of electric and magnetic fields and its realization with staggered grids.

motivates the definition of the discrete gradient operator $\mathbf{grad}_\mathrm{h}$:

$$\mathbf{grad}_\mathrm{h}\, \phi\big|_{\mathbf{e}} \;:=\; \frac{\mathbf{e}^0}{h}\left(\phi(v_2) - \phi(v_1)\right) \text{ forall edges e,}$$

where $\mathbf{e}^0$ is the normalization of **e** to unit length. The $\mathbf{grad}_\mathrm{h}$ -operator maps quantities defined on *vertices* to quantities defined on *edges*.

52.3.2.3 Discrete Curl

The definition of the discrete **curl** -operator, $\mathbf{curl}_\mathrm{h}$, relies on the Stokes theorem

$$\int_{\mathbf{f}} \mathbf{curl}\,\mathbf{E}\ \mathrm{d}\mathbf{f} = \int_{\partial \mathbf{f}} \mathbf{E}\ \mathrm{d}\mathbf{e}$$

with **f** a plane, rectangular face of a cube and $\partial\mathbf{f}$ its boundary consisting of four edges $\mathbf{e}_i$. The discrete version reads

$$\mathbf{curl}_\mathrm{h}\ \mathbf{E}\big|_{\mathbf{f}}\ :=\frac{\mathbf{f}^0}{h_1 h_2}\sum_{i=1}^{4}\iota_{e_i,f}\left|\mathbf{e}_i\right| E_i \text{ forall faces } \mathbf{f}.$$

The iota-symbol $\iota_{e_i,f}$ denotes the incidence number of edge $\mathbf{e}_i$ with respect to the orientation of the face **f**. It holds $\iota_{e_i,f}\ =\ 1$, if edge $\mathbf{e}_i$ shares the same direction as that induced by the face vector **f**, otherwise $\iota_{e_i,f}\ =\ -1$. The $\mathbf{curl}_\mathrm{h}$ -operator maps quantities defined on *edges* to quantities defined on *faces*. For the light gray loop of Figure 52.7, e.g., starting with the edge in front of the figure and assuming the edge lengths of the rectangle are all the same $h_1 = h_2 = h$, this is

$$\mathbf{curl}_\mathrm{h}\ \mathbf{E}\big|_{\mathbf{f}}\ :=\frac{\mathbf{f}^0}{h}\left(E_1+E_2-E_3-E_4\right).$$

52.3.2.4 Discrete Divergence

The definition of the discrete **div** -operator, $\mathbf{div}_\mathrm{h}$, relies on the Gauss theorem

$$\int_{V} \mathbf{div}\,\mathbf{D}\ dV = \int_{\partial V} \mathbf{D}\ \mathrm{d}\mathbf{f}.$$

Here, V is the volume of the cube, ∂V its boundary consisting of 6 faces, and **f** is the outward pointing surface vector. The discrete version reads

$$\mathbf{div}_\mathrm{h}\ \mathbf{D}\big|_{c}=\frac{1}{h_1 h_2 h_3}\sum_{i=1}^{6}\iota_{e_f,c}\left|\mathbf{f}_i\right| D_i \text{ for all cells } c.$$

The iota-symbol $\iota_{f_i,c}$ denotes the incidence number of face $\mathbf{f}_i$ with respect to the orientation of cell c, defined positive by the outward orientation of all surface vectors. Hence, it holds $\iota_{f_i,c} = 1$, if face $\mathbf{f}_i$ points outward c, otherwise $\iota_{f_i,c} = -1$. The $\mathbf{div}_\mathrm{h}$ -operator maps quantities defined on *faces* to quantities defined on *cells*.

52.3.2.5 Exact Sequences

As a direct consequence of the fact that the Yee mesh is orientable, it follows

$$\begin{aligned}\mathbf{curl}_\mathrm{h}\ \cdot \mathbf{grad}_\mathrm{h}\ &= 0\\ \mathbf{div}_\mathrm{h}\ \cdot \mathbf{curl}_\mathrm{h}\ &= 0.\end{aligned} \tag{52.23}$$

Hence, the discrete operators share the same properties as the continuous ones. For the Yee mesh, the property follows easily from a direct computation.

52.3.3 Spatial Discretization of Maxwell's Equations

Let us denote by $\mathbf{E}_\mathrm{h}$ the column vector of all DoFs associated with the primal mesh and by $\mathbf{H}_\mathrm{h}$ the column vector of all DoFs associated with the dual mesh. Then, the spatial discrete version of Equation 52.3, setting $\sigma = 0$ for simplicity, reads

$$\varepsilon \partial_t \mathbf{E}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{p}} = \mathbf{curl}_\mathrm{h}\, \mathbf{H}_\mathrm{h}\big|_{\mathbf{f}_\mathrm{d}} \qquad \text{for all faces } \mathbf{f}_\mathrm{d} \text{ of the dual mesh,} \tag{52.24}$$

$$\mu \partial_t \mathbf{H}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{d}} = -\,\mathbf{curl}_\mathrm{h}\, \mathbf{E}_\mathrm{h}\big|_{\mathbf{f}_\mathrm{p}} \qquad \text{for all inner faces } \mathbf{f}_\mathrm{p} \text{ of the primal mesh.} \tag{52.25}$$

Here, $\mathbf{e}_\mathrm{p}$ denotes the edge of the primal mesh that penetrates face $\mathbf{f}_\mathrm{d}$ of the dual mesh and $\mathbf{e}_\mathrm{d}$ denotes the edge of the dual mesh going through $\mathbf{f}_\mathrm{p}$. For implementations, it is convenient to consider both the primal and the dual meshes as submeshes of a common Cartesian mesh.

52.3.4 Time Evolution

The time evolution of the semidiscretized equations (Equations 52.24 and 52.25) can be computed by a number of different time integrators. Frequently used is the so-called leapfrog scheme. It fits perfectly to the spatial discretization and offers the advantage of being an explicit method.

52.3.4.1 Leapfrog Scheme in Time

We introduce a uniform mesh in time

$$0 = t_0 < t_{\frac{1}{2}} < t_1 < \ldots < t_i < t_{i+\frac{1}{2}} < t_{i+1} \ldots < t^N = T,$$

with $t_{i+1} - t_i = \Delta t$ for all time steps and indicate the discrete time by a superscript at the quantities to use, e. g., the discrete electric field vector at time $t_{n+\frac{1}{2}}$ is $\mathbf{E}_\mathrm{h}^{n+\frac{1}{2}}$. Using a central finite difference approximation of the time derivative in Equation 52.24, we obtain its fully discretized version

$$\mathbf{E}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{p}}^{n+\frac{1}{2}} = \mathbf{E}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{p}}^{n-\frac{1}{2}} + \frac{\Delta t}{\varepsilon}\, \mathbf{curl}_\mathrm{h}\, \mathbf{H}_\mathrm{h}\big|_{\mathbf{f}_\mathrm{d}}^{n} \qquad \text{for all faces } \mathbf{f}_\mathrm{d} \text{ of the dual mesh.} \tag{52.26}$$

The time derivative is centered around n, hence it is of second-order accuracy. Using the updated electric field of Equation 52.26, Equation 52.25 is updated accordingly:

$$\mathbf{H}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{d}}^{n+1} = \mathbf{H}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{d}}^{n} - \frac{\Delta t}{\mu}\, \mathbf{curl}_\mathrm{h}\, \mathbf{E}_\mathrm{h}\big|_{\mathbf{f}_\mathrm{p}}^{n+\frac{1}{2}} \qquad \text{for all inner faces } \mathbf{f}_\mathrm{p} \text{ of the primal mesh.} \tag{52.27}$$

Equations 52.26 and 52.27 are typical FDTD update equations.

52.3.5 Dispersion, Stability, and Courant–Friedrichs–Lewy Condition

Let us consider a monochromatic plane wave of frequency ω in a homogeneous medium, with speed of light $c = 1/\sqrt{\varepsilon\mu}$, ε and μ independent of position, and a wavevector $\mathbf{k} = (k_x, k_y, k_z)$. Then the plane wave

$$\mathbf{E}e^{-i\omega t + i\mathbf{k}\mathbf{r}}$$

with a constant vector **E** is a solution of Maxwell's equations without sources, if the dispersion relation

$$\omega^2 = \mathbf{k}^2 c^2 \tag{52.28}$$

and orthogonality condition

$$\mathbf{kE} = 0$$

hold true. The first one follows from the **curl**-**curl** equations, the second one from the divergence condition. For the Yee mesh, there is a discrete counterpart. The discrete plane wave satisfies Equations 52.26 and 52.27 takes the natural form

$$\mathbf{E}_\mathrm{h} e^{-i\omega t_j + i\mathbf{k}\mathbf{r}_\mathrm{h}}.$$

The corresponding discrete dispersion relation, however, contains the stepsizes in time, Δt, and space, h_x, h_y, h_z,

$$\frac{1}{\Delta t^2}\sin^2\frac{\omega\Delta t}{2} = c^2\left(\left(\frac{1}{h_x^2}\sin^2\frac{k_x h_x}{2}\right) + \left(\frac{1}{h_y^2}\sin^2\frac{k_y h_y}{2}\right) + \left(\frac{1}{h_z^2}\sin^2\frac{k_z h_z}{2}\right)\right). \tag{52.29}$$

For a derivation see, e. g.,[17] or [1]. Clearly, this discrete dispersion relation approaches Equation 52.28 as $\Delta t, h_x, h_y, h_z \to 0$ as it must be. But the sin-function of the left-hand side of Equation 52.29 has a further consequence. Let k_x, k_y, k_z be real, hence the plane wave is spatially bounded everywhere as in the case of the the continuous plane wave. Multiplying with Δt^2 from the left-hand side of Equation 52.29 might result in a real number larger than 1, if Δt is large enough. Then ω onthe left-hand side becomes imaginary, and since the expression involves $\sin^2$, both positive and negative imaginary frequencies occur. The positive one will cause an exponential increase of the discrete plane wave in time, hence the solution becomes unstable. Consequently, for stability we must require

$$c^2\Delta t^2\left(\frac{1}{h_x^2} + \frac{1}{h_x^2} + \frac{1}{h_z^2}\right) \leq 1$$

or, for uniform meshes, $h = h_x = h_y = h_z$,

$$\Delta t \leq \frac{h}{c\sqrt{3}}.$$

These inequalities are the Courant–Friedrichs–Lewy (CFL) conditions for the wave equation.

52.3.6 Discrete Div-Conditions

It is a remarkable property that the Yee scheme preserves the **div**-conditions of Equation 52.1 for homogeneous media in discrete form [18]. Computing the discrete divergence of the updated Equation 52.26

$$\mathbf{div}_\mathrm{h}\, \varepsilon\mathbf{E}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{p}}^{n+\frac{1}{2}} = \mathbf{div}_\mathrm{h}\, \varepsilon\mathbf{E}_\mathrm{h}\big|_{\mathbf{e}_\mathrm{p}}^{n-\frac{1}{2}} + \Delta t\,\mathbf{div}_\mathrm{h}\ \mathbf{curl}_\mathrm{h}\ \mathbf{H}_\mathrm{h}\big|_{\mathbf{f}_\mathrm{d}}^{n} \quad \text{for all cells of the dual mesh.}$$

The last term vanishes due to property (Equation 52.23), hence we have the convergence

$$\mathbf{div}_\mathrm{h}\, \varepsilon \mathbf{E}_\mathrm{h}\Big|_{\mathbf{e}_\mathrm{p}}^{n+\frac{1}{2}} = \mathbf{div}_\mathrm{h}\, \varepsilon \mathbf{E}_\mathrm{h}\Big|_{\mathbf{e}_\mathrm{p}}^{n-\frac{1}{2}} = \dots = \mathbf{div}_\mathrm{h}\, \varepsilon \mathbf{E}_\mathrm{h}\Big|_{\mathbf{e}_\mathrm{p}}^{-\frac{1}{2}} \quad \text{for all cells of the dual mesh.}$$

The divergence attributed to a cell is preserved over all time steps. For the magnetic field, we obtain the corresponding result the same way:

$$\mathbf{div}_\mathrm{h}\, \mu \mathbf{H}_\mathrm{h}\Big|_{\mathbf{e}_\mathrm{d}}^{n+1} = \mathbf{div}_\mathrm{h}\, \mu \mathbf{H}_\mathrm{h}\Big|_{\mathbf{e}_\mathrm{d}}^{n} = \dots = \mathbf{div}_\mathrm{h}\, \mu \mathbf{H}_\mathrm{h}\Big|_{\mathbf{e}_\mathrm{d}}^{0} \quad \text{for all cells of the primal mesh.}$$

If the initial data at time step $n = -1/2$ and $n = 0$ are divergence-free,

$$\mathbf{div}_\mathrm{h}\, \varepsilon \mathbf{E}_\mathrm{h}^{-\frac{1}{2}} = 0 \qquad \mathbf{div}_\mathrm{h}\, \mu \mathbf{H}_\mathrm{h}^{0} = 0,$$

this property is preserved for all following time steps.

52.3.7 Convergence

Let, as before, $\mathbf{E}_\mathrm{h}$ be the vector of field values attributed to the edges of the primal mesh with values computed at the midpoint of edges. Let us denote the vector of the exact field values, projected to the edges, by $\mathbf{E}$; hence $\mathbf{E}_\mathrm{h}$ approximates $\mathbf{E}$ pointwise. We define the following, mesh-dependent discrete norm

$$\left\|\mathbf{E} - \mathbf{E}_h\right\|^2 \;:\; = h^3 \left(\mathbf{E}_\mathrm{h} - \mathbf{E}\right)^2,$$

and the analog for $\mathbf{H}$ with values attributed to the centroids of the cell faces.The following holds true. Let $\mathbf{E}$ and $\mathbf{H}$ be three times continuously differentiable in Ω. Then, there exists a constant C independent of h such that

$$\left\|\mathbf{E} - \mathbf{E}_\mathbf{h}\right\| + \left\|\mathbf{H} - \mathbf{H}_\mathbf{h}\right\| \leq Ch^2.$$

FDTD converges with second order. This classical result extends also to nonuniform meshes as shown by Monk and Süli [19].

52.4 Fourier Modal Method

This summary of FMMs follows Chapter 6 of [1]; for more details see that textbook and the references given therein.

52.4.1 Historical Notes

The FMM is a family of methods that belong to the class of spectral methods for the frequency domain and has been well established for more than three decades now. The term spectral method means that the field approximation is realized with functions which have a global support. In contrast, in FDTD methods, we consider only the pointwise-defined functions, and in FEM methods, we have basis functions with a very small, local support.

The origins of the FMM dates back to the 1960s and were devoted to grating analyses [20,21]. In 1981, Moharam and Gaylord established in their seminal paper [22] a computational framework they called rigorous coupled wave analysis (RCWA). Other synonyms or variations of the method include the PWE, the eigenmode expansion (EME), and eigenmode expansion technique (EET). A typical observation in RCWA methods has been that the convergence of TM fields is notably worse than that of TE fields. In 1996, Lalanne and Granet and coworkers [23,24] found a solution to overcome this TM convergence problem and in the same year Li published his factorization rules [25] formalizing the results. Currently, different proposals are being investigated to establish a fast convergent 3D FMM, e. g., [26].

52.4.2 Concept: Segmentwise Waveguide Treatment

The family of FMMs are characterized by the following building blocks:

1. The geometry is subdivided into a sequence of waveguide-type sections, see Figure 52.8b and c.
2. The field within each waveguide section is represented by up and down propagating modes as well as evanescent modes.
3. The transversal fields of the waveguide-type eigenmodes are computed in Fourier representation.
4. The interaction of the individual sections is realized by posing proper continuity conditions at the interfaces and a scattering-matrix formalism.

52.4.3 Transfer-Matrix Approach

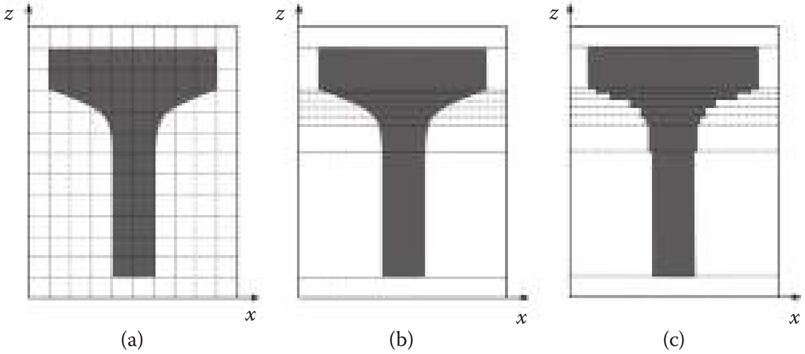
We consider the fields in two subsequent layers i and $i+1$, cf. Figure 52.9. At the moment, we do not focus on the special structure of the electromagnetic field except that we can decompose it into upward ($\mathbf{E}^+$) and downward ($\mathbf{E}^-$) propagating parts with respect to the z-direction. Let $\mathbf{E}_i^{\pm}(x,z)$ be arbitrary normalized solutions of Maxwell's equations split into propagation directions and $a_i^{\pm} \in \mathbb{C}$ their weights. It holds that

$$\begin{array}{rcccc} \text{Field in layer } i & : & a_i^+\mathbf{E}_i^+(x,z) & + & a_i^-\mathbf{E}_i^-(x,z) \\ \Downarrow \text{ implies} & & & & \\ \text{Field in layer } i+1 & : & a_{i+1}^+\mathbf{E}_{i+1}^+(x,z) & + & a_{i+1}^-\mathbf{E}_{i+1}^-(x,z). \end{array} \tag{52.30}$$

Given the complete field in one layer, Maxwell's equations including continuity conditions enforce the complete field in the other layer. This can be expressed by a transfer-matrix relation of the coefficients

$$\mathbf{A}_{i+1}\begin{pmatrix} a_{i+1}^+ \\ a_{i+1}^- \end{pmatrix} = \mathbf{A}_i\begin{pmatrix} a_i^+ \\ a_i^- \end{pmatrix}, \tag{52.31}$$

where $\mathbf{A}_i$, $\mathbf{A}_{i+1}$ are complex 2×2 matrices. We introduced this symmetric form since the count sequence of layers (top-down or bottom-up) is arbitrary. The typical application of the 1D transfer-matrix method is, given n layers of finite thickness with indices 0 and $n+1$ referring to substrate and superstrate, the



(a) (b) (c)

FIGURE 52.8 (a) A schematic geometry and a Cartesian mesh as used in FDTD methods. (b) Slicing of the geometry into segments along z. (c) Waveguide-like approximation of the geometry used in Fourier modal methods.

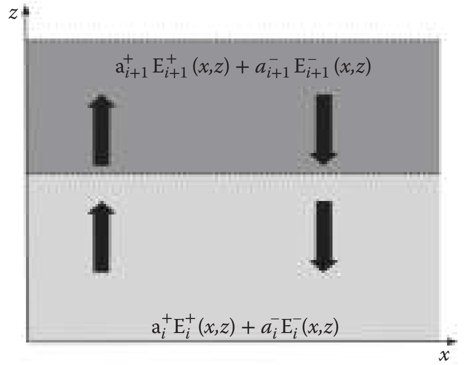

z
$a_{i+1}^{+} E_{i+1}^{+}(x,z) + a_{i+1}^{-} E_{i+1}^{-}(x,z)$
$a_{i}^{+} E_{i}^{+}(x,z) + a_{i}^{-} E_{i}^{-}(x,z)$
x

FIGURE 52.9 Superposition of up and down propagating fields in two subsequent layers.

update procedure

$$\begin{pmatrix} a_{n+1}^{+} \\ a_{n+1}^{-} \end{pmatrix} = \mathbf{T}_{n,n+1} \cdot \ldots \cdot \mathbf{T}_{i,i+1} \cdot \ldots \cdot \mathbf{T}_{0,1} \begin{pmatrix} a_{0}^{+} \\ a_{0}^{-} \end{pmatrix}.$$

This procedure may become unstable due to the exponentials involved in $T_{i,i+1}$.

52.4.4 Scattering-Matrix Approach

The scattering-matrix approach uses merely the same representation as the transfer-matrix approach, except that the computation of the coefficients $a_i^{\pm}$ is reordered (Figure 52.10). The scattering-matrix notation corresponding to Equation 52.31 is

$$\begin{pmatrix} a_{i}^{-} \\ a_{i+1}^{+} \end{pmatrix} = \mathbf{S}_{i,i+1} \begin{pmatrix} a_{i}^{+} \\ a_{i+1}^{-} \end{pmatrix} \tag{52.32}$$

with the obvious computation of the coefficients of $\mathbf{S}_{i,i+1}$ from $\mathbf{A}_i$ and $\mathbf{A}_{i+1}$. The notion scattering follows from the fact that we can consider the coefficients $\left(a_i^{+}, a_{i+1}^{-}\right)$ as fields *incident* to the layer separating boundary and the coefficients $\left(a_i^{-}, a_{i+1}^{+}\right)$ as fields *scattered* away from the boundary, respectively. The scattering matrix for an n-layer system is derived as follows. We rewrite Equation 52.31 as

$$\begin{pmatrix} 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_i\left(d_i\right) & -\mathbf{A}_{i+1}(0) \end{pmatrix} \begin{pmatrix} a_i^{+} \\ a_i^{-} \\ a_{i+1}^{+} \\ a_{i+1}^{-} \end{pmatrix}.$$

Here, the top coefficient a_i^{+} of the column vector and the bottom coefficient a_{i+1}^{-} are the incident coefficients. We repeat this for the next layer, increasing the index i by 1

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} = \begin{pmatrix} \mathbf{A}_i\left(d_i\right) & -\mathbf{A}_{i+1}(0) & \\ & \mathbf{A}_{i+1}\left(d_{i+1}\right) & -\mathbf{A}_{i+2}(0) \end{pmatrix} \begin{pmatrix} a_i^{+} \\ a_i^{-} \\ a_{i+1}^{+} \\ a_{i+1}^{-} \\ a_{i+2}^{+} \\ a_{i+2}^{-} \end{pmatrix}.$$

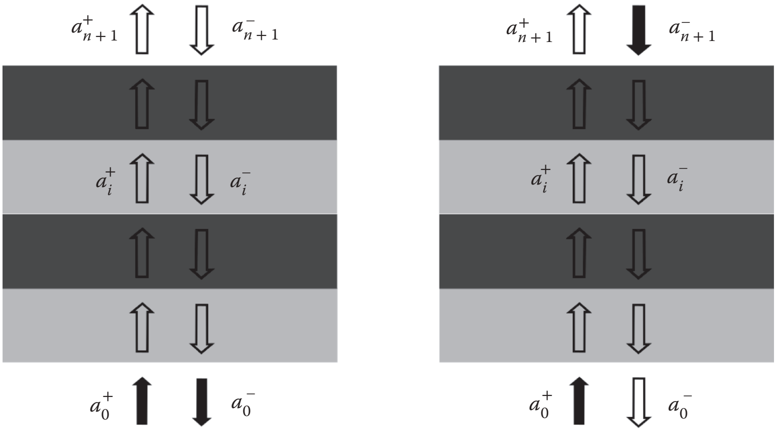


FIGURE 52.10 (a) Transfer-matrix approach. The data a_0^+, a_0^- are given (filled arrows), the others are computed. (b) Scattering-matrix approach. The data a_0^+, a_{n+1}^- are given, the others are computed.

Again, the top and bottom coefficients of the column vector are the incident coefficients of the layer system. Moving these coefficients to the left-hand side, deleting the first and last column of the system matrix, and inverting it yields the scattering-matrix formulation with scattering matrix **S**:

$$\mathbf{S}\begin{pmatrix} a_0^+ \\ 0 \\ \vdots \\ 0 \\ a_{n+1}^- \end{pmatrix} = \begin{pmatrix} a_0^- \\ a_1^+ \\ a_1^- \\ \vdots \\ \\ a_{n+1}^+ \end{pmatrix}. \tag{52.33}$$

52.4.5 Layered Media Stack as Model Case

A layered media stack is a 1D problem, if we choose the z-axis normal to the faces of the layers. To discuss the structure of FMM algorithms, we construct the standard transfer-matrix algorithm in scattering-matrix formulation from the viewpoint of the FMM. To find the field representation in layer i, we have to solve Maxwell's equations in layer i, i. e., either the equation for TE or TM. Let us consider the TE case. The modal ansatz

$$\text{layer } i: \quad E_y(x,z) = E(x)e^{ik_z z} = E_0 e^{ik_x x + ik_z z}$$

satisfies Equation 52.9 if $n_i^2 k_0^2 - k_x^2 = k_{z,i}^2$. Hence, given k_x, we have

$$\text{layer } i: \quad E_{y,i}(x,z) = a_i^+ e^{ik_x x + ik_{z,i} z} + a_i^- e^{ik_x x - ik_{z,i} z}, \tag{52.34}$$

which is the first line in Equation 52.30. Equally, we have for layer $i+1$

$$\text{layer } i+1: \quad E_{y,i+1}(x,z) = a_{i+1}^+ e^{ik_x x + ik_{z,i+1} z} + a_{i+1}^- e^{ik_x x - ik_{z,i+1} z}.$$

We compute matrices $\mathbf{A}_i$ and $\mathbf{A}_{i+1}$. Let the boundary between the two layers be placed at $z = 0$. The tangential electric and magnetic fields must be continuous. It follows for the electric field

$$a_i^+ e^{ik_x x} + a_i^- e^{ik_x x} = a_{i+1}^+ e^{ik_x x} + a_{i+1}^- e^{ik_x x} \quad \text{for all } x \in \mathbb{R},$$

hence

$$a_i^+ + a_i^- = a_{i+1}^+ + a_{i+1}^-. \tag{52.35}$$

Given the electric field in each layer, we compute the tangential component of the magnetic field

$$\begin{array}{rrcll} \text{layer } i: & \omega\mu H_{z,i}(x,z) & = & a_i^+ k_{z,i} e^{ik_x x + ik_{z,i} z} & - \; a_i^- k_{z,i} e^{ik_x x - ik_{z,i} z} \\ \text{layer } i+1: & \omega\mu H_{z,i+1}(x,z) & = & a_{i+1}^+ k_{z,i+1} e^{ik_x x + ik_{z,i+1} z} & - \; a_{i+1}^- k_{z,i+1} e^{ik_x x - ik_{z,i+1} z} \end{array}$$

and obtain the continuity condition

$$k_{z,i} a_i^+ - k_{z,i} a_i^- = k_{z,i+1} a_{i+1}^+ - k_{z,i+1} a_{i+1}^-. \tag{52.36}$$

Consequently, taking Equations 52.35 and 52.36,

$$\begin{pmatrix} 1 & 1 \\ k_{z,i+1} & -k_{z,i+1} \end{pmatrix} \begin{pmatrix} a_{i+1}^+ \\ a_{i+1}^- \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ k_{z,i} & -k_{z,i} \end{pmatrix} \begin{pmatrix} a_i^+ \\ a_i^- \end{pmatrix}. \tag{52.37}$$

Repeating this calculation with a finite layer thickness d_i between the two boundaries, we obtain

$$A_i(d_i) = \begin{pmatrix} 1 & 1 \\ k_{z,i} & -k_{z,i} \end{pmatrix} \begin{pmatrix} e^{ik_{z,i} d_i} & 0 \\ 0 & -e^{ik_{z,i} d_i} \end{pmatrix}. \tag{52.38}$$

Collecting the steps, we obtain the classical transfer-matrix algorithm in scattering formulation Algorithm 52.1. In 1D it is not necessary to save the eigenvectors since they are the same for all layers. But in 2D and 3D, the eigenvectors are different in each layer, so we introduced the step here for full parallelism with Algorithm 52.1.

Algorithm 52.1: Transfer-Matrix Algorithm in Scattering Formulation

---

```
Require: $a_0^+$ and $a_{n+1}^-$ to define source fields.
  for all layers do
    Solve each layer's eigenvalue problem (Equation 52.9).
    Save: eigenvectors $\exp(ik_x x)$ and eigenvalues $k_{z,i}$
    for later field representation (Equation 52.34).
    Compute the matrices $A_i$ by (Equation 52.38).
  end for
  Assemble the system (Equation 52.33).
  Solve.
  return coefficients $a_0^-, \dots, a_{n+1}^+$.
```

52.4.6 FMM in 2D

We generalize the foregoing, where ε has been constant in each layer. Now, let ε vary with x, each layer may have a different dependence $\varepsilon_i(x)$. The eigenvalue problems are Equations 52.9 and 52.10 with layerwise z-independent permittivities. In RCWA and FMM methods, these eigenvalue equations are solved based on a Fourier representation of the fields and the permittivities. This implies essentially two properties:

- Advantage: There is no lateral mesh needed. The nonlocal (or spectral) basis functions of fields and permittivities are simply constructed which results in a straightforward implementation.
- Disadvantage: The convergence of RCWA and FMM depend on the convergence properties of the Fourier representation which are not advantageous for nonsmooth functions.

52.4.6.1 Fourier Series Representation of Functions and Products of Functions

We consider a function $f(x)$ which is periodic with L: $f(x + L) = f(x)$. Hence, $f(x)$ has a Fourier series representation

$$f(x) = \sum_{m=-\infty}^{\infty} f_m e^{im\frac{2\pi}{L}x}. \tag{52.39}$$

We denote its Fourier transform by $\hat{f}$

$$\hat{f}(k) = \frac{1}{L}\int_0^L f(x)e^{-ikx}\,dx \text{ and compute } f_m = \hat{f}\left(\frac{2\pi}{L}m\right), \quad m \in \mathbb{Z}.$$

We consider a second function $g(x)$, also periodic with interval L, and study the product $h(x) = f(x)g(x)$, which is also L-periodic and consequently has a Fourier representation

$$h(x) = \sum_{m=-\infty}^{\infty} h_m e^{im\frac{2\pi}{L}x}.$$

In terms of the Fourier coefficients of f and g, this is

$$\begin{aligned} h(x) &= \left(\sum_{m=-\infty}^{\infty} f_m e^{im\frac{2\pi}{L}x}\right)\left(\sum_{n=-\infty}^{\infty} g_n e^{in\frac{2\pi}{L}x}\right) = \sum_{m=-\infty}^{\infty}\left(\sum_{n=-\infty}^{\infty} f_{m-n}g_n\right)e^{im\frac{2\pi}{L}x} \\ h_m &= \sum_{n=-\infty}^{\infty} f_{m-n}g_n. \end{aligned} \tag{52.40}$$

We write the last equation in matrix-vector form

$$\begin{aligned} \mathbf{h} &= \mathbf{Fg} \\ \mathbf{F} &= \begin{pmatrix} \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & f_0 & f_{-1} & f_{-2} & f_{-3} & \cdots \\ \cdots & f_1 & f_0 & f_{-1} & f_{-2} & \cdots \\ \cdots & f_2 & f_1 & f_0 & f_{-1} & \cdots \\ \cdots & f_3 & f_2 & f_1 & f_0 & \cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \end{pmatrix}, \end{aligned}$$

where the convolution matrix $\mathbf{F}$ is called the Toeplitz matrix.

52.4.6.2 Fourier Series Convergence in Dependence of the Smoothness of the Functions

In electromagnetics, we have to deal with piecewise smooth functions which have different global smoothness properties. Roughly, we have three classes:

Discontinuous: Functions that are piecewise smooth, but may have jumps or singularities. Example: Normal field components in a 3D setting with material discontinuities.

Continuous: Functions that are continuous but not differentiable at positions, where different materials meet. Example: H_y in TM fields as solution of, e.g., Equation 52.10.

Smooth: Functions that are continuous *and* the first derivative exists everywhere, even at positions, where different materials meet. Example: E_y in s-polarization, e.g., Equation 52.9.

52.4.6.3 Truncation

For an algorithmic application, the infinite series has to be truncated. The truncated Fourier series (Equation 52.39) then reads

$$f^M(x) = \sum_{m=-M}^{M} f_m e^{im\frac{2\pi}{L}x}, \quad f_m = \frac{1}{L}\int_0^L f(x)e^{-ik_m x}\,dx, \quad k_m = \frac{2\pi}{L}m \tag{52.41}$$

and the symmetric truncation of (Equation 52.40) yields

$$h^M(x) = \sum_{m=-M}^{M}\left(\sum_{n=-M}^{M} f_{m-n}g_n\right)e^{im\frac{2\pi}{L}x}, \quad h_m^M = \sum_{n=-M}^{M} f_{m-n}g_n.$$

Not only does the number of terms of the convolution sum change, but the individual coefficients h_n^M change too. The truncation changes the problem from an infinite to a finite one. The finite dimensional problem will differ from the exact, infinite dimensional one, as always in numerics, and the quality of this difference depends on the nature of the applied series truncation. The truncated matrix-vector form reads

$$\mathbf{h}^M = \mathbf{F}^M\mathbf{g}^M \qquad \text{“direct rule”} \tag{52.42}$$

$$\mathbf{F}^M = (f_{m-n})_{-M\le m,n\le M} \tag{52.43}$$

with F^M the finite dimensional Toeplitz matrix and $f_{m-n} = 0$ for $|m-n| > M$. As shorthand for $\mathbf{F}^M$ generated by coefficients f_m we also write $[\![f]\!] := \mathbf{F}^M$.

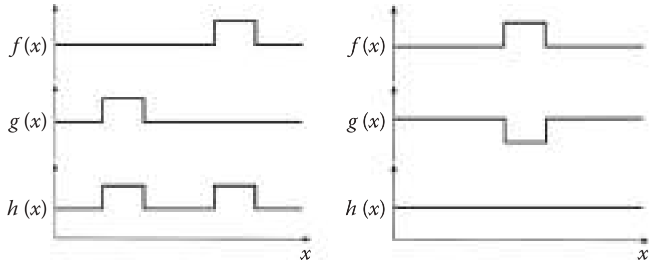
52.4.6.4 Li's Factorization Rules

It has turned out that the originally observed slow convergence rate of TM problems is related to the smoothness properties of the products of field and permittivity and a special, polarization-dependent treatment can be used to remove the difficulty [23–25]. To this end, we introduce the discontinuity-types shown in Figure 52.11:

1. Type: I The product $f(x)g(x)$ results in $h(x)$ which exhibits also jump discontinuities.
2. Type: II The product $f(x)g(x)$ yields a continuous function $h(x)$. The discontinuities are removed in $h(x)$.

For a formulation of the FMM algorithm, taking into account the accelerated TM technique, we need additionally, the finite Fourier series representation of the inverse function $f^\dagger$:

$$f^\dagger(x) := \frac{1}{f(x)}$$



Type I Type II

FIGURE 52.11 Two different types of jump discontinuities. Type I: The product $f(x)g(x)$ results in $h(x)$ which exhibits also jump discontinuities. Type II: The product $f(x)g(x)$ yields a continuous function $h(x)$.

together with its matrix-vector representation as counterpart to Equation 52.42

$$\mathbf{h}^M = \left(\mathbf{F}^{M\dagger}\right)^{-1} \mathbf{g}^M \qquad \text{“inverse rule.”} \tag{52.44}$$

The factorization rules are as follows:

1. Type: I If the discontinuities are not removable (Type I), apply the direct rule (Equation 52.42) for fastest convergence.
2. Type: II If the discontinuities are removable (Type II), apply the inverse rule (Equation 52.44) for fastest convergence.

52.4.6.5 Eigenvalue Problems in Fourier Space and Application of the Factorization Rules

The eigenvalue problems Equations 52.9 and 52.10 with ε independent of z hold true in each layer. We consider the TE equation Equation 52.9. Its continuous Fourier transform reads

$$\partial_x^2 E_y + \varepsilon(x)\mu\omega^2 E_y = k_z^2 E_y \quad \overset{\text{Fourier transform}}{\Longrightarrow} \quad (ik_x)^2\widehat{E}_y + \mu\omega^2\widehat{\varepsilon} * \widehat{E}_y = k_z^2\widehat{E}_y. \tag{52.45}$$

Restricting $E_y(x)$ to functions periodic over L (discretization in k-space), truncating the number of Fourier basis functions, and using the Toeplitz matrix to compute the convolution we obtain the eigenvalue equation in Fourier space

$$(i\mathbf{K}_x)^2\widehat{\mathbf{E}}_y + \mu\omega^2 [\![\widehat{\varepsilon}]\!] \widehat{\mathbf{E}}_y = k_z^2\widehat{\mathbf{E}}_y. \tag{52.46}$$

Equation 52.46 yields the generalization of the plane waves (TE) used in Equation 52.34 for the derivation of the transfer-matrix algorithm. The same procedure applies to the TM case, where strict consideration of the factorization rules is important. We compute

$$\begin{aligned} & \varepsilon(x)\partial_x \frac{1}{\varepsilon(x)}\partial_x H_y + \varepsilon(x)\mu\omega^2 H_y = k_z^2 H_y. \\ \overset{\text{Fourier transform}}{\Longrightarrow} \quad & \widehat{\varepsilon} * (ik_x)\widehat{\left(\frac{1}{\varepsilon}\right)} * (ik_x)\widehat{H}_y + \mu\omega^2\widehat{\varepsilon} * \widehat{H}_y = k_z^2\widehat{H}_y. \end{aligned} \tag{52.47}$$

For proper discretization based on the factorization rules, we rewrite Equation 52.47, factoring out $\widehat{\varepsilon}$,

$$\widehat{\varepsilon} * \left((ik_x)\widehat{\left(\frac{1}{\varepsilon}\right)} * (ik_x) + \mu\omega^2 \right) \widehat{H}_y = k_z^2 \widehat{H}_y. \tag{52.48}$$

First, $\partial_x H_y$ and $1/\varepsilon$ are discontinuous, their product is continuous, hence factorization rule II applies. Second, the large parenthesis on the left-hand side of Equation 52.48 must be discontinuous since ε is discontinuous. The product, however, is continuous, since the right-hand side is continuous. Hence, the factorization rule type II applies again:

$$\left[\!\left[\widehat{\left(\frac{1}{\varepsilon}\right)}\right]\!\right]^{-1} \left(i\mathbf{K}_x \left[\!\left[\widehat{(\varepsilon)}\right]\!\right]^{-1} (i\mathbf{K}_x) + \mu\omega^2 \right) \widehat{\mathbf{H}}_y = k_z^2 \widehat{\mathbf{H}}_y. \tag{52.49}$$

52.4.6.6 Continuity Condition

We derive the continuity condition Equation 52.31 for the 2D case. In two subsequent waveguide sections, the multimode generalization of Equation 52.30 holds:

$$\begin{array}{rcccc} \text{field in section } i & : & \sum_m a^+_{i,m}\mathbf{E}^+_{i,m}(x,z) & + & \sum_m a^-_{i,m}\mathbf{E}^-_{i,m}(x,z) \\ \Downarrow \text{implies} & & & & \\ \text{field in section } i+1 & : & \sum_m a^+_{i+1,m}\mathbf{E}^+_{i+1,m}(x,z) & + & \sum_m a^-_{i+1,m}\mathbf{E}^-_{i+1,m}(x,z). \end{array} \tag{52.50}$$

The generalization from plane waves (Equation 52.34) to waveguide modes reads

$$\text{section } i: \quad E_{y,i}(x,z) = \sum_m a^+_{i,m}\mathbf{E}_{i,m}(x)e^{ik_{z,i,m}z} + \sum_m a^-_{i,m}\mathbf{E}_{i,m}(x)e^{-ik_{z,i,m}z}. \tag{52.51}$$

Each mth normalized mode in section i has a Fouriermode representation $\mathbf{E}_{i,m} = \sum_l b_{i,m,l} \exp ik_{x,l}x$. Here, the weights $b_{i,m,l}$ (ith section, mth waveguide mode, lth Fouriermode of this waveguide mode) are, in case of TE fields, the coefficients of the eigenvectors (Equation 52.46), i.e., $b_{i,m,l} = \left(\widehat{\mathbf{E}}_y\right)_l$, if $\widehat{E}_y$ is the mth waveguide mode in the ith section. Hence,

$$E_{y,i}(x,z) = \sum_m a^+_{i,m} \sum_l b_{i,m,l}e^{ik_{x,i,m,l}x}e^{ik_{z,i,m}z} + \sum_m a^-_{i,m} \sum_l b_{i,m,l}e^{ik_{x,i,m,l}x}e^{-ik_{z,i,m}z}.$$

Setting the interface position between sections i and $i+1$ to $z = 0$ and requiring continuity of E_y yields

$$\sum_{m=-M}^{M} b_{i,m,l}\left(a^+_{i,m} + a^-_{i,m}\right) = \sum_{m=-M}^{M} b_{i+1,m,l}\left(a^+_{i+1,m} + a^-_{i+1,m}\right), \qquad -M \le l \le M.$$

In the same way, one computes the 2D counterpart to Equation 52.36

$$\sum_{m=-M}^{M} b_{i,m,l}\left(k_{z,i,m}a^+_{i,m} - k_{z,i,m}a^-_{i,m}\right) = \sum_{m=-M}^{M} b_{i+1,m,l}\left(k_{z,i+1,m}a^+_{i+1,m} - k_{z,i+1,m}a^-_{i+1,m}\right)$$

$$-M \le l \le M$$

Introducing the matrix

$$\mathbf{B}_i = \left(b_{i,l,m}\right)^T = \begin{pmatrix} \hat{\mathbf{E}}_{y,-M}, & \cdots & , \hat{\mathbf{E}}_{y,m}, & \cdots & , \hat{\mathbf{E}}_{y,M} \end{pmatrix}^T$$

allows writing the two continuity conditions in compact block matrix form:

$$\begin{pmatrix} \mathbf{B}_i & \\ & \mathbf{B}_i \end{pmatrix} \begin{pmatrix} \mathbf{I}_\mathrm{d} & \mathbf{I}_\mathrm{d} \\ \operatorname{diag} \mathbf{k}_{z,i} & -\operatorname{diag} \mathbf{k}_{z,i} \end{pmatrix} \begin{pmatrix} \mathbf{a}_i^+ \\ \mathbf{a}_i^- \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{B}_{i+1} & \\ & \mathbf{B}_{i+1} \end{pmatrix} \begin{pmatrix} \mathbf{I}_\mathrm{d} & \mathbf{I}_\mathrm{d} \\ \operatorname{diag} \mathbf{k}_{z,i+1} & -\operatorname{diag} \mathbf{k}_{z,i+1} \end{pmatrix} \begin{pmatrix} \mathbf{a}_{i+1}^+ \\ \mathbf{a}_{i+1}^- \end{pmatrix}$$

In this form, the similarity to the continuity condition Equation 52.37 of the transfer-matrix algorithm becomes obvious. For compact algorithmic notation, we define

$$\mathbf{A}_i := \begin{pmatrix} \mathbf{B}_i & \\ & \mathbf{B}_i \end{pmatrix} \begin{pmatrix} \mathbf{I}_\mathrm{d} & \mathbf{I}_\mathrm{d} \\ \operatorname{diag} \mathbf{k}_{z,i} & -\operatorname{diag} \mathbf{k}_{z,i} \end{pmatrix} \tag{52.52}$$

and have exactly as in the 1D case

$$\mathbf{A}_i \begin{pmatrix} \mathbf{a}_i^+ \\ \mathbf{a}_i^- \end{pmatrix} = \mathbf{A}_{i+1} \begin{pmatrix} \mathbf{a}_{i+1}^+ \\ \mathbf{a}_{i+1}^- \end{pmatrix}.$$

Collecting the steps, we obtain the FMM algorithm in 2D, Algorithm 52.2.

Algorithm 52.2: Fourier Modal Method in 2D

```
Require: $\mathbf{a}_0^+$ and $\mathbf{a}_{n+1}^-$ to define source fields.
  for all sections $i$ do
    Solve each sections's eigenvalue problem, either (Equation 52.46), TE, or (Equation 52.49), TM.
    Save: eigenvectors $\hat{\mathbf{E}}_{y,i,m}$, eigenvalues $k_{z,i,m}$
    for later field representation (Equation 52.51).
    Compute the matrices $A_i$ by (Equation 52.52).
  end for
  Assemble the system (Equation 52.33).
  Solve.
  return vectors $\mathbf{a}_0^-, \ldots, \mathbf{a}_{n+1}^+$.
```

52.4.7 Convergence

The two significant approximation errors in FMM are the geometrical approximation by staircasing and the trunction in the number of Fourier modes. These problems have been addressed in numerous papers (see also the introduction). Especially, the representation of permittivities, which are not continuous in lateral direction, by global Fourier modes poses a convergence problem. Since in this case, the corresponding

Fourier coefficients decay only as $1/n$, where n is the number of the Fouriermode, the spatial approximation is worse than the $1/n^2$ we know, e. g., from the FDTD method. On the other hand, physics very often behaves in a way that only a few modes are sufficient to describe reasonably wave propagation and scattering. In these cases, FMM has been proven to be a very efficient method.

52.5 Finite Element Method

Several excellent books treating the finite element analysis of Maxwell's equations are available; the books of Peter Monk [27] and Leszek Demkowicz [28] are very recommendable. The summary of FEM here follows Chapter 8 of [1].

52.5.1 Historical Notes

The deep understanding of the relation between the scalar Helmholtz case, which enforces the use of classical, nodal based Lagrange shape functions, and the vectorial Maxwell case relies greatly on the work of Whitney and Nédélec [29,30], Bosssavit [31], and Webb [32]. Many important contributions of other authors have added to the field, which makes the FEM today a very flexible, effective, and well-studied method for electrodynamical simulations. An early survey of finite element studies with focus on nano-optical applications has been given in [33].

52.5.2 Concept: Patchwise Tangential Continuous, Polynomial Approximation on Unstructured Grids

The FEM is based on a *variational formulation* of Maxwell's equations that involves integral expressions on the computational domain. The FEM does not use any approximation to Maxwell's equations itself, except the discretization of the geometry, but approximates the solution space in which one tries to find a reasonable approximation to the exact solution. This is in contrast to the finite difference method which approximates Maxwell's equations directly via an approximation of the differential operators by finite difference stencils. In some respects, the FEM is comparable to FMMs: one takes a number of trial functions (the Fourier modes in case of FMM) and asks in which way a superposition of them gives a reasonable approximation to the exact solutions.

The solution space of FEM, however, is obtained by subdividing the computational domain into small patches and by providing a number of polynomials on each patch for the approximation of the solution. The patches together with the local polynomials defined on them are called *finite elements*. The most common examples of finite elements are triangles and rectangles in 2D and tetrahedrons and cubes in 3D together with constant, linear, quadratic, and cubic polynomials. These locally defined polynomial spaces have to be pieced together to ensure tangential continuity of the electric and magnetic field across the boundaries of neighboring patches. Once these local approximations with proper continuity conditions have been defined, these are inserted into the variational equation. This results in a linear system whose solution is a piecewise polynomial approximation of the exact solution. Hence, the two basic ingredients in each FEM for Maxwell's equations are

1. A variational formulation of Maxwell's equations.
2. A suitable construction of finite elements based on polynomials defined on local geometric patches to transform the variational formulation into a discrete, algebraic problem.

This concept has a number of remarkable properties:

- Complex geometrical shapes can be treated without geometrical approximations, e.g., rounding can be well and easily approximated, cf. Figure 52.12.

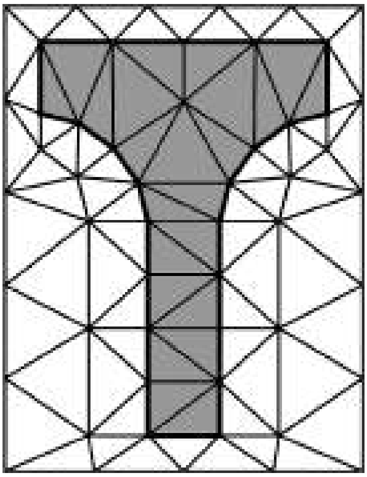


FIGURE 52.12 Sketch of a 2D triangulation for FEM, cf. also Figure 52.8 with staircase approximation. The triangles are shaped and placed such that the geometry is well covered by relatively few triangles.

- The finite element mesh can easily be adapted to the behavior of the solution, e.g., to singularities at corners.
- High-order approximations are available and ensure fast convergence.

52.5.3 1D: Weak Form and Variational Formulation

For the introduction of the finite element concept, we start with the most simple approximation of the time-harmonic scattering problem—the scalar Helmholtz equation in 1D, defined on the whole real axis, see Figure 52.3. The exterior domain should have a spatially invariant k. We let our computational domain Ω be the interval $(0, L)$. The source field is a plane wave coming from the left. A possible problem formulation reads: Find $u(x)$, $x \in \Omega = (0, L)$ such that

$$\partial_x^2 u(x) + k^2 u(x) = 0, \qquad x \in \mathbb{R} \tag{52.53}$$

$$u_{\text{source}}(x) = e^{ikx}, \qquad x \in \mathbb{R} \setminus \Omega. \tag{52.54}$$

Let us choose the interval $(0, L)$. We multiply Equation 52.53 with the complex conjugate of a so-called test function $v(x)$ and integrate over the domain Ω to arrive at

$$\int_0^L v^*(x) \left(\partial_x^2 u(x) + k^2 u(x)\right) dx = 0 \tag{52.55}$$

where nothing has been specified so far about $v(x)$, except that we must be allowed to compute the integral. This formulation cannot be equivalent to Equation 52.53, since very special choices of v would equate the integral to zero even for functions, which do not satisfy Equation 52.53. Hence, in order to have the chance to get something equivalent to Equation 52.53, Equation 52.55 must hold true for many functions $v(x)$. We call the class of functions from which we take $v(x)$, the *test space*. In contrast, the space of functions containing the desired solution $u(x)$ is called the *trial* or *ansatz space*. In general, both spaces can be different. It is subject to the convergence theory of finite elements to define precisely the test and trial spaces and to conclude convergence rates. For our purpose here, it is sufficient to define these spaces in a way that makes the construction of the numerical algorithms reasonable. In the first attempt, we require $\int_0^L v^* v\,dx < \infty$, which is the class of square-integrable functions, the $L^2(0,L)$-functions. We will make the choice of the space more precise below, but only in a framework that introduces usual FEM notions and gives a first look at the topic. A comprehensive treatment can be found, e. g., in [27]. The functions $u(x)$ should be twice differentiable since the second derivative is contained in Equation 52.53. Hence, we seek them in the class $C^2(0,L)$. So in a first attempt to construct a variational equation, we formulate find a function $u(x) \in C^2(0,L)$ such that Equation 52.55 holds true for all $v \in L^2(0,L)$.

52.5.3.1 Weak Form

We perform an integration by parts on the left-hand side of Equation 52.55:

$$\int_0^L v^*(x)\left(\partial_x^2 u(x) + k^2 u(x)\right) dx = \int_0^L \left(-\partial_x v^*(x)\partial_x u(x) + k^2 v^*(x)u(x)\right) dx + \left. v^*(x)\partial_x u(x)\right|_0^L . \tag{52.56}$$

It is convenient to abbreviate the negative of the integral without boundary terms by

$$a_{\text{int}}(v,u) := \int_0^L \left(\partial_x v^*(x)\partial_x u(x) - k^2 v^*(x)u(x)\right) dx \tag{52.57}$$

to rewrite (Equation 52.56) to

$$\int_{\Omega_{\text{int}}} v^*(x)\left(\partial_x^2 u(x) + k^2 u(x)\right) dx =: -a_{\text{int}}(v,u) + \sum_{x\in\partial\Omega_{\text{int}}} v^*(x)\partial_n u(x). \tag{52.58}$$

The term Equation 52.57 is called a *sesquilinear form* because it is antilinear in its first argument (the v, due to complex conjugation) and linear in its second argument (the u). Hence, $a_{\text{int}}(v,u)$ maps two complex functions via integration to a complex number. We may repeat this for the scattered part of the exterior field

$$a_{\text{int}}(v,u) - \sum_{x\in\partial\Omega_{\text{int}}} v^*(x)\partial_n u(x) = 0 \quad \text{in } \Omega_{\text{int}} \tag{52.59}$$

$$a_{\text{ext}}(v,u_s) - \sum_{x\in\partial\Omega_{\text{int}}} v^*(x)\partial_n u_s(x) = 0 \quad \text{in } \mathbb{R}\setminus\Omega_{\text{int}}, \tag{52.60}$$

where we ensure (see discussion below) that the boundary term at infinity vanishes and only the boundary $\partial\Omega_{\text{int}}$ contributes to the equation. Using on this boundary, $u = u_{\text{source}} + u_s$ and introducing an auxiliary function g with

$$\begin{aligned} u_s &= \underbrace{u_s + g}_{w} - g \\ g(\partial\Omega_{\text{int}}) &= u_{\text{source}}(\partial\Omega_{\text{int}}) \\ g(\infty) &= 0 \end{aligned} \tag{52.61}$$

we obtain

$$a_{\text{int}}(v,u) - \sum_{x\in\partial\Omega_{\text{int}}} v^*(x)\partial_n\left(u_{\text{source}} + u_{\text{s}}\right) = 0 \quad \text{in } \Omega_{\text{int}} \tag{52.62}$$

$$a_{\text{ext}}(v,w-g) - \sum_{x\in\partial\Omega_{\text{int}}} v^*(x)\partial_n u_{\text{s}}(x) = 0 \quad \text{in } \mathbb{R}\setminus\Omega_{\text{int}}. \tag{52.63}$$

Adding these two equations, taking into account the different normal directions, we end up with

$$a_{\text{int}}(v,u) + a_{\text{ext}}(v,w) = \sum_{x\in\partial\Omega_{\text{int}}} v^*(x)\partial_n u_{\text{source}} + a_{\text{ext}}(v,g). \tag{52.64}$$

Next, using the function u, which lives only on the interior domain, and the function w, which lives only on the exterior domain, we define a joint, continuous function, again called u. This gives rise to the variational formulation of the scattering problem.

52.5.3.2 Variational Formulation

Let the source field data u_{source} and $\partial_n u_{\text{source}}$ at the boundary $\partial\Omega$ of the computational domain Ω be given. The variational form of the scattering problem reads

$$\text{Find } u \in V: \quad a(v,u) = \sum_{x\in\partial\Omega_{\text{int}}} v^*(x)\partial_n u_{\text{source}} + a_{\text{ext}}(v,g) \quad \text{for all } v \in V, \tag{52.65}$$

with $a(\cdot,\cdot) = a_{\text{int}}(\cdot,\cdot) + a_{\text{ext}}(\cdot,\cdot)$, with a_{int} and a_{ext} given by Equation 52.59, Equation 52.60 and g defined by Equation 52.61. Note that in Equation 52.60 we neglected the contribution from the outer boundary $\partial\Omega$ext. The reason is that the expression is established for the scattered field u_{s} and we perform the integration over the complex extended exterior domain as described in Section 52.2.4. The complex extension leads to an exponential decay of u_{s}, hence the contribution of $\partial\Omega$ext vanishes.

52.5.4 1D: Linear Finite Elements, Local Matrices, and Assembly Process

The discretization of a given variational form to get the FEM equations is a straight forward procedure: just restrict the infinite dimensional test and trial space V to a finite dimensional space V_h:

$$\begin{array}{rcll} \text{Find } u \in V: \ a(v,u) & = & b(v) & \text{for all } v \in V. \\ & \downarrow & & \\ \text{Find } u_h \in V_h: \ a(v_h,u_h) & = & b(v_h) & \text{for all } v_h \in V_h. \end{array} \tag{52.66}$$

The space V_h is not uniquely predetermined. The different types of Galerkin-type methods differ in the construction of the elements of V_h. Let V_h be spanned by N linearly independent functions $v_j(x)$, $j = 1,\ldots,N$. We represent the discrete approximation of the exact solution u by $u_h(x) = \sum_{j=1}^{N} u_j v_j(x)$, insert this into Equation 52.66:

$$a\left(v_h, \sum_{j=1}^{N} u_j v_j(x)\right) = b(v_h) \tag{52.67}$$

and get the linear system

$$i = 1, \dots, N: \qquad \sum_{j=1}^{N} a_{ij} u_j = b_i \qquad \text{with } a_{ij} = a\left(v_i, v_j\right) \quad \text{and } b_i = b\left(v_i\right). \tag{52.68}$$

52.5.4.1 Linear Finite Elements

We want to construct the matrix $\left(a_{ij}\right)$ as a sparse matrix, i.e., a matrix with predominantly zero entries, which is a very useful numerical property. To this end we construct test functions v_j that exist only on small patches. The most common finite elements are the *piecewise linear "hat"*-functions shown in Figure 52.13a:

$$v_i(x) = \begin{cases} \frac{x - x_{i-1}}{h_{i-1}} & \text{for} \quad x \in \left[x_{i-1}, x_i\right] \\ -\frac{x - x_{i+1}}{h_i} & \text{for} \quad x \in \left[x_i, x_{i+1}\right] \\ 0 & \text{elsewhere,} \end{cases} \tag{52.69}$$

where we introduced the lengths of the subdomains with $h_{i-1} = x_i - x_{i-1}$ and $h_i = x_{i+1} - x_i$.

52.5.4.2 Assembly Process: Patchwise Computation of $A = \left(a_{ij}\right)$

We start from Equation 52.67 with the summation moved to front of $a(\cdot, \cdot)$:

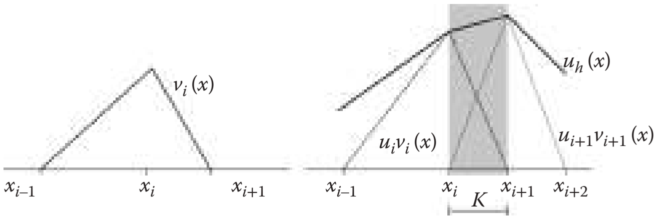
$$i = 1, \dots, N \qquad \sum_{j=1}^{N} a\left(v_i(x), v_j(x)\right) u_j = b(v_i).$$

Next we denote $a(\cdot, \cdot)$ restricted to patch K by $a(\cdot, \cdot)|_K = a_k(\cdot, \cdot)$ and rewrite the above in terms of a sum over patches:

$$i = 1, \dots, N \qquad \sum_{j=1}^{N} \sum_{k} a_k\left(v_i(x), v_j(x)\right) u_j = b(v_i).$$

Interchanging the order of summation we have

$$i = 1, \dots, N \qquad \sum_{k} \sum_{j=1}^{N} a_k\left(v_i(x), v_j(x)\right) u_j = b(v_i).$$



(a) (b)

FIGURE 52.13 (a) Piecewise linear test function $v_i(x)$. (b) Piecewise linear function u_h composed from trial functions $v_i(x)$ and $v_{i+1}(x)$. Only trial functions v_i and v_{i+1} have support on the interval K, but no other trial function.

In principle, this is a summation over k different matrices $A_k = a_k\left(v_i(x), v_j(x)\right)$, each of dimension $N \times N$ multiplied with a vector $u = (u)_i$ of dimension N, hence

$$\left(\sum_k A_k\right) u = b(v_i).$$

The matrices A_k have full dimension. However, since they are computed only over a single patch K they have entries different from zero only if both v_i and v_j have support over K, hence most entries but a few are zero. Suppose we have a set of (few) trial functions $v_{i_1}, v_{i_2}, \dots, v_{i_n}$ with index set $\{i_1, i_2, \dots, i_n\}$ which contribute to A_k assigned to patch K. We count these indices as follows:

$$\begin{array}{ccc} \text{local DOFs} & \rightarrow & \text{global DOFs} \\ (1, 2, \dots, n) & \rightarrow & \left(i_1, i_2, \dots, i_n\right) \end{array}$$

and store the connectivity data, using MATLAB ® notation, in an element connectivity vector

$$c = [\mathrm{i}_1, \mathrm{i}_2, \dots, \mathrm{i}_n]. \tag{52.70}$$

We have such a mapping for each patch connecting the *local* DOFs to the *global* ones. With the help of the local DOFs we define dense, so-called *local* matrices $A_k^{\mathrm{loc}} \in \mathbb{C}^{n \times n}$ of small dimension n and corresponding *local* vectors $b_k^{\mathrm{loc}} \in \mathbb{C}^n$:

$$\begin{aligned} A_k^{\mathrm{loc}}(l, m) \quad &: = \quad A_k\left(i_l, i_m\right) \quad \text{for all indices with non-vanishing values of } A_k, \text{ and} \\ b_k^{\mathrm{loc}}(l) \quad &: = \quad b_k(i_l). \end{aligned}$$

The process of distributing the entries of the local matrices to the global matrices is called index scattering. The MATLAB ® statement

$$A(c, c) = A^{\mathrm{loc}}$$

scatters the entries of the local matrix A^{loc} to their global counterparts.

52.5.4.3 Computation of Local Matrices

For a convenient computation of the local vectors and matrices, it is useful to decompose the form $a(v, u)$, e.g., Equation 52.57, into the so-called *stiffness* and *mass* parts $s(v, u)$, $m(v, u)$, respectively:

$$a(v, u) = s(v, u) - m(v, u) \quad \text{with}$$

$$s(v, u) : = \int_0^L \partial_x v^* \partial_x u \quad \text{and} \quad m(v, u) : = \int_0^L k^2 v^* u.$$

The terms stiffness and mass come from mechanical systems and have no meaning in electrodynamics. But, as we will see, both parts have very different mathematical properties.

Let us compute the stiffness matrix based on the hat-functions. We consider the patch K from Figure 52.13 which gives us the local⟶global assignment of indices

$$(1, 2) \longrightarrow (i, i + 1).$$

Using Equation 52.69, we get

$$s^{\text{loc}} = \begin{pmatrix} s\left(v_i, v_i\right)\big|_K & s\left(v_i, v_{i+1}\right)\big|_K \\ s\left(v_{i+1}, v_i\right)\big|_K & s\left(v_{i+1}, v_{i+1}\right)\big|_K \end{pmatrix}.$$

We compute directly

$$s^{\text{loc}} = \frac{1}{h_i}\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \quad \text{and the in same way} \quad m^{\text{loc}} = h_i \begin{pmatrix} \frac{1}{3} & \frac{1}{6} \\ \frac{1}{6} & \frac{1}{3} \end{pmatrix}.$$

The structure of a finite element algorithm is given in Algorithm 52.3.

52.5.5 2D and 3D: Weak Form and Variational Formulation

The Maxwell scattering problem with PML is derived in complete analogy to Section 52.5.3. Let the source field data $\mathbf{E}_{\text{source}}$ and $\nabla \times \mathbf{E}_{\text{source}}$ at the boundary $\partial\Omega$ of the computational domain Ω be given. The variational form of the scattering problem reads:

$$\text{Find } \mathbf{E} \in V: \quad a(\mathbf{v}, \mathbf{E}) = \int\limits_{\partial\Omega} d\mathbf{s} \cdot \left(\mathbf{v}^* \times \left[\left(\mu^c\right)^{-1} \nabla \times \mathbf{E}_{\text{source}}\right]\right) + a_{\text{ext}}\left(\mathbf{v}, \mathbf{g}\right) \quad \text{for all } \mathbf{v} \in V, \tag{52.71}$$

with

$$\begin{aligned} a(\cdot,\cdot) &= a_{\text{int}}(\cdot,\cdot) + a_{\text{ext}}(\cdot,\cdot) \\ a_{\text{int}}(\mathbf{v}, \mathbf{E}) &= \int\limits_{\Omega} dx\, dy\, \left(\nabla \times \mathbf{v}^*\right) \cdot \mu^{-1} \left(\nabla \times \mathbf{E}\right) - \omega^2 \int\limits_{\Omega} dx\, dy\, \mathbf{v}^* \varepsilon \mathbf{E} \\ a_{\text{ext}}(\mathbf{v}, \mathbf{g}) &= \int\limits_{\Omega_{\text{ext}}} dx\, dy\, \left(\nabla \times \mathbf{v}^*\right) \cdot \left(\mu^c\right)^{-1} \left(\nabla \times \mathbf{g}\right) - \omega^2 \int\limits_{\Omega_{\text{ext}}} dx\, dy\, \mathbf{v}^* \varepsilon^c \mathbf{E} \end{aligned}$$

and $\mathbf{g}$ is an arbitrary tangentially continuous function with

$$\begin{aligned} \mathbf{g}(\partial\Omega) &= \mathbf{E}_{\text{source}}(\partial\Omega) \\ \mathbf{g}(\infty) &= 0. \end{aligned}$$

The transformed permeabilities and permittivities are

$$\left(\mu^c\right)^{-1} = \frac{1}{|J|} J^T \mu^{-1} J, \qquad \varepsilon^c = |J|\, J^{-1} \varepsilon J^{-T},$$

with J the Jacobian of the complex extension. Here, we let Ω_{ext} be the original, real exterior domain and the PML-complexification is hidden in the transformed materials μ^c and ε^c, see Equations 52.18 and 52.19.

52.5.6 2D and 3D: Linear Finite Elements, Local Matrices, and Assembly Process

The basic scheme of deriving the numerical algorithm is exactly the same as in 1D, see Section 52.5.4, and also Algorithm 52.3:

1. Construct a mesh that describes the geometry plus surrounding domain. The meshes in 2D are triangular or quadrilateral meshes, in 3D these are meshes based on tetrahedrons, bricks, prisms, and pyramids. Typically, special mesh generators are employed. The construction of high-quality meshes is often a nontrivial task and a research field in its own right.
2. On each patch of the mesh, say on a triangle, a polynomial space is defined which is the basis of the construction of so-called shape functions. The local space of shape functions must be capable of approximating the solutions on the patch. Given a polynomial space, different families of shape functions can be constructed which all span the same space. A local matrix and a local right-hand side vector is attributed to each patch. The computation of these involves mapping from reference elements.
3. The shape functions of neighboring patches are glued together to build the discrete test and trial functions v_i. These functions must be tangentially continuous across the patch boundaries to be a global basis function of the solution space for Maxwell's equations. One DoF is attributed to each v_i.
4. Having obtained all test and trial functions v_i, the assembly process is carried out as in 1D, see Equations 52.67 and 52.68. The scattering of local to global matrices depends on the neighborhood relation of patches.

Algorithm 52.3: Basic Finite Element Algorithm

```
construct a tessellation $\bar{\Omega} = \cup_i \bar{K}_i$
for all patches $K_i$ do
    map $\widehat{K} \to K_i$
    compute local matrices $\mathbf{A}_i^{\text{loc}}$
    compute local vectors $\mathbf{b}_i^{\text{loc}}$
    scatter the local matrices and vectors to global ones
    update global matrices $\mathbf{A}$ and right-hand side $\mathbf{b}$
end for
solve $\mathbf{A}\mathbf{u}_h = \mathbf{b}$
return $\mathbf{u}_h$
```

52.5.6.1 Finite Elements

Figure 52.14 recalls the construction of the global test and trial functions from the local shape functions. The process is the same in higher dimensions, but there are many more possibilities for how to do it.

We illustrate these possibilities by means of some examples in 2D. First, we consider a mesh built from quadrilaterals, see Figure 52.15a. For a technical simplification of finite element algorithms, each quadrilateral is considered to be a map from a unit square, see Figure 52.16. The test and trial functions are constructed over the unit square and then mapped to their original positions. The lowest order trial functions are linear functions, in the case of a unit square, we use simply tensor products of 1D linear functions, see Section 52.5.4.

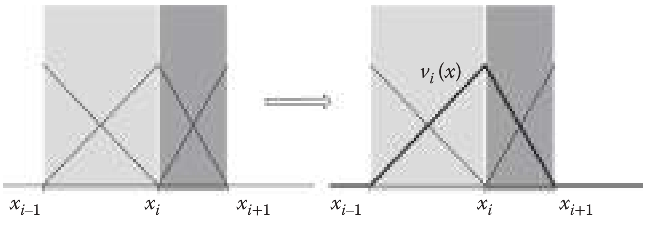


FIGURE 52.14 Construction of a global trial function v_i. Two neighboring intervals with two linear shape functions, each have the common boundary point x_i. One of the shape functions of each interval is taken and glued to the other and continued to a global, continuous trial function with local support.

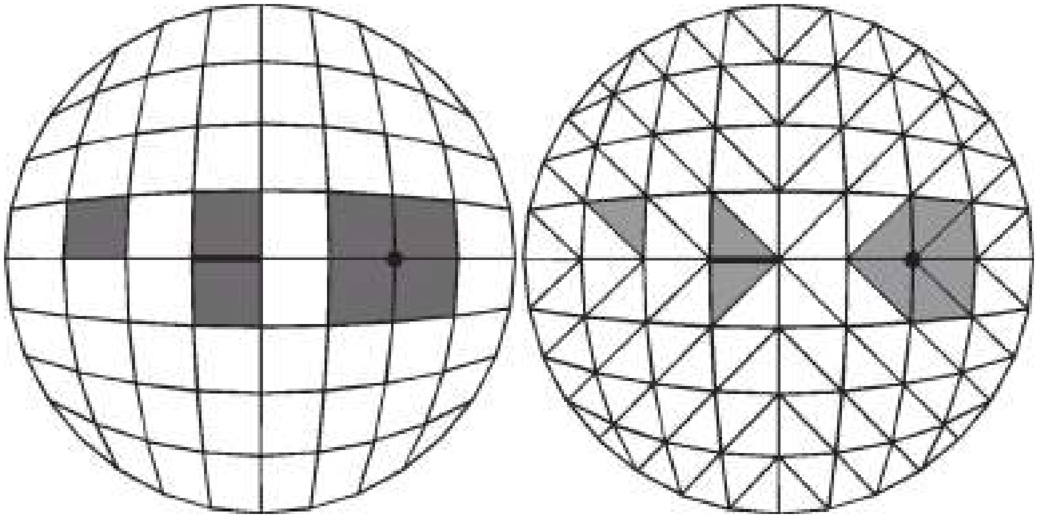

(a) (b)

FIGURE 52.15 Structured mesh based on quadrilaterals (a) and a triangular mesh with the same vertices (b). The gray shaded areas show the geometric possibilities of constructing trial functions based on the interior of a single patch, neighboring patches with a common edge and neighboring patches with a common vertex.

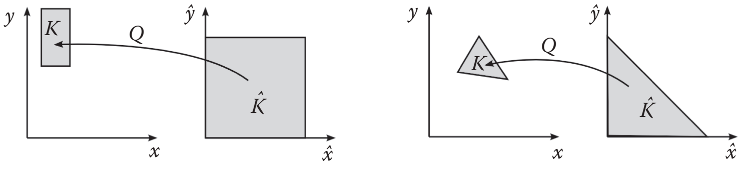


FIGURE 52.16 Mappings from reference elements to desired shapes.

52.5.6.2 Scalar Helmholtz Equation

Scalar trial functions v_i are used to approximate the solution of Helmholtz equations (Equation 52.8). In the case of a quadrilateral mesh with a unit square as reference element, the tensor product of the 1D shape functions gives four bilinear shape functions:

$$\{N_i \,:\, i = 1, \dots, 4\} = \{(1-\hat{x})(1-\hat{y}), \hat{x}(1-\hat{y}), \hat{x}\hat{y}, (1-\hat{x})\hat{y}\}.$$

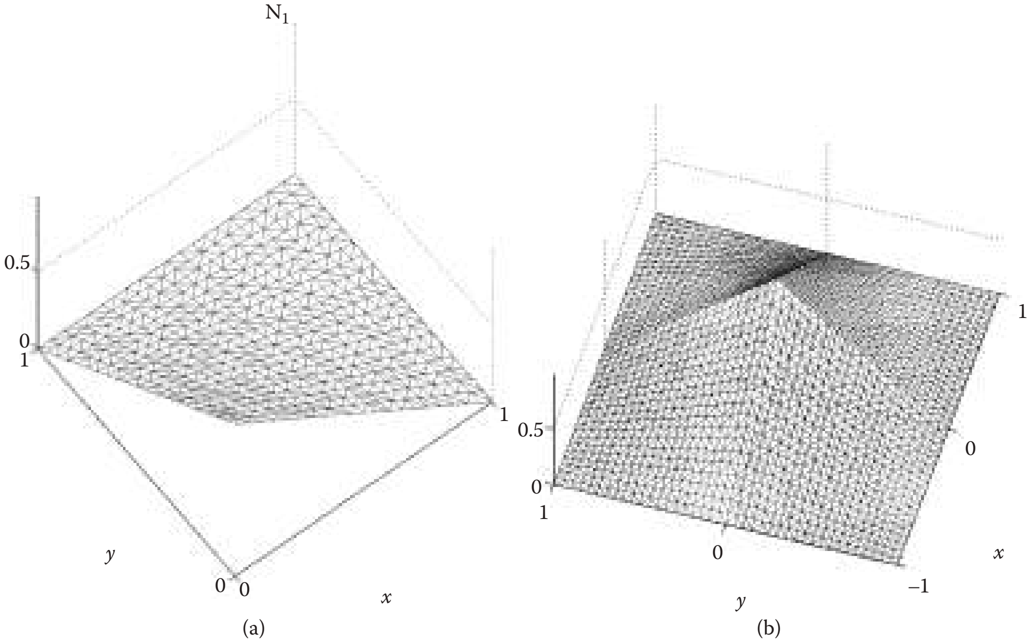


FIGURE 52.17 Bilinear shape function N_1 on a unit square and the composed trialfunction.

Figure 52.17a shows the shape function N_1, which is associated to the vertex $(0, 0)$ of the reference coordinate system $\widehat{K}$. In the spirit of the 1D construction (Figure 52.14), neighboring shape functions will be connected to continuous, global trial functions v_i as shown in Figure 52.17b. The common vertex is the defining feature of the resulting global trial function. The DoF attributed to this function is usually taken as its nodal value at the vertex position.

52.5.6.3 Maxwell's Equations

In the case of Maxwell's equations, the fields are generally no longer continuous. They are tangentially continuous but may have jumps or even singularities in normal field components at material discontinuities. Hence, nodal values, at least along material discontinuities, do not represent the physical situation properly. Therefore, the lowest order vectorial shape functions for Maxwell'equations are so-called edge functions. They are constructed in a way that they allow for an easy extension to tangential continuous, global trial functions. The four lowest order edge functions attributed to the unit square are shown in Figure 52.18.

52.5.6.4 Computation of Local Matrices and Assembly Process

This is very similar to the procedure discussed for the 1D Helmholtz problem Section 52.5.4. According to the Maxwell scattering formulation (Equation 52.71), we have to evaluate the variational formulation (Equation 52.71) on a finite dimensional test and trial space V_h. The computation of the local matrices and vectors goes the same way as in Section 52.5.4 with

$$a_{i,j} = a\left(\mathbf{v}_i, \mathbf{v}_j\right) = \underbrace{\int_{\Omega_{\mathrm{cd}}} dx\,dy\,\left(\nabla \times \mathbf{v}_i^*\right) \cdot \mu^{-1}\left(\nabla \times \mathbf{v_j}\right)}_{\text{stiffness part } s_{ij}} - \underbrace{\omega^2 \int_{\Omega_{\mathrm{cd}}} dx\,dy\,\mathbf{v}_i^* \varepsilon \mathbf{v}_j}_{\text{mass part } m_{ij}}$$

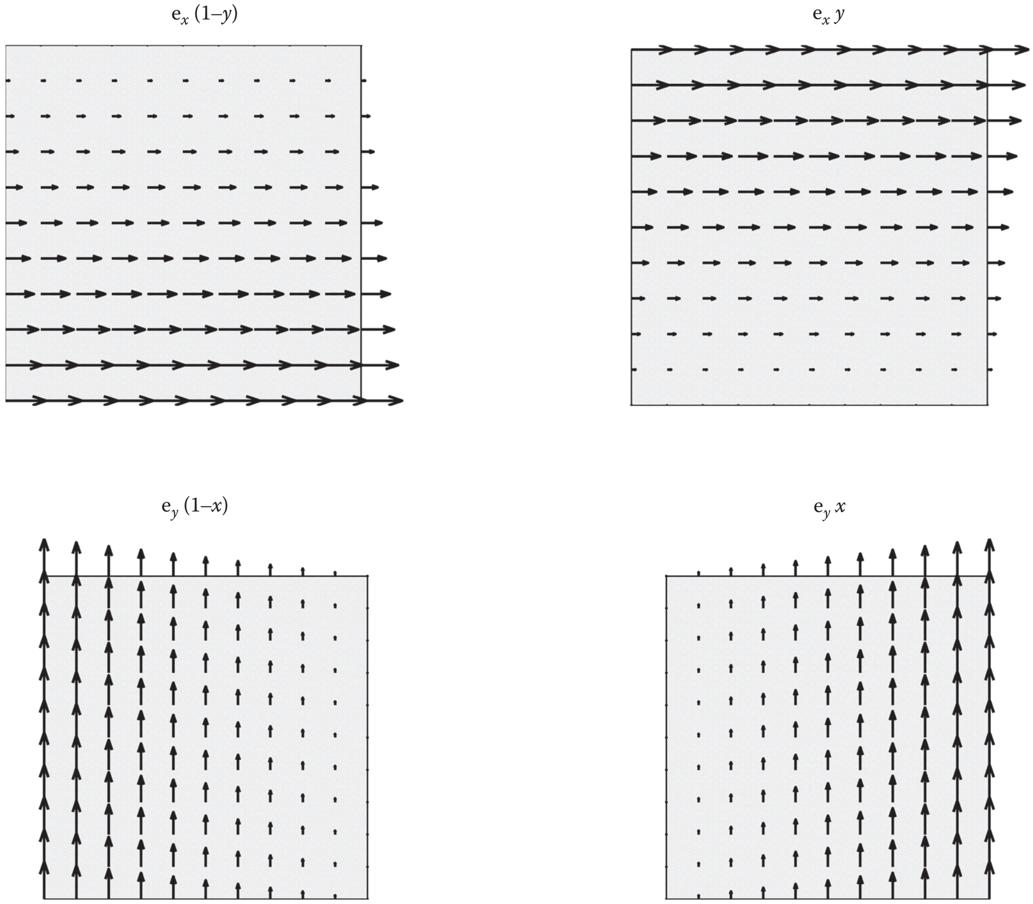


FIGURE 52.18 Vectorial, first-order edge elements of the unit square.

and

$$(b)_i = \int_{\partial\Omega} d\mathbf{f}\cdot\left(\mathbf{v}_i^{*}\times\left[\mu^{-1}\nabla\times\mathbf{E}_{\text{source}}\right]\right) + a_{\text{ext}}\left(\mathbf{v}_i,\mathbf{g}\right). \tag{52.72}$$

52.5.7 Convergence

The convergence analysis for edge element discretizations of time-harmonic scattering problems yields very useful insights into the behavior of numerical codes. The analysis itself is a challenging task since even in the lossless case the variational equations are neither elliptic nor definite. As a typical example we refer to Monk [34], who investigated a lossless scattering problem with sources on a bounded domain with electric boundary conditions. The error estimation uses the **curl** -norm

$$\|\mathbf{E}-\mathbf{E}_h\|_{\mathbf{curl}}^2 := \|\mathbf{E}-\mathbf{E}_h\|^2 + \|\nabla\times(\mathbf{E}-\mathbf{E}_h)\|^2,$$

where $\|\cdot\|$ is the standard L^2-norm, h denotes again a typical meshsize and as subscript it indicates discrete approximation. Depending on the properties of the computational domain and the smoothness of the exact solution there are constants C and δ independent of h and a constant $h_0 > 0$ independent of $\mathbf{E}$ and $\mathbf{E}_h$ such that for all $0 < h < h_0$

$$\|\mathbf{E} - \mathbf{E}_h\|_{\mathbf{curl}} \leq \frac{1}{1 - Ch^{1/2+\delta}} \inf_{\mathbf{v}_h \in V_h} \|\mathbf{E} - \mathbf{v}_h\|_{\mathbf{curl}}.$$

This result shows that for h less than a threshold value h_0, the FEM scheme converges to the exact solution, but in the worst case without a polynomial rate. However, if the solution is smooth quantified as $\mathbf{E} \in H^s(\mathbf{curl}\,;\Omega)$ and s is a smoothness parameter, then, we obtain in fact the desired polynomial convergence if we use high-order finite elements

$$\|\mathbf{E} - \mathbf{E}_h\|_{\mathbf{curl}} \leq Ch^s.$$

52.6 DG Method

The FDTD method has become the method of choice in physics and engineering when time-dependent Maxwell's equations have to be solved. The reasons are its simplicity, robustness and sufficient accuracy for most practical problems. On the other hand, it has severe limitations. These are mainly the difficultie incorporating complicated geometry due to its link to staggered grids unless the grid is prohibitively fine, the inherent second-order accuracy in time and space. Both topics have been addressed successfully in the frame work of FDTM methods, see the section about FDTD. However, DG methods present a successful alternative with the two main features geometrical flexibility and high-order approximations both in time and space inherently anchored in the method from the beginning.

52.6.1 Historical Notes

As any other method described in this chapter, the DG method has a long history with many contributions from many sides. Reed and Hill [35] applied a first form of the DG method to the neutron transport equation in 1973.

Since then, it has undergone a fast development and finds numerous applications to e.g., the Euler equations of gas dynamics, the shallow water equations, the equations of magneto-hydrodynamics, the compressible Navier–Stokes equations, and Maxwell's equations only to mention a few; see e.g., [36] for an overview and analysis of many of these applications.

The combination of the DG method for spatial discretization and of an explicit Runge–Kutta scheme for time-integration has been introduced by Cockburn and Shu in several papers, [37] maybe being one of the latest. [38] gives an introduction to the DG method and contains a more complete list of references on the (RKDG) method; it can be downloaded online. We also mention [39–41], which is a review paper about RKDG methods, as references on the DG method. [42] gives a review of the DG method applied to nanophotonics.

52.6.2 Concept: Combine Adaptive Spatial FEM Discretization with the Time Treatment in Finite Volume Methods

The FEM treats successfully complicated geometries in a natural way. There is no basic obstacle in combining local high-order approximations on unstructured meshes composed from different elements like tetrahedrons, prisms, bricks, and pyramids. The difficulties here lie on the side of implementations and is linked to the complexity of the problem: nontrivial geometries have to be described, decomposed into geometric patches, a detailed and efficient bookkeeping has to be realized for fast traversing. A direct application of FEM to time-dependent problems are possible but has a severe drawback: due to the occurrence of the mass matrix in conjunction with the time-derivative, an inversion of the mass matrix is needed. This can be done in several ways; nevertheless, it causes a problem not present in FDTD methods. The reason for the necessity to invert a mass matrix lies in the construction of finite elements. Typically, neighboring elements couple to each other in a way that the resulting system matrices are sparse but yield a global coupling of DoF which does not break down to structures which are cheaply invertible.

Finite volume methods, on the other hand, offer explicit time-integration schemes like those in FDTD methods. The key idea is to localize most of all spatial discretization into cells which are treated completely independent from each other. The cells are connected to each other by an explicit kind of continuity statement. This connection has to satisfy natural requirements like continuity or convergence properties but is not derived from first principles. There are many choices which can be adapted to given situations. The most intriguing feature is that by construction, each cell "sees" its own time evolution which results in an explicit scheme. The key idea of DG methods now is to combine this kind of time-treatment with a high-order spatial discretization localized to individual cell which are connected only by explicitly forced interface conditions on the cell boundaries.

52.6.3 Scalar Advection Equation as Model Equation: The Finite Volume Method in 1D

Consider the scalar advection equation

$$\begin{aligned} \partial_t u + \partial_x f(u) &= 0 \\ f(u) &= au, \quad a \text{ is a real constant}, \\ u(x,0) &= u_0(x), \quad x \in \mathbb{R}, \end{aligned} \tag{52.73}$$

where $\partial_x f(u)$ is the 1D model for divergence in higher dimensions $\nabla \cdot \mathbf{F}(\mathbf{u})$. The solution to Equation 52.73 is $u(x,t) = u_0(x - at)$. If $a > 0$, this is a right traveling wave, if $a < 0$, this is a left traveling wave. In order to analyze the stability of the FVM described below, we consider energy conservation. We multiply the wave equation (Equation 52.73) with u and integrate over a subdomain $\Omega_i = (x_i, x_{i+1})$:

$$\begin{aligned} \frac{d}{dt}\frac{1}{2}u^2 + \frac{d}{dx}\frac{1}{2}au^2 &= 0 \\ \frac{d}{dt}\int_{x_i}^{x_{i+1}} \frac{1}{2}u^2\,dx + \int_{x_i}^{x_{i+1}} \frac{d}{dx}\frac{1}{2}au^2\,dx &= 0 \\ \frac{d}{dt}\|u\|^2_{\Omega_i} &= -a\left(u^2\left(x_{i+1}\right) - u^2\left(x_i\right)\right). \end{aligned} \tag{52.74}$$

52.6.3.1 Direction of Information Flow

From Equation 52.74 we deduce the direction of information flow and stability under perturbation. Let $a > 0$, the wave is right propagating. At a fixed time, we consider the function $u(x)$ on the interval (x_i, x_{i+1}). Since (Equation 52.73) is linear, we can subtract a constant such that we obtain a new $u(x)$ with $u(x_i) = 0$. When we consider this value as a given zero boundary condition at the left side, we read from (Equation 52.74) $\frac{d}{dt}\|u\|^2_{\Omega_i} \leq 0$ independently from the right value. This right value might be the result of a computation, hence it might be contaminated by an error. Independent of contamination, there is no increase in the norm of u, hence a perturbation does not lead to instability. If we had chosen the right boundary to be the place for the exact boundary condition, the opposite would be the case. There would always be an increase in $\|u\|^2_{\Omega_i}$ in time.

52.6.3.2 Link to Finite Volume Methods

The whole computational domain Ω is decomposed into small control volumes $\bar{\Omega} = \cup\bar{\Omega}_i$, $\Omega_i \cap \Omega_j = 0$ for $i \neq j$. We consider one subdomain Ω_i and use the conservation form of Equation 52.73 applying Gauss's theorem:

$$\partial_t \int\limits_{\Omega_i} u\, dV + \int\limits_{\partial\Omega_i} f dS = 0, \tag{52.75}$$

where in higher dimensions the surface integral takes the form $\int\limits_S \mathbf{f} \cdot \mathbf{n}\, dS$ and $\mathbf{n}$ is the outwardly directed unit normal vector. Next, let the surface $\partial\Omega_i$ of cell Ω_i be subdivided into surfaces S_k which connect cell i to neighboring cells. Let $|\Omega_i|$ be the volume of Ω_i. We introduce the cell average u_i as an unknown target quantity

$$\partial_t u_i + \frac{1}{|\Omega_i|} \sum_k \int\limits_{S_k} f dS = 0 \tag{52.76}$$

$$u_i = \frac{1}{|\Omega_i|} \int\limits_{\Omega_i} u\, dV. \tag{52.77}$$

On the 1D mesh this results in the numerical scheme:

$$\partial_t u_i + \frac{1}{|\Omega_i|} \left(f_{i+1/2} - f_{i-1/2}\right) = 0; \tag{52.78}$$

see Figure 52.19. As in the FDTD method, the solution is based on a combination of quantities defined in primal meshes (the function u_h at integer positions) and on dual meshes (the discrete flux f_h at half counting positions). The main question is how do we define the values of $f = au$ at positions $x_{i-1/2}$ and $x_{i+1/2}$?

The answer depends, as discussed above, on the direction of wave propagation. It is natural to select the points in the solution at current time that are "upwind" of the solution at the position i for increasing time to maintain the causal connection. Depending then on the direction in which the solution is translated, we have different numerical schemes.

52.6.3.3 Finite Volume Method with First-Order Upwind Scheme

If $a > 0$, there is propagation in the positive x-direction. It holds $u_{i-1/2} \approx u_{i-1}$ and $u_{i+1/2} \approx u_i$, hence $f_{i+1/2} = au_i, f_{i-1/2} = au_{i-1}$, and the scheme (Equation 52.78) would be

$$\partial_t u_i + \frac{a}{\Delta x} \left(u_i - u_{i-1}\right) = 0.$$

In contrast, if $a < 0$, there is propagation in the negative x-direction. It holds $u_{i-1/2} \approx u_i$ and $u_{i+1/2} \approx u_{i+1}$, hence $f_{i+1/2} = au_{i+1}, f_{i-1/2} = au_i$, and the scheme (Equation 52.78) would be

$$\partial_t u_i + \frac{a}{\Delta x} \left(u_{i+1} - u_i\right) = 0.$$

Let us fix position $x_{i+1/2}$. The approximation of the flux just used is called the numerical flux, f^{num},

$$f^{\text{num}}_{i+1/2}\left(u_i, u_{i+1}\right) = \begin{cases} au_i & \text{if } a > 0 \\ au_{i+1} & \text{if } a < 0. \end{cases} \tag{52.79}$$

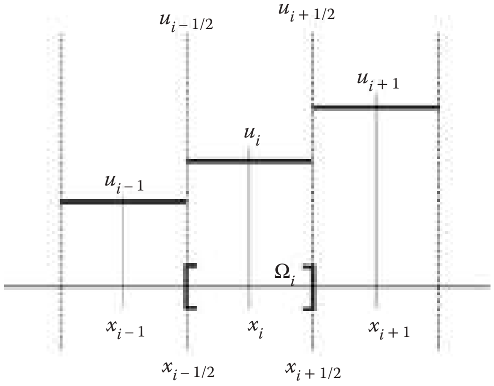


FIGURE 52.19 Vertex-centered finite volume method in 1D.

This can be compactly denoted as

$$f^{\text{num}}_{i+1/2}\left(u_i, u_{i+1}\right) = a \underbrace{\frac{1}{2}\left(u_i + u_{i+1}\right)}_{:=\{u\}_{i+1/2}} + \frac{|a|}{2} \underbrace{\left(u_i - u_{i+1}\right)}_{:=[\![u]\!]_{i+1/2}}$$

$$= a\{\!\{u\}\!\}_{i+1/2} + \frac{|a|}{2}\,[\![u]\!]_{i+1/2}\,.$$

We describe the typical finite volume discretization and give explicitly the 1D scheme:

1. Subdivide the domain into finite control volumes Ω_i. In one space dimension this is $\Omega_i = \left(x_{i-1/2}, x_{i+1/2}\right)$.
2. The conservation law is computed independently on each individual control volume Ω_i (52.76)

   $$\partial_t u_i = -\frac{1}{|\Omega_i|} \sum_k \int_{S_k} f dS$$

   with u_i the approximation to the average (Equation 52.77)

   $$u_i = \frac{1}{|\Omega_i|} \int_{\Omega_i} u\, dV.$$

   The resulting upwind scheme is, with flux (Equation 52.79),

   $$\begin{aligned}\partial_t u_i &= -\frac{1}{|\Omega_i|}\left(f_{i+1/2} - f_{i-1/2}\right)\\ &= -\frac{1}{|\Omega_i|}\left(f^{\text{num}}_{i+1/2}\left(u_i, u_{i+1}\right) - f^{\text{num}}_{i+1/2}\left(u_{i-1}, u_i\right)\right).\end{aligned} \tag{52.80}$$

3. Solve the semidiscrete system (Equation 52.80) with a suitable time-integrator. There are many possibilities, the most often used scheme in conjunction with time-dependent Maxwell's equations are low-storage Runge–Kutta methods. The resulting scheme after discretization in time must be a so-called monotonic scheme. Monotonic numerical schemes are ones which, given an initial distribution which is monotonic, produce a monotonic distribution after advection. A consequence of this property is that monotonic schemes neither create new extrema in the solution nor amplify existing extrema.

52.6.3.4 DG Method in 1D

The goal is to replace the low-order (piecewise constant) solution by a high-order approximation. We consider cell Ω_i. Exactly as in the FEM, we multiply (Equation 52.73) with a test function $v \in V$:

$$\int_{\Omega_i} v\partial_t u\,dx + \int_{\Omega_i} v\partial_x f(u)\,dx = 0.$$

An integration by parts gives

$$\int_{\Omega_i} v\partial_t u\,dx - \int_{\Omega_i} \partial_x v f(u)\,dx = -\int_{\partial\Omega_i} vf(u)\,ds.$$

Again as in FEM, we obtain the numerical method in *weak form* if we restrict the general space V to a finite dimensional space (of possibly high order) $V_h \subset V$:

$$\int_{\Omega_i} v_h\partial_t u_h\,dx - \int_{\Omega_i} \partial_x v_h f(u_h)\,dx = -\int_{\partial\Omega_i} v_h f^{\mathrm{num}}(u_h)\,ds \quad \text{for all } v_h \in V_h. \tag{52.81}$$

The corresponding strong version results, if we undo the partial integration:

$$\int_{\Omega_i} v_h\partial_t u_h\,dx + \int_{\Omega_i} v_h\,\partial_x f(u_h)\,dx = \int_{\partial\Omega_i} v_h\left(f\left(u_h\right) - f^{\mathrm{num}}(u_h)\right)\,ds \quad \text{for all } v_h \in V_h. \tag{52.82}$$

52.6.4 Maxwell's Equations in Conservation Form

We generalize the above step by step and present the framework of DG for Maxwell's equation's in 3D. The matrix-valued flux function $\mathbf{F} := \left(\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3\right)$ is composed of three column vectors of length n, $\mathbf{F}_i : \Omega \subset \mathbb{R}^3 \to \mathbb{R}^n$, with $i = 1, 2, 3$, and $\mathbf{F}_i = \left(F_{1i}, \ldots, F_{ni}\right)^T$. The divergence $\nabla \cdot \mathbf{F}$ of this matrix-valued field $\mathbf{F} : \Omega \subset \mathbb{R}^3 \to \mathbb{R}^{n\times 3}$ is defined as

$$\nabla \cdot \mathbf{F} := \begin{pmatrix} \partial_{x_1} F_{11} & +\partial_{x_2} F_{12}+ & \partial_{x_3} F_{13} \\ & \vdots & \\ \partial_{x_1} F_{n1} & +\partial_{x_2} F_{n2}+ & \partial_{x_3} F_{n3} \end{pmatrix}.$$

The generalization of the 1D advection equation (Equation 52.73) reads

$$\begin{aligned} \mathbf{Q}\partial_t u + \nabla \cdot \mathbf{F}(\mathbf{u}) &= 0 \\ u(x,0) &= u_0(x), \quad x \in \mathbb{R}^3. \end{aligned} \tag{52.83}$$

Next, we bring Maxwell's equations into this conservation form. From

$$\begin{pmatrix} \varepsilon & \\ & \mu \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix} = \begin{pmatrix} & \nabla\times \\ -\nabla\times & \end{pmatrix} \begin{pmatrix} \mathbf{E} \\ \mathbf{H} \end{pmatrix}$$

we read

$$\mathbf{Q} = \begin{pmatrix} \varepsilon & \\ & \mu \end{pmatrix} \qquad \mathbf{F}_i = \begin{pmatrix} -\mathbf{e}_i \times \mathbf{H} \\ \mathbf{e}_i \times \mathbf{E} \end{pmatrix}$$

with $\mathbf{e}_i$ being one of the three Cartesian unit vectors. Hence, there is in fact a conservation form (Equation 52.83) of Maxwell's equations.

52.6.5 Concept of DG-Algorithms for 3D Maxwell's Equations

The 3D algorithm follows the 1D concept. The only missing link is the upwind flux for the situation of Maxwell's equations. Following the same strategy as in 1D, Hesthaven and Warburg [43] derive the following expression:

$$\mathbf{n} \cdot \left(\mathbf{F} - \mathbf{F}^{\text{num}}\right) = \begin{pmatrix} \bar{Z}^{-1}\mathbf{n} \times \left(-Z^{+} [\![\mathbf{H}]\!] + \mathbf{n} \times [\![\mathbf{E}]\!]\right) \\ \bar{Y}^{-1}\mathbf{n} \times \left(-Y^{+} [\![\mathbf{E}]\!] + \mathbf{n} \times [\![\mathbf{H}]\!]\right) \end{pmatrix}. \tag{52.84}$$

Here Z and Y are the wave impedance and admittance, respectively,

$$Z^{\pm} = \sqrt{\frac{\mu^{\pm}}{\varepsilon^{\pm}}}, \quad Y^{\pm} = \left(Z^{\pm}\right)^{-1}, \quad \bar{Z} = \frac{Z^{+} + Z^{-}}{2}, \quad \bar{Y} = \frac{Y^{+} + Y^{-}}{2}.$$

The superscript "−" denotes the limit from the interior to the cell boundary, and the superscript "+" the limit from the neighboring cell to the boundary.

Algorithm 52.4: Basic Discontinuous Galerkin Algorithm

```
1. Discretization in space
construct a tessellation $\bar{\Omega} = \cup_i \bar{\Omega}_i$
for all patches $\Omega_i$ do
   map $\widehat{K} \to K_i$
   compute local matrices on local volumes
   assemble them (independently) to a global matrix
   compute local matrices on interior interfaces
   connect them to the global matrix
   compute local matrices on boundary interfaces
   connect them to build up the right-hand side
end for
2. Time evolution
use a time integrator (e. g., Runge–Kutta, leap-frog) to solve the semidiscrete, block-diagonal system in time
return  vector of field values and fluxes in time
```

52.6.6 Convergence

We consider Maxwell's equations with homogeneous, isotropic materials and a convex domain $\Omega \subset \mathbb{R}^3$ and arbitrary smooth initial data. This is a simplified model case which has no singularities. In this case optimal L^2-convergence for a fixed time interval is obtained [44]:

$$\left\|\mathbf{u} - \mathbf{u}_h\right\| \leq Ch^{p+1}.$$

This is the same as we would have, under the same conditions, for finite elements.

Symbol Definition

$\mathbf{r}^0$	Vector of unit length.
Ω	Computational domain.
$\partial\Omega$	Boundary of Ω.
$\mathbf{E}_\perp$	$\mathbf{e}_x E_x + \mathbf{e}_y E_y$
$\nabla_\perp$	$\mathbf{e}_x \partial_x + \mathbf{e}_y \partial_y$
∇_{k_z}	$\mathbf{e}_x \partial_x + \mathbf{e}_y \partial_y + ik_z \mathbf{e}_z$
$\iota_{e,f}$	incidence number edge-face
c	speed of light
$\mathbf{k}$	wavevector
$\mathbf{r}$	position
$\llbracket f \rrbracket$	Töplitz matrix generated by f
$\llbracket \cdot \rrbracket$	jump across a face
$\{\!\{ \cdot \}\!\}$	average across a face

References

1. Andrei V Lavrinenko, Jesper Lagsgaard, Niels Gregersen, Frank Schmidt, and Thomas Søndergaard. *Numerical Methods in Photonics*. Boca Raton, FL: CRC Press, 2014.
2. Junji Yamauchi. *Propagating Beam Analysis of Optical Waveguides*. Electronic and Electrical Engineering Research Studies. Optoelectronics and Microwaves Series. Exeter, UK: Research Studies Press, 2003.
3. Jean-Claude Nédélec. *Acoustic and Electromagnetic Equations: Integral Representations for Harmonic Problems*, Vol. 144. Springer Science & Business Media, 2013.
4. Richard A Norton and Robert Scheichl. Planewave expansion methods for photonic crystal fibres. *Applied Numerical Mathematics*, 63:88–104, 2013.
5. Jean-Pierre Berenger. A perfectly matched layer for the absorption of electromagnetic waves. *Journal of Computational Physics*, 114(2):185–200, 1994.
6. Saul Abarbanel and David Gottlieb. On the construction and analysis of absorbing layers in CEM. *Applied Numerical Mathematics*, 27(4):331–340, 1998.
7. Stephen D Gedney. An anisotropic perfectly matched layer-absorbing medium for the truncation of FDTD lattices. *IEEE Transactions on Antennas and Propagation*, 44(12):1630–1639, 1996.
8. Weng Cho Chew and William H Weedon. A 3D perfectly matched medium from modified Maxwell's equations with stretched coordinates. *Microwave and Optical Technology Letters*, 7(13):599–604, 1994.
9. Fernando L Teixeira and Weng C Chew. General closed-form PML constitutive tensors to match arbitrary bianisotropic and dispersive linear media. *IEEE Microwave and Guided Wave Letters,* 8(6):223–225, 1998.

10. Kane S Yee. Numerical solution of initial boundary value problems involving Maxwell's equations in isotropic media. *IEEE Transactions on Antennas Propagation*, 14(3):302–307, 1966.
11. Thomas Weiland. A discretization model for the solution of Maxwell's equations for six component fields. *Archiv Elektronik und Uebertragungstechnik*, 31:116–120, 1977.
12. Adi Ditkowski, K Dridi, and Jan S Hesthaven. Convergent Cartesian grid methods for Maxwell's equations in complex geometries. *Journal of Computational Physics*, 170(1):39–80, 2001.
13. Jan S Hesthaven. High-order accurate methods in time-domain computational electromagnetics: A review. *Advances in Imaging and Electron Physics*, 127:59–125, 2003.
14. Allen Taflove. *Advances in Computational Electrodynamics, the Finite-Difference time Domain*. Boston: Artech House, 1998.
15. Alain Bossavit. 'Generalized finite differences' in computational electromagnetics. *Progress in Electromagnetics Research*, 32:45–64, 2001.
16. Markus Clemens and Thomas Weiland. Discrete electromagnetism with the finite integration technique. *Progress in Electromagnetics Research*, 32:65–87, 2001.
17. Allen Taflove and Susan C Hagness. *Computational Electrodynamics: The Finite Difference Time-Domain Method*, 3rd edition. Boston, MA: Artech House Publishers, 2005.
18. Anders Bondeson, Thomas Rylander, and Par Ingelstrm. *Computational Electromagnetics*, Vol. 51. New York, NY: Springer Verlag, 2005.
19. Peter Monk and Endre Süli. A convergence analysis of Yee's scheme on nonuniform grids. *SIAM Journal on Numerical Analysis*, 31(2):393–412, 1994.
20. Theodor Tamir, HC Wang, and AA Oliner. Wave propagation in sinusoidally stratified dielectric media. *IEEE Transactions on Microwave Theory and Techniques*, 12(3):323–335, 1964.
21. C Yeh, KF Casey, and ZA Kaprielian. Transverse magnetic wave propagation in sinusoidally stratified dielectric media. *IEEE Transactions on Microwave Theory and Techniques*, 13(3):297–302, 1965.
22. MG Moharam and TK Gaylord. Rigorous coupled-wave analysis of planar-grating diffraction. *Journal of the Optical Society of America*, 71(7):811–818, 1981.
23. Philippe Lalanne and G Michael Morris. Highly improved convergence of the coupled wave method for tm polarization. *JOSA A*, 13(4):779–784, 1996.
24. G Granet and B Guizal. Efficient implementation of the coupled-wave method for metallic lamellar gratings in tm polarization. *Journal of the Optical Society of America A*, 13(5):1019–1023, 1996.
25. Lifeng Li. Use of Fourier series in the analysis of discontinuous periodic structures. *Journal of the Optical Society of America A*, 13(9):1870–1876, 1996.
26. Jens Küchenmeister. Three-dimensional adaptive coordinate transformations for the Fourier modal method. *Optics Express*, 22(2):1342–1349, 2014.
27. Peter Monk. *Finite element Methods for Maxwell's Equations*. Oxford University Press, 2003.
28. Leszek Demkowicz. *Computing with hp-Adaptive Finite Elements: One and Two Dimensional Elliptic and Maxwell Problems*, Vol. 1. CRC Press, 2006.
29. Jean-Claude Nédélec. Mixed finite elements in R^3. *Numerische Mathematik*, 35(3):315–341, 1980.
30. Jean-Claude Nédélec. A new family of mixed finite elements in R^3. *Numerische Mathematik*, 50:57–81, 1986.
31. Alain Bossavit. Whitney forms: A class of finite elements for three-dimensional computations in electromagnetism. *IEE Proceedings A (Physical Science, Measurement and Instrumentation, Management and Education, Reviews)*, 135(8):493–500, 1988.
32. Jon P Webb. Hierarchal vector basis functions of arbitrary order for triangular and tetrahedral finite elements. *IEEE Transactions on Antennas and Propagation*, 47(8):1244–1253, 1999.
33. Jan Pomplun, Sven Burger, Lin Zschiedrich, and Frank Schmidt. Adaptive finite element method for simulation of optical nano structures. *Physica Status Solidi (B)*, 244(10):3419–3434, 2007.
34. Peter Monk. A simple proof of convergence for an edge element discretization of Maxwell's equations. In *Computational Electromagnetics*, pp. 127–141. Springer, 2003.

35. WH Reed and TR Hill. Triangular mesh methods for the neutron transport equation. Report LA-UR-73-479. Los Alamos, NM: Los Alamos Scientific Laboratory, 1973.
36. Bernardo Cockburn, George E Karniadakis, and Chi-Wang Shu. *The Development of Discontinuous Galerkin Methods*. Lecture notes in computational science and engineering. Berlin, Heidelberg: Springer-Verlag, 2000.
37. Bernardo Cockburn and Chi-Wang Shu. The Runge–Kutta discontinuous Galerkin method for conservation laws V: Multidimensional systems. *Journal of Computational Physics*, 141(2):199–224, 1998.
38. Bernardo Cockburn. An introduction to the discontinuous Galerkin method for convection-dominated problems. In *Advanced Numerical Approximation of Nonlinear Hyperbolic Equations*, pp. 151–268. Berlin Heidelberg: Springer, 1998.
39. Bernardo Cockburn, George E Karniadakis, and Chi-Wang Shu. *Discontinuous Galerkin Methods: Theory, Computation and Applications*. Berlin Heidelberg: Springer, 2000.
40. Bernardo Cockburn and Chi-Wang Shu. Runge–Kutta discontinuous Galerkin methods for convection-dominated problems. *Journal of Scientific Computing*, 16(3):173–261, 2001.
41. Jan S Hesthaven and Tim Warburton. *Nodal Discontinuous Galerkin Methods: Algorithms, Analysis, and Applications*. Springer Science & Business Media, 2007.
42. Kurt Busch, Michael Koenig, and Jens Niegemann. Discontinuous Galerkin methods in nanophotonics. *Laser & Photonics Reviews*, 5(6):773–809, 2011.
43. Jan S Hesthaven and Timothy Warburton. Nodal high-order methods on unstructured grids: I. Time-domain solution of Maxwell's equations. *Journal of Computational Physics*, 181(1):186–221, 2002.
44. Marcus J Grote, Anna Schneebeli, and Dominik Schtzau. Interior penalty discontinuous Galerkin method for Maxwell's equations: Optimal l2-norm error estimates. *IMA journal of Numerical Analysis*, 28(3):440–468, 2008.

Index

A

Absorber relaxation time, 212–213
Absorber-to-gain cross section ratio, 214
Absorption dip, 724
Absorption spectrum, 220, 319, 364–367, 504
ACML, *see* Anticolliding mode-locking
Active ML, 184, 185, 187
Active nonlinear optical waveguide/cavity properties, 678–680
Additive pulse ML, 184
AEI, *see* Anisotropic exchange interaction
AM ML, *see* Amplitude-modulation ML
Amorphous silicon (a-Si), 498–499
Amplification, 87, 187, 215, 680
Amplified spontaneous emission (ASE), 84, 92, 623
Amplitude-modulation ML (AM ML), 183, 226
Andronov–Hopf bifurcation, 197–198
Angle of incidence (AOI), 353, 389, 392, 421, 478
Anisotropic exchange interaction (AEI), 703
Antibunching, 662
Anticolliding mode-locking (ACML), 216
Antireflection coating (ARC), 393, 446, 479, 480 485, 487
AOI, *see* Angle of incidence
APDs, *see* Avalanche photodiodes
ARC, *see* Anti reflection coating
Artificial interfaces, 747
ASE, *see* Amplified spontaneous emission
Auger recombination, 6
  process, 453, 542, 545, 550, 552
  rate, 317
Avalanche photodiodes (APDs), 295–297, 633

B

BA, *see* Broad-area
Backward differentiation formula (BDF) methods, 763
Backward propagation, 65, 73
Bandgap edge, 715–716, 723
Band structure, 249–250, 338, 363, 424–429, 529, 695, 776
Beam parameter product (BPP), 59, 60, 82, 83
Beam propagation factor, 82
Beam propagation method (BPM), 41–42, 65, 310, 314–315, 674, 808
Beam propagation ratio, 59–61
Beam quality, 59, 61, 66, 76, 82–83
Beam spoilers, 64
  GG tapered laser with, 68–72
  GG tapered laser without, 73–75
Beer–Lambert law, 482, 485, 504, 505, 510
Bernoulli function, 741, 757
Bessel functions, 271
BHJ, *see* Bulk heterojunction
Bidirectional symmetric filter, 135
Bipolar junction transistor (BJT), 300
Blakemore approximation, 757, 758
Bloch spin vector equation of motion, 667
Bloch theorem, 561, 566, 567
Boltzmann approximation, 52, 735, 740, 746
  numerical charge carrier fluxes, 756
Boltzmann transport equation (BTE), 244, 451, 779
Box–Müller method, 708
BPM, *see* Beam propagation method
BPP, *see* Beam parameter product
Bragg grating, 16, 32, 116, 117, 119, 156, 159, 209
Bragg stop-band width, 117–118
Bragg waveguide gratings, 37
Bragg waveguide SLM laser, 141–144
Brillouin zone (BZ), 695
Broad-area (BA)
  devices, 59
  lasers, 66, 85–86
BTE, *see* Boltzmann transport equation
Built-in potential, 736, 748
Built-in voltage, 737, 749
Bulk heterojunction (BHJ), 449, 502
Butler–Volmer equation, 527
BZ, *see* Brillouin zone (BZ)

C

Cadmium telluride (CdTe) cells, 497, 499–501
Carrier capture, 34–35, 210, 451, 454

Carrier density, 10, 71, 128, 163–165, 744
Carrier dynamics, in quantum-dot systems
  carrier–carrier Coulomb interaction, 643–646
  carrier–phonon interaction, 642–644
  carrier population functions, 641
  Fermi–Dirac distribution functions, 639, 640
  scattering process, 640
Carrier leakage, 8, 9, 19, 20
Carrier rate equations, 167–169
Carrier reservoir (CR), 167–169
Carrier-selective contacts, 451–452
Cartesian coordinate system, 265, 273, 566
Cascading scheme, 236, 240
Casimir effect, 662
Catastrophic optical mirror damage (COMD) threshold, 85
Cavity field quantization, 665
Cavity modes, 41–42, 96, 144, 623, 809
  in Lang–Kobayashi (LK) model, 178
  resonant, 676
Cavity quantum electrodynamics (CQED), 594–596, 665–666
CdTe cells, *see* Cadmium telluride cells
CFL conditions, *see* Courant–Friedrichs–Lewy conditions
Charge carrier escape, 449–451
Chiral carbon nanotubes (CNTs), optical rotation in
  chirality and optical activity, 712
  density of states, 716
  dipole optical intersubband transitions, 713, 715
  energy dispersion, 713–714
  Faraday rotation and magneto-chiral effects, 721–724
  single-walled CNTs, 712–713, 717
Chirality, 712–714, 717
CIC scheme, *see* Cloud-in-cell scheme
CIGS cells, *see* Copper indium gallium diselenide cells
Circularly polarized electric-field components, 704–705
Circularly polarized optical-pulse interactions
  degenerate four-level systems, 695–699
  two-level quantum systems, 692–695
Circularly polarized ultrashort pulse, 715, 720, 724
Cleaved-coupled-cavity (C^3) laser, 136–137
Cloud-in-cell (CIC) scheme, 792, 796
CMOS devices, *see* Complementary metal-oxide-semiconductor devices
Coherent control, semiconductor nanostructures, 664
Coherent Maxwell-pseudospin equations, 673–674
Coherent population
  effects, 225–226
  trapping, 724
Coherent pulse propagation, 664
Coherent ray tracing, 93
  laser diodes simulation tool coupling with, 94–95
Coherent spin manipulation, 701–707
Coherent vector Maxwell-pseudospin formalism, 717
Colliding pulse mode-locking (CPM), 216
COMD threshold, *see* Catastrophic optical mirror damage threshold
Complementary metal-oxide-semiconductor (CMOS) devices, 773, 775, 776
Complex refractive index, 320, 421–422
Compound cavity, 205, 216
Compressively strained multiple quantum wells (CS-MQW), 127
Computed band edge energy, 744
Concentrating photovoltaics (CPV), 410
Confinement factor, 84
Contact terminal currents, 763–765
Continuous wave (CW)
  excitation, 638
  operation, 8, 16
  simulation tool, 82
Continuous-wave-excited lasers, 638–639
Control volumes, 753–754
Conventional lasers, 629
Conversion efficiency, 23
Copper indium gallium diselenide (CIGS) cells, 497, 500, 501
Coupled-cavity SLM lasers
  band-pass filtering model, 133–136
  cleaved-coupled-cavity laser, 136–137
  discrete mode laser, 139
  etched slotted laser, 137–139
  general optical filtering model, 133–136
  governing equations, 139–140
  performance comparisons among, 140–141
Coupled laser systems, 170–172
Coupled Schrödinger–Poisson solver, 240, 458
Coupling losses, 20, 25, 69, 193, 203
Courant–Friedrichs–Lewy (CFL) conditions, 821–822
CPM, *see* Colliding pulse mode-locking
CPV, *see* Concentrating photovoltaics
CQED, *see* Cavity quantum electrodynamics
CR, *see* Carrier reservoir
Cross-talk heating effects, 163
Crystalline silicon solar cells
  models for, 403–406
  numerical simulation of, 402–403
CS-MQW, *see* Compressively strained multiple quantum wells
CW, *see* Continuous wave

D

Dark current, 338–341
DBR, *see* Distributed Bragg reflector
DBR-tapered laser, *see* Distributed Bragg reflector-tapered laser
DDE model, *see* Delay-differential equation model
Decimation technique, 220, 223–225

Defect voltage, 18, 19, 22
Delaunay triangulation, 761–762
Delay-differential equation (DDE) model, 154, 193–206
Density-dependent nonlinear factor, 751
Density matrix model, 239
Density of states (DOS), 222, 242, 342, 403, 442, 513, 563, 749–750
Dephasing time, 226, 239, 664, 669, 679, 708, 710
Detector QE modeling, 347–348
  edge coupling, for quantum well infrared photodetectors, 351–353
  InAs/GaInSb IR photodiode, 350–351
  modeling validation, 348–349
  photonic crystal slab quantum well infrared photodetectors, 355–356
  planar solar cell, 349–350
  plasmonic-enhanced quantum well infrared photodetectors, 353–355
  quantum grid infrared photodetector, 353
Device-level electrothermal simulation, 247–249
  band structure, 249–250
  heat generation rate, 251–252
  J-F curve, 250
  modal gain, 250–251
  results, 252–254
  threshold, 250–251
DFB laser, *see* Distributed feedback laser
DG method, *see* Discontinuous galerkin method
DG transistor, *see* Double-gate transistor
Dielectric permittivity, 209, 218, 422, 717, 735
Diffusion enhancement, 751
Diffusion time, of carriers, 288
Diffusion voltage, 395
Dipole length scale, 672
Dipole matrix element, 250, 668, 689, 703, 705, 723
Direct-diode laser systems, 81, 82, 86, 87, 106
Direct gap semiconductors, 285
Dirichlet boundary conditions, 800–801
Discontinuous galerkin (DG) method, 808, 809
  algorithm, 849
  convergence analysis, 849–850
  finite volume methods, 844–847
  in one dimensional, 848
  for three dimensional Maxwell's Equations, 848–849
Discrete mode laser, 139
Discrete nonlinear Poisson problem, 738–739
Discretizing van Roosbroeck system
  control volumes, 753–754
  existence, uniqueness, and convergence, 759
  finite volume method in higher dimensions, 754–756
  flux expressions with thermodynamic equilibrium, 758
  free energy and dissipation rate, 758–759
  maximum principle, 759
  Scharfetter–Gummel fluxes and non-Boltzmann generalizations, 756–758
  2D simulation, 753–754
Dispersion, 38, 821–822
Dispersive element, 189, 223
Displacement current density, 54, 764
Dissipation rate, 751
Dissipative dynamics, 669–671
Distributed Bragg reflector (DBR), 60, 261, 563, 571, 595–596
  laser, 12, 16, 118–120, 226
  mirror, 676
Distributed Bragg reflector (DBR)-tapered laser, 87, 89, 91
  analysis and discussion, 102–105
  parasitic reflections with off-axis alignment, 102
  parasitic reflections with on-axis alignment, 97–102
  unintentional reflections on, 95–97
Distributed feedback (DFB) laser, 12, 16, 111, 120–125
Distributed time-domain, 206–210
Doping, 250, 296, 346, 755
  optimization, 767
Double-gate (DG) transistor, 776
Double-trench resonant tunneling SLM laser, 144–146
Drift-diffusion calculation, 423–424
  interface recombination velocities, 429–430
  intrinsic carrier concentration, 424–429
  tunnel diodes, 430–431
Drift-diffusion models, 309, 530, 543–544
  contact terminal currents, 763–765
  definition of, 511
  density of states, 514–516
  finite element method, 765–766
  finite volume method, 761–763
  hopping process, 517, 518
  nonlinear solvers, 759–761
  1D van Roosbroeck Model, 735–736
  parabolic approximation, 514
  semiconductor device equations, 512–513
  thermodynamic formulation, 517
  time discretization, 763
  van Roosbroeck system, *see* van Roosbroeck system
Drift velocity, 289
Drude–Lorentz model, 610–611
DSC, *see* Dye-sensitized solar cell
Dyadic Green's function, 591, 592, 665
Dye-sensitized solar cell (DSC), 500–502, 527–528

E

EA, *see* Evolutionary algorithm
EAM, *see* Electroabsorption modulator
EBL, *see* Electron-beam lithography
Edge coupling, for quantum well infrared photodetectors, 351–353

Edge-emitting lasers, 36, 133, 193, 202, 261–263, 563–564, 631
Edge-emitting semiconductor laser diodes, 110, 125, 153
Effective interface roughness, 244–245
Efficiency analysis, 8–9
Einstein relation, 403, 423, 513, 514, 516, 742
Einstein summation convention, 567
Electrical energy, 395–396
Electrical phenomena, 262
Electric displacement flux, 755–756
Electricfield envelope, 684
Electroabsorption modulator (EAM), 187, 364
Electroluminescent (EL) refrigeration
  active layer thickness, 551–552
  boundary conditions, 546–547
  cooling cycling for, 542–543
  device structure and equations, 543–545
  doping types and concentrations on, 552–553
  feasibility of, 541–542
  GaAs-based, 542, 557–558
  leakage current and carrier blocking layer, 549–551
  light extraction, 553–554
  photon recycling, 554–557
  recombination and generation processes, 545–546
  steady-state analysis, 547–549
Electromagnetic theory, 477, 478
Electron-beam lithography (EBL), 118
Electron emission, of longitudinal optical phonon, 236
Electron-hole pairs, 3, 7
  photogeneration of, 445–449
Electronic bandstructure solver, 240–241
Electronic nanostructures, in solar cell, 481
Electronic transport, 238
  in mid-infrared quantum cascade lasers, 237
  nonequilibrium phonons, ensemble Monte Carlo with, 240–243
  quantum techniques, 239–240
  semiclassical techniques, 238–239
  in terahertz quantum cascade lasers, 237
Electron leakage current, 33–34
Electron spin-flip rate, 701
Electro-optic effect modulators, 371–372
  Mach–Zehnder modulator, 372–374
  traveling-wave design, 374
Electrorefractive modulator, 372, 374–375
EL refrigeration, *see* Electroluminescent refrigeration
EMC, *see* Ensemble Monte Carlo
Energy conservation, 53–55
Energy-level diagram, two-level quantum systems, 692
Energy yield, 409
Ensemble Monte Carlo (EMC), 239, 779, 799
  with nonequilibrium phonons, 240
Entanglement, 663, 665
EQE, *see* External quantum efficiency
Etched slotted laser, 137–139
Evolutionary algorithm (EA)
  description of, 616–617
  flowchart, 617
  zero group velocity points, 617–618
Exciton dissociation, 449
External cavity simulation tool, 93
External differential efficiency, 26
External quantum efficiency (EQE), 288, 398, 433, *See also* Slope efficiency

F

Fabry–Pérot (FP) laser, 16, 28, 111–113
  cavity, 5, 6
  edge-emitting, 203
Faraday rotation, 721–724
Fast-axis collimation (FAC) lens, 96
Fast Fourier transform (FFT), 315
FCA, *see* Free-carrier absorption
FCCC model, *see* Full cavity coherent coupling model
FCIC model, *see* Full cavity incoherent coupling
FD-SOI device technology, *see* Fully-depleted silicon-on-insulator device technology
FDTD, *see* Finite-difference time-domain (FDTD)
FEM, *see* Finite element method
Fermi–Dirac distribution function, 504, 513, 546, 639, 640, 748, 749
Fermi's golden rule, 242, 446, 504, 592
Fermi voltage, 18, 29, 34
FET, *see* Field-effect transistor
FF, *see* Fill factor
FFT, *see* Fast Fourier transform
Field-effect transistor (FET), 774
Filamentation
  longitudinal-lateral approximate projected equations, 42–43
  nonlinearities due to virtual transitions, 44–45
  nonlinearity induced by lateral spatial hole burning, 45–47
Fill factor (FF), 398
Final field equation, 39–41
FinFET transistor, 776
Finite-difference method, *see* Shooting method
Finite-difference time-domain (FDTD)
  algorithmic features and drawbacks, 817
  convergence, 823
  discrete differential operators, 818–820
  discrete div-conditions, 822–823
  dispersion and stability, 821–822
  historical notes, 818
  leapfrog scheme in time, 821
  method, 313–314, 674, 676, 680, 808, 809
  nanoplasmonic laser, 618–620
  numerical implementation, 683–684
  single CNT, 717

Finite element method (FEM), 765–766
    convergence analysis, 843–844
    historical notes, 833
    in one dimensional, 834–839
    patchwise tangential continuous and polynomial spaces, 833–834
    in two and three dimensional, 839–843
Finite volume method (FVM), 737
    mesh generation
        boundary conforming Delaunay property, 761–762
        open problems, 762–763
        tensor product approaches, 762
    in one dimensional, 845–847
    time treatment in, 844–845
First-principles model, 127
FK effect, *see* Franz–Keldysh effect
Flared unstable cavity lasers, *see* Tapered lasers
Flicker noise, 292
Fluorine-doped tin oxide (FTO), 501
FMM, *see* Fourier modal method
FO, *see* Fourier optics
Fourier modal method (FMM), 809
    convergence, 832–833
    historical notes, 823–824
    layered media stack, 826–827
    scattering-matrix approach, 825–826
    segmentwise waveguide treatment, 824
    transfer-matrix approach, 824–826
    in two dimensional, 828–832
Fourier optics (FO), 93
Fourier transform, 36, 240, 315, 348, 676, 680, 710, 712, 809
Fox–Li approach, 42
FP laser, *see* Fabry–Pérot laser
Franz–Keldysh (FK) effect, 363, 365
Free-carrier absorption (FCA), 6, 8, 33, 84, 365
Free energy, 517, 751, 758–759, 763
FTO, *see* Fluorine-doped tin oxide
Full cavity coherent coupling (FCCC) model, 65
Full cavity incoherent coupling (FCIC) model, 65
Full discretized van Roosbroeck system, 742–743
Full width at half maximum (FWHM), 133, 250
Fully-depleted silicon-on-insulator (FD-SOI) device technology, 800, 802
Fully microscopic quantum-kinetic models, 459–461
FVM, *see* Finite volume method
FWHM, *see* Full width at half maximum

G

Gain compression, 4, 169–170
Gain dispersion, 187, 207–208
    and group velocity dispersion parameters, 214
    model of, 158–161
    self-phase modulation in, 215
Gain-guided (GG) tapered laser
    with beam spoilers, 68–72
    feature, 61–62
    without beam spoilers, 73–75
Gain-to-absorber saturation energy ratio, 191
Galerkin-type methods, 836
GaN/AlGaN PIN photodiode, 327–331
Gate contact, 747, 762, 800
Gate operations, 663
Gauss–Fermi integral, 766
Gaussian beam, 61, 68, 77
Gaussian density of states (GDOS), 515
Gaussian function, 68, 723
Gaussian noise, 708
Gauss's law, 266, 367, 820, 846
GDOS, *see* Gaussian density of states
Generalized finite difference method, 818
Generation-recombination (GR) noise, 292
GFs, *see* Green's functions
GG tapered laser, *see* Gain-guided tapered laser
Global detailed balance theories, for ideal systems, 454–457
Gm, *see* Modal gain
Governing equations, 17–19
Green's functions (GFs), 643
    dyadic, 591, 592, 665
    nonequilibrium, 238–240
Green's identity, 53
GR noise, *see* Generation-recombination noise
Group velocity dispersion (GVD), 187, 214
Gummel map, 753
Gummel's method, nonlinear solvers, 760–761

H

Half-cavity (HC) model, 65
Hanbury, Brown, and Twiss (HBT)
    experiment, 585–586
    measurement, 633, 634
Harmonic oscillator, 662
    potentials, 699–700
Hartree–Fock factorization, 637–638
Haus theory, 190, 196–197
HBT, *see* Hanbury, Brown, and Twiss
HCGs, *see* High-contrast gratings
HC model, *see* Half-cavity model
Heat sources, spatial distributions of, 51–52
HED, *see* Horizontal electric dipole
Heterojunctions, incorporation of, 303
HgCdTe-Based p-n junction photodiode, 323–327
High-contrast gratings (HCGs), 575–576
    grating mirrors, cavities with, 578–580
    manufacturing sensitivity of, 580–582
    subwavelength reflectivity, 576–579

High-power laser diodes, 81
  beam quality, role of, 82–83
  external cavity simulation tool, 93
  lateral cavity design, 85–86
  power scaling, 82–83
  simulation tool coupling with coherent ray tracing tools, 94–95
  Speclase, 92–93
  tapered laser diode, 86–92
  vertical cavity design, 83–85
High-power lasers
  high-power broad-area semiconductor laser, 16
  nonlinear modes and filamentation
    longitudinal-lateral approximate projected equations, 42–43
    nonlinearities due to virtual transitions, 44–45
    nonlinearity induced by lateral spatial hole burning, 45–47
  optical field
    balance of radiative energy, 39–41
    beam propagation method, 41–42
    cavity modes, 41–42
    dispersion, 38
    final field equation, 39–41
    nonlinear susceptibility, 38–39
    roundtrip operator, 41–42
    spontaneous emission, 39
    three-dimensional equations, 35–37
  optical power
    basic relations, 24–26
    LSH and impact of series resistance and internal loss, 28–31
    nonthermal power saturation effects, 31–35
    Rigrod model, 26–28
    treat-power-as-a-parameter method, 28
  phenomenological model
    governing equations, 17–19
    parameters determination, 19–22
    simulated versus. measured CW characteristics, 23–24
  thermodynamic-based energy-transport model
    basic equations, 48–51
    energy conservation, 53–55
    spatial distributions of heat sources, 51–52
Hole spin-flip relaxation, 701
Homogeneously broadened (HB) two-level quantum system
  analytically computed gain coefficient, 720
  density-matrix method, 680
  gain saturation dynamics of, 681–682
  linear small-signal gain in, 681
  theoretical gain coefficient of, 724
Hong–Ou–Mandel (HOM) experiment, 586, 587
Horizontal electric dipole (HED), 600
Horn ridge waveguide (HRW) DFB laser, 146–147
Hot-charged exciton states, 701, 703–707
Huygen's integral, 42
Hybrid ML, 184, 187, 194
Hybrid models, 457–459

I

IAM, *see* Incident angle modifier
ICRs, *see* Intracavity reflectors
Ideal fill factor, 407
Ideal *I-V* model, 396–398
  ideal single-junction solar cells, 399
  quantum efficiency, 398–399
IG laser, *see* Index-guided laser
IIR, *see* Infinite impulse response
InAs/GaInSb IR photodiode, 350–351
Incident angle modifier (IAM), 392–393
Incident photocurrent efficiency (IPCE), 511
Index-guided (IG) laser, 60–62
  narrow tapered section, 75–76
Indirect gap semiconductors, 285–286
Indium tin oxide (ITO), 499–501
Infinite impulse response (IIR), 208
Infrared photodiodes
  HgCdTe-Based p-n junction photodiode, 323–327
  InP/$In_{0.53}Ga_{0.47}As$/InP PIN photodiode, 322–323
Injection barrier, 236
Injector level, 236, 239
Inorganic semiconductors, 529, 748, 749
InP/$In_{0.53}Ga_{0.47}As$/InP PIN photodiode, 322–323, 325
Instantaneous optical modes, 174–176
Integrated circuits (ICs), 774
Interface recombination velocities, 429–430
Internal loss, 28–31
Internal quantum efficiency (IQE), 398, 434, *see also* Quantum efficiency
Intersymbol interference (ISI), 111
Intracavity reflectors (ICRs), 216
Intrinsic carrier density/concentration, 424–429, 748
Inverse Fermi integrals, 29, 49
Ionization thresholds, 295
IPCE, *see* Incident photocurrent efficiency
IQE, *see* Internal quantum efficiency
ISI, *see* Intersymbol interference
ITO, *see* Indium tin oxide

J

J-F curve, 250, 252
Joule heat, 50, 52, 84, 105, 800
Junction capacitance effect, 289, 293, 296
J-V curve calculation, 418, 419, 431–432

K

Kerr lens ML, 184
Kinetic Monte Carlo (kMC) method, 530–531
  algorithm, 520–525
  1D disordered system by, 525–527
  infrequent-event system, 519
  transport models, hierarchy of, 518
k·p method, 240–241
Kramers–Kronig relationship, 127, 363–364

L

Lambert's law, 285
Langevin formalism, 707
LA phonons, *see* Longitudinal acoustic phonons
Large optical cavity (LOC) tapered laser, 82, 84
  ELoD2 structure, 85, 95, 96
Large-signal lumped iterative model, 203
Large-signal time-domain approach, 193–206
Laser diode
  efficiency analysis, 8–9
  high-power, *see* High-power laser diodes
  laser cavity designs, 12
  optical gain and optical loss, 3–6
  photonic crystal, *see* Photonic crystal laser diodes
  rate equation analysis, 9–12
  semiconductor laser diodes, 3
  temperature effects, 7–8
  threshold current and slope efficiency, 6–7
Laser geometry, 215–217
Lasing filaments, 42
Lasing frequency, 110
Lateral modes, intensity distributions of, 275–276
LCOE, *see* Levelized cost of electricity
LDOS, *see* Local density of states
Leakage currents, 33–34
LED, *see* Light-emitting diode
Levelized cost of electricity (LCOE), 410
Light-emitting diode (LED), 541–542
Light intensity behavior, of solar cells, 406–408
Light states, 662
$LiNbO_3$, 372
Linear electro-optic effect, *see* Pockel's effect
Linear finite elements, 837, 840–841
Liouville equation, 669–670
Local charge neutrality, 736–737
Local density of states (LDOS), 462–463, 488, 592–593
LOC tapered laser, *see* Large optical cavity tapered laser
Longitudinal acoustic (LA) phonons, 242, 642
Longitudinal-lateral approximate projected equations, 42–43
Longitudinal optical (LO) phonon, 236, 642
Longitudinal spatial horn burning (LSH/LSHB) effect, 28–31, 121
Long-wavelength/mid-wavelength (LW/MW) two-color HgCdTe, 331–334
LO phonon, *see* Longitudinal optical phonon
Lorentzian approximation, 159–160
Lorentzian form, 218
LSH/LSHB effect, *see* Longitudinal spatial horn burning effect
Luminescent coupling, 434–437
Lumped series resistance, 400, 407
LW/MW two-color HgCdTe, *see* Long-wavelength/mid-wavelength two-color HgCdTe
Lyapunov function, 751, 752

M

Mach–Zehnder modulator, 372–374
Macroscopic polarization vector, 671, 672
Magneto-chiral effects, 721–724
$MAPbX_3$, *see* Methylammonium lead halide perovskites
Masetti model, 425
Master pseudospin equation, 669–671
MA system, *see* Mode approximation system
Material absorption modeling
  material parameters, 343–344
  numerical examples, 344–347
Material gain dispersion, 158–161
Maxwell–Bloch approach, 610–613
Maxwell–Bloch equations, 166, 635, 707
Maxwell-pseudospin method
  coherent equations in 2D, 673–674
  coupling to pseudospin equations, 671–673
  dissipative dynamics and master pseudospin equation, 669–671
  numerical methodology, 674
  pseudospin equation, 667–669
  vector Maxwell equations, 671–673
Maxwell's equations, 264
  curl equations, 314
  edge functions, 842, 843
  homogeneous, 35
  perfectly matched layers, 814–817
  resonance and mode propagation problem, 811
  scattering problem, 810–811
  spatial discretization of, 821
  time-dependent, 808, 809
  transformation rules, 812–814
  two dimensional geometries and fields, 811–812
MBE, *see* Molecular beam epitaxy
MCPM, *see* Multiple colliding pulse mode-locking
MD, *see* Modal discrimination
MESFET, *see* Metal-semiconductor field-effect transistor
Mesh discretization, 419–420
Metal-organic chemical vapor deposition (MOCVD), 235

Metal-organic vapor phase epitaxy (MOVPE), 15
Metal-oxide-semiconductor field-effect transistor (MOSFET), 774, 803
Metal-semiconductor field-effect transistor (MESFET), 792, 793
Metal-semiconductor-metal photodetectors (MSM photodetectors), 298–299
Methylammonium lead halide perovskites ($MAPbX_3$), 503
Microcavity exciton–polaritons, 665
Microcavity pillar geometry, 595–596
Microcrystalline silicon (μc-Si), 498–499
Micropillar lasers, 629
Mid-infrared quantum cascade lasers (mid-IR QCLs), 235
  electron transport in, 237
mid-IR QCLs, *see* Mid-infrared quantum cascade lasers
Minority carrier lifetime, 315–317
MJSCs, *see* Multijunction solar cells
ML semiconductor lasers, *see* Mode-locked semiconductor lasers
MMM, *see* Mode-matching method
MOCVD, *see* Metal-organic chemical vapor deposition
Modal discrimination (MD), 84, 100
Modal gain (Gm), 250–251
Mode approximation (MA) system, 178–179
Mode intensity, 176, 269
Model-dependent generation-recombination rate, 748
Modeling cavity-emitter systems, 666, 707–712
Modeling coherent spin, in QDs
  hot-charged exciton states, 701, 703–707
  single-charged particles in waveguide, 699–701
Mode-locked (ML) semiconductor lasers
  absorber relaxation time, 212–213
  coherent population effects, 225–226
  delay-differential equations model, 193–206
  features, 185–187
  frequency and time-frequency treatment, 192
  gain dispersion
    and absorber compression coefficients, 215
    and group velocity dispersion parameters, 214
  general principles of, 183–185
  large-signal time-domain approach, 193–206
  laser geometry and saturable absorber location, 215–217
  microscopic input in, 217–222
  operating point, 210–212
  pumping current, 212
  self-consistent pulse profile, 187–192
  self-phase modulation, 215
  in single-section lasers, 226–227
  small-signal time-domain models, 187–192
  traveling-wave models, 206–210, 223–225
Mode-matching method (MMM), 115
Mode partition noise (MPN), 105, 111
Molecular beam epitaxy (MBE), 15, 235
  grown QD, 700
Monolithic diode lasers, 214
Monte Carlo algorithm, 523–525
Monte Carlo device simulations
  of different materials, 796–799
  electron-polar optical phonon scattering, 787
  electron transport, scattering tables for, 784
  energy and wavevector histograms, 791
  ensemble Monte Carlo techniques, 779
  final angle selection, 787–788
  flow chart description of, 786
  initial carrier distribution, 785
  modeling and simulation, 776–779
  particle-based device simulation, *see* Particle-based device simulation
  pseudo-random number generator, 780
  rejection technique and drift velocity, 789
  scattering rate expressions, 783
  self-scattering mechanism, 781–782
  steady-state drift velocity versus. electric field, 790
Moore's law, 774–775
More than Moore (MtM) devices, 775–776
MOSFET, *see* Metal-oxide-semiconductor field-effect transistor
MOVPE, *see* Metal-organic vapor phase epitaxy
MPN, *see* Mode partition noise
MQW, *see* Multiple quantum well
MSLs, *see* Multisection semiconductor lasers
MSM photodetectors, *see* Metal-semiconductor-metal photodetectors
MtM devices, *see* More than Moore devices
Muller's algorithm, 121
Multijunction solar cells (MJSCs)
  complex refractive index, 421–422
  drift-diffusion calculation, *see* Drift-diffusion calculation
  J-V curve calculation, 431–432
  luminescent coupling, 434–437
  mesh discretization, 419–420
  overview, 415–417
  QE calculation, 433–434
  quantum yield, 422–423
  simulation workflow, 418–419
  transfer matrix method, 420–421
Multiple colliding pulse mode-locking (MCPM), 216
Multiple quantum well (MQW), 235, 338, 366
Multisection semiconductor lasers (MSLs), 154, 163, 167–168
  and coupled laser systems, 170–172
  nontrivial MSL device, 172–174

N

Nanoplasmonic laser
  block diagram of, 610
  Drude–Lorentz model, 610–611
  evolutionary algorithm, 616–618
  finite-difference time-domain method, 618–620
  lasing dynamics, 623
  loss compensation, 620–622
  multilevel Maxwell–Bloch theory, 611–613
  stopped-light lasing, 623–625
  transfer matrix method, 613–616
Nanostructure solar cells (NSSC)
  charge separation, 449–452
  electron-hole pairs, photogeneration of, 445–449
  fully microscopic quantum-kinetic models, 459–461
  global detailed balance theories, for ideal systems, 454–457
  hybrid models, 457–459
  in photovoltaics, 441–444
  recombination, 452–454
  single quantum well, 461–465
Nanowire solar cells
  current density, 476
  efficiency evaluation, 479–481
  electronic model, 482–483
  material parameters, 493
  modal analysis, 485–487
  nonequilibrium, 477–479
  optical models, 481–482
  with p-n junction, 475–476
  Purcell effect, 488
  radial versus. axial junction arrangement, 489–491
  Shockley–Queisser efficiency, 487
  spectral absorptivity, 484–485
  surface recombination, 491–492
  thermal equilibrium, 477
Narrow broad-area (NBA) laser diodes, 86
Narrow index-guided tapered laser, 75–76
National Renewable Energy Laboratory (NREL), 498
Natural optical activity, *see* Optical activity
NBA laser diodes, *see* Narrow broad-area laser diodes
Nearest-element-cell (NEC) scheme, 792, 796
Nearest-grid-point (NGP) scheme, 792, 796
NEC scheme, *see* Nearest-element-cell scheme
NEGF, *see* Nonequilibrium Green's functions
NEP, *see* Noise equivalent power
Neumann boundary conditions, 528, 747, 748, 801
New's model, 195–197
Newton's method, nonlinear solvers, 760, 763
NGP scheme, *see* Nearest-grid-point scheme
N-level quantum system, 667, 670
Noise equivalent power (NEP), 289–290
Non-concentrated solar cells, 388–389
  irradiation, 389–390
  temperature, 390–391
Nonequilibrium Green's functions (NEGF), 238–240, 483
Nonequilibrium phonons
  electronic bandstructure solver, 240–241
  ensemble Monte Carlo with, 240
  transport kernel with, 241–243
Non-Gaussian beams, 61
Nonideal *I-V* model
  parallel resistance, 399–400
  recombination, 400–402
Nonlinear gain saturation, 165–167
Nonlinearity
  due to virtual transitions, 44–45
  gain compression, 31–32
  induced by lateral spatial hole burning, 45–47
  susceptibility, 38–39
Nonlinear modes
  due to virtual transitions, 44–45
  induced by lateral spatial hole burning, 45–47
  longitudinal-lateral approximate projected equations, 42–43
Nonlinear Poisson problem, 738–739
Nonlinear solvers
  Gummel's method, 760–761
  Newton's method, 760
Nonthermal power saturation effects, 31–35
Normalized saturated field intensity, 681
NREL, *see* National Renewable Energy Laboratory
NSSC, *see* Nanostructure solar cells

O

OBPF, *see* Optical band-pass filter
ODE, *see* Ordinary differential equation
Ohmic contacts, 28, 289, 350, 395, 418, 432, 735–737, 747, 752, 754, 794–795
One-dimensional (1D) laser simulator, 63
One-dimensional scalar approach, 266–270
Optical activity
  bandgap edge, 715–716
  chirality, 714
  cholesteric liquid crystals and sculptured thin films, 721
  circularly polarized ultrashort pulse, 715
  coherent vector Maxwell-pseudospin formalism, 717
  electric-field components, time evolution, 717–718
  Fourier spectra, 717–718
  linearly polarized pulse, 719
  polarization-resolved transmission spectra, 720
  rotation angle, 714
  wavevector, 720

Optical band-pass filter (OBPF), 134
Optical beam absorption method, 310–311
Optical feedback, 4, 12, 82, 98, 102, 153, 172, 205, 261
Optical field
  balance of radiative energy, 39–41
  beam propagation method, 41–42
  cavity modes, 41–42
  dispersion, 38
  final field equation, 39–41
  nonlinear susceptibility, 38–39
  roundtrip operator, 41–42
  spontaneous emission, 39
  three-dimensional equations, 35–37
Optical gain, 3–6
Optical generation models, 309–310
  beam propagation method, 314–315
  FDTD method, 313–314
  optical beam absorption method, 310–311
  raytracing, 311
  transfer matrix method, 312–313
Optical loss, 3–6
Optical microconcentrators, of solar cell, 480–481
Optical modulators, 363–364
  electro-optic effect, 371–374
  plasma-effect, 374
  QCSE-based, *see* QCSE-based modulators
  resonant, 377–378
  Si electroabsorption versus electrorefractive effect, 374–376
Optical phenomena, 262
Optical power
  basic relations, 24–26
  LSH and impact of series resistance and internal loss, 28–31
  nonthermal power saturation effects, 31–35
  Rigrod model, 26–28
  treat-power-as-a-parameter method, 28
Optical Rabi oscillations, 665, 691
Optical source, brightness of, 59
Optical waveguides and semiconductor microcavities
  active nonlinear optical waveguide/cavity properties, 678–680
  passive cavity properties, 676–678
  resonant nonlinearities and gain saturation, 681–683
  2D nonlinear optical waveguides, 683–687
  2D semiconductor microcavities, 687–691
  two-level quantum system, 675–676
OpticStudio, 93–95
OPV cells, *see* Organic photovoltaics cells
Ordinary differential equation (ODE), 154
Organic photovoltaics (OPV) cells, 500, 502, 528–529
Organic semiconductors, 748, 766
Oxide aperture diameter, impact of, 277–278

P

Parallel-mirror slab waveguide, 685–686
PARDISO, direct solvers, 760
Partial differential equations (PDEs), 154
Partially gain-coupled antiphase DFB, 122, 123, 133
Partially gain-coupled in-phase DFB, 122, 123
Partially loss-coupled antiphase-grating DFB, 122, 124
Partially loss-coupled in-phase DFB, 122, 124
Particle-based device simulation
  current calculation, 792, 794, 795
  flowchart of, 793
  ohmic contacts, 794–795
  particle-mesh coupling, 791–792, 796, 797
  thermal effects in silicon-on-insulator devices, 799–803
  time-step and mesh size, 795–796
Particle current flow, 48
Particle-mesh (PM) coupling, 791–792, 796, 797
Passive cavity properties, 676–678
Passive ML (PML), 184
PAT, *see* Pulse-area theorem (PAT)
PBCs, *see* Periodic boundary conditions
PBG, *see* Photonic bandgap
PC-EELs, *see* Photonic-crystal edge-emitting lasers
PCS QWIP, *see* Photonic crystal slab quantum well infrared photodetectors
PDA, *see* Planar dielectric antenna
PDEs, *see* Partial differential equations
PDF, *see* Probability density function
p-doped QDs, 701, 704
PDs, *see* Photodetectors
PECVD, *see* Plasma-enhanced chemical vapor deposition
Peltier heat, 50, 52, 54
Perfectly matched layer (PML), 37, 814–817
Periodic boundary conditions (PBCs), 240, 484, 565
Permittivity spectrum, 370
Perovskites-based solar cells, 500, 503, 529–530
PhC waveguide, *see* Photonic crystal waveguide
Phenomenological model
  governing equations, 17–19
  parameters determination, 19–22
  simulated versus measured CW characteristics, 23–24
Photoconductive detectors, 290–292
Photoconductive gain, 291, 341
Photocurrent, 341–342, 394
Photodetectors (PDs), 283, *see also* Quantum well infrared photodetectors
Photodiode
  p-i-n, 293–294
  p-n, *see* p-n junction photodiodes
  visible, 327

Photogeneration of electron-hole pairs
  absorption coefficient, 446–449
  photon flux, 445–446
Photoluminescent (PL) refrigeration, 541
Photon, 4
  blockade, 665
  density, 10
  flux, 445–446
Photon recycling, 433–437, 546–548, 553–558
Photonic bandgap (PBG), 563, 564
Photonic-crystal edge-emitting lasers (PC-EELs), 563–565
Photonic crystal laser diodes
  periodic structures, 561–563
  photonic bandgap, 563, 564
  photonic-crystal edge-emitting lasers, 563–565
  vertical-cavity surface-emitting lasers, *see* Vertical-cavity surface-emitting lasers
Photonic crystal slab quantum well infrared photodetectors (PCS QWIP), 355–356
Photonic crystal (PhC) waveguide, 598–599
Photonic nanowire geometry, 597–598
Photonics, numerical methods of, 808–809
Photon-number (Fock) states, 662
Photon recycling, 434–437, 546, 554–557
Phototransistors, 290, 300
Photovoltaics, nanostructure solar cells in, 441–444
Photovoltaic (PV) technologies, 497
Physical phenomena, 262–263
Picosecond pulse propagation, 664
p-i-n-diode, 744
p-i-n photodiode, 293–294
Planar dielectric antenna (PDA), 600
Planar GaAs solar cell, 349–350
Planar semiconductor microcavity, 683–684
Planck constant, 17, 156, 310, 434, 445, 504, 545, 749
Plane-wave admittance method (PWAM), 565–570
Plane-wave reflection transfer (PWRT), 569–570
Plasma-effect modulators, 363, 374
Plasma-enhanced chemical vapor deposition (PECVD), 499
Plasmonic-enhanced quantum well infrared photodetectors, 353–355
PL refrigeration, *see* Photoluminescent refrigeration
PM coupling, *see* Particle-mesh coupling
PML, *see* Passive ML; Perfectly matched layer
P-N junction photodiodes, 293
  absorption coefficient model, 318–320
  basic equations, 309
  beam propagation method, 314–315
  FDTD method, 313–314
  infrared photodiodes, 322–327
  minority carrier lifetime, 315–317
  optical beam absorption method, 310–311
  overwiew, 307–308
  physical mechanism, 320–322
  raytracing, 311
  transfer matrix method, 312–313
  trap concentration versus energy, 317–318
  two-color heterojunction photodiodes, 331–334
  ultraviolet photodiodes, 327–331
  visible photodiodes, 327
Pockel's effect, 363, 371
Poisson-drift-diffusion model, 367–368
Poisson equation, 48, 53
Poissonian statistics, 662
Polarization-decay, 669
Polarization density, 35, 38
Population relaxation time, 669
Power-current (P-I) characteristics, 19, 20, 23, 30, 66–67
Power saturation, 23
  laser parameters for, 30
Power scaling, 82–83
Probability density function (PDF), 521, 522
Pseudospin equation, 667–669
Pseudospin vector, 669
Pulse-area theorem (PAT), 664
Pulse center frequency, 704
Pumping current, 212
Purcell effect, 590, 594, 595, 602
  nanowire solar cells, 488
Purely gain-coupled DFB, 122, 124, 125, 128, 133
Purely refractive index-coupled uniform-grating DFB laser, 122, 123, 128–131
PV technologies, *see* Photovoltaic technologies
PWAM, *see* Plane-wave admittance method
PWRT, *see* Plane-wave reflection transfer

Q

QCLs, *see* Quantum cascade lasers
QCSE, *see* Quantum-confined Stark effect
QCSE-based modulators
  absorption spectrum, 367
  Franz–Keldysh effect, 365
  numerical results, 370–371
  permittivity spectrum, 370
  Poisson-drift-diffusion model, 367–368
  quantum-confined stark effect, 365–367
  Schrödinger equation, for QW states, 368–369
  structure, 364–365
QDIP, *see* Quantum dot infrared photodetector
QD nanolasers, *see* Quantum-dot nanolasers
QE, *see* Quantum efficiency
QED, *see* Quantum electrodynamics
QGIP, *see* Quantum grid infrared photodetector
Q-switching instability, 198, 203, 206, 211
Quantum cascade lasers (QCLs)
  lasing in, 236–237
  models, 237–238

Quantum cascade lasers (QCLs) (*Continued*)
  overview, 235–236
  recent developments in, 237
  thermal transport in, 238
Quantum coherent control, 664
Quantum-confined Stark effect (QCSE), 220–221
Quantum defect heat, 52
Quantum dot infrared photodetector (QDIP), 300–301
Quantum dot (QD) lasers, carrier rate equations, 167–169
Quantum-dot (QD) nanolasers
  carrier dynamics, *see* Carrier dynamics, in quantum-dot systems
  inhomogeneous broadening in, 651
  input-output curves and autocorrelation function, 646–647
  interemitter coupling effects and superradiance in, 647–651
  laser rate equations and β factor, 628–631
  lasing regime and threshold, 635
  nonresonant cavity feeding, 652–653
  photonic and electronic density of states, 632
  Purcell enhancement, 631
  semiconductor laser theory, *see* Quantum-optical semiconductor laser theory
  spontaneous emission, 627–628
  streak-camera measurements, 633
Quantum dots (QDs)
  antibunching, 662
  embedding, 665
  modeling coherent spin in, 699–707
  spontaneous emission
    Box–Müller method, 708
    intracavity electric-field build-up, 708–709
    output electric-field trace, 710–712
    refractive-index profile, 707–708
    semiconductor microcavity, electric field, 708–710
    slowly-varying envelope approximations, 710
    steady-state population inversion, 709, 711
Quantum efficiency (QE), 288, 322, 398–399
  calculation of, 433–434
Quantum electrodynamics (QED), 628
Quantum grid infrared photodetector (QGIP), 353
Quantum nanostructures, 286–287
  incorporation of, 303
Quantum-optical semiconductor laser theory
  continuous-wave-excited lasers, 638–639
  correlation-expansion technique, 635
  dynamical laser equations, 637–638
  semiclassical and fully quantum-mechanical approach, 635–637
Quantum photonic structures
  chiral carbon nanotubes, optical rotation in, 712–724
  Maxwell-pseudospin method, 667–674
  modeling cavity-emitter systems, 707–712
  modeling coherent spin in QDs, 699–707
  optical waveguides and semiconductor microcavities, 674–691
  quantum wells and QDs, 691–699
Quantum regression theorem, 602, 634, 666
Quantum stochastic formalism, modeling cavity-emitter systems
  Langevin formalism, 707
  QD spontaneous emission, 707–712
  quantum fluctuations, 707
  quantum noise, 707
Quantum techniques, 239–240
Quantum well infrared photodetectors (QWIPs), 300–301
  absorption coefficient, 342
  dark current, 338–341
  detector QE modeling, *see* Detector QE modeling
  material parameters, 343–344
  numerical examples, 344–347
  overwiew, 337–338
  photoconductive gain, 341
  photocurrent, 341–342
  resonator, *see* Resonator-quantum well infrared photodetector
Quantum wells (QWs), 286–287
  energy levels, 66
Quantum wells (QWs) and QDs
  circularly polarized optical-pulse interactions
    degenerate four-level systems, 695–699
    two-level quantum systems, 692–695
  optical orientation, 691
  Rabi flop, 691
Quantum yield, 422–423
Quarter-wavelength phase-shifted (QWPS) DFB, 122, 123, 125, 128–133
Quasi-3D (three-dimensional) model, 62–65
Quasi-Fermi level separation, 417, 478
Quasi-Fermi potentials, 735–736, 740
Quasi-monochromatic electromagnetic field, 109–110
QWIPs, *see* Quantum well infrared photodetectors
QWPS DFB, *see* Quarter-wavelength phase-shifted DFB
QWs, *see* Quantum wells

R

Radiative energy, balance of, 39–41
Radiative recombination, 452–453, 545
Radio frequency (RF) linewidth, 187
Raman effect, 44
Random-phase approximation (RPA), 643
Rate equation analysis, 9–12
Raytracing method, 310, 311
RCE structure, *see* Resonant-cavity-enhanced structure
RCWA, *see* Rigorous coupled wave analysis

Recombination phenomena, 263
Refractive index, 71, 169–170
  nonlinearities, 184
Refractive-index profile, 707–708
Representative elementary volumes, 753
Resonant-cavity-enhanced (RCE) structure, 302–303
Resonant modulators, 377–378
Resonant nonlinearities, 681–683
Resonator-quantum well infrared photodetector (R-QWIP)
  design, 356–358
  realization, 358–361
Responsivity, 288
RF linewidth, *see* Radio frequency linewidth
Ridge waveguide (RW)
  lasers, 59, 60, 68–75
  and tapered laser diode, 86–92
Rigorous coupled wave analysis (RCWA), 576, 823, 828
Rigrod model, 26–28
Robin boundary conditions, 747
Rotating-wave approximation (RWA), 664
Roundtrip operator, 41–42
RPA, *see* Random-phase approximation
R-QWIP, *see* Resonator-quantum well infrared photodetector
RW, *see* Ridge waveguide
RWA, *see* Rotating-wave approximation (RWA)

S

SA, *see* Saturable absorber
SAC lens, *see* Slow-axis collimation lens
SAGCM structure, *see* Separate absorption, grading, charge sheet, and multiplication structure
SAGM, *see* Separate absorption grading and multiplication
SAM, *see* Separate absorption and multiplication
SA Speclase, *see* Stand-alone Speclase
Saturable absorber (SA)
  implementation, 187
  location, 215–217
  mechanism, 225–226
  reverse bias, 186
  self-starting mechanism, 184
SB, *see* Schottky barrier
Scalar Helmholtz equation, 841–842
Scattering-matrix approach, 825–826
Scharfetter–Gummel fluxes, 756–758
Schottky barrier (SB), 298
Schrödinger equation, for QW states, 368–369
SCOWL diodes, *see* Slab-coupled optical waveguide laser diodes
SCP, *see* Self-consistent profile
Self-consistent profile (SCP), 187–192
Self-heating, 8, 799, 801–803
  due to free-carrier absorption/optical pumping, 101
  due to Joule heating, 100
Self-induced transparency (SIT), 664
  and polarized SIT soliton formation, 693–695
  soliton, 664, 685, 687
    polarized, formation, 693–695
Self-phase modulation, 184, 187, 215, 218
Semiclassical techniques, 238–239
Semiconductor diode lasers, 184
Semiconductor Industry Association (SIA), 774
Semiconductor laser
  brightness, 59
  devices, 155
  diodes, 3
Semiconductor luminescent refrigerator, 541, 542
Semiconductor optical amplifier (SOA), 60
Semiconductors, 284–285
Semiconductor saturable absorber mirror (SESAM) chip, 202
Separate absorption and multiplication (SAM), 296
Separate absorption grading and multiplication (SAGM), 296
Separate absorption, grading, charge sheet, and multiplication (SAGCM) structure, 296
Series resistance, 28–31
SESAM chip, *see* Semiconductor saturable absorber mirror chip
SHB, *see* Spatial hole burning
Shockley-Queisser model
  of nanowire array, 476, 481, 487
Shockley–Read–Hall (SRH), 401
  process, 128
  recombination, 6, 545
Shooting method, 369
Short-scale spatial hole burning, 209
SIA, *see* Semiconductor Industry Association
Side-mode-suppression-ratio (SMSR), 119
Si electroabsorption versus electrorefractive effect, 374–376
Silicon-based cells, 498–500
Silicon-on-insulator (SOI)
  technology, 775, 799–803
  waveguide, 120
Single-charged QD embedded in waveguide, 699–701
Single longitudinal mode (SLM) laser, 111–113
  Bragg waveguide SLM laser, 141–144
  characteristics, examples, 128–133
  classification, 113
  coupled-cavity lasers, *see* Coupled-cavity SLM lasers
  distributed Bragg reflector laser, 118–120
  distributed feedback lasers, 120–125
  double-trench resonant tunneling SLM laser, 144–146
  governing equations
    for design, modeling, and simulation, 125–128
    and solution technique, 139–140

Single-charged QD embedded in waveguide (*Continued*)
  grating analysis, 113–118
  horn ridge waveguide DFB laser, 146–147
  problems in, 141
Single-mode photon density, 10
Single-photon emission events, 662, 587, 662
Single-photon source (SPS)
  cavity quantum electrodynamics, 594–596
  classical dipole, 590–592
  coupling strength, 603–604
  definition of, 585
  efficiency, 587–588
  emitter-cavity master equation, 602–603
  indistinguishability, 586–587, 601
  photon extraction strategies, 590
  planar dielectric antenna, 600
  pure single-photon emission, 585–586
  quantum emitter master equation, 601–602
  simulation techniques, 593–594
  spectral wandering, 603
  spontaneous emission, 592–593
  spontaneous parametric down-conversion process, 588–590
  waveguide quantum electrodynamics approach, 596–599
Single quantum well (SQW), 461
  characteristics, 464–465
  dynamics, 463–464
  states, 462–463
Si P-Njunction photodiode, 327, 328
SIT, *see* Self-induced transparency
SL, *see* Superlattice
Slab-coupled optical waveguide laser (SCOWL) diodes, 85
SLM laser, *see* Single longitudinal mode laser
Slope efficiency, 6–9, 237
Slotboom variables, 752
Slow-axis collimation (SAC) lens, 96
Slowly-varying envelope (SVEA) approximations, 664, 710
Small-signal time-domain models, 187–192
Small wavelength chirp, 119
SMSR, *see* Side-mode-suppression-ratio
SOA, *see* Semiconductor optical amplifier
SOI, *see* Silicon-on-insulator
Solar cells optics, photocurrent of, 394
Solar spectrum, 387–388
Solid-state optical refrigeration, 541
Space-time folding technique, 223–225
Spatial hole burning (SHB), 62, 71, 163–165
  nonlinearity induced by, 45–47
Spatially inhomogeneous carrier density, 163–165
SPDC process, *see* Spontaneous parametric down-conversion process
Specific rotatory power, 712
Speclase, 92–95
Spectrally selective laser cavity, 216
Spin-B interaction energy, 722
Spin decoherence mechanisms, 703
Spin-filtering effect, 701, 706–707
Spin-polarized drift diffusion models, 766
Spontaneous emission, 39, 128
Spontaneous parametric down-conversion (SPDC) process, 588–590
SPP modes, *see* Surface plasmon polariton modes
SPS, *see* Single-photon source
SQW, *see* Single quantum well
SRH, *see* Shockley–Read–Hall
Stand-alone (SA) Speclase, 93
Standing wave-induced gratings, 209
Stark effects, 44
Stationary 1D semiconductor equations
  discrete nonlinear Poisson problem, 738–739
  discretize Poisson problem, 737–738
  1D van Roosbroeck system, 735–736
  full discretized van Roosbroeck system, 742–743
  linear electrostatic potential, 743–745
  Scharfetter–Gummel scheme, 739–742
  thermodynamic equilibrium and local charge neutrality, 736–737
Stationary ML equation, 189
Statistical distribution function, 735
Steady state, 176–178
  above threshold, 25–26
  around threshold, 25
  conditions, 10–11
Steady-state population inversion, 709, 711
Stranski–Krastanov mode, 651
Superlattice (SL), 244
Surface plasmon polariton (SPP) modes, 609, 616
SVEA, *see* Slowly-varying envelope approximations
Synchronous ML, 184

T

Tapered laser diode
  DBR-tapered laser, *see* Distributed Bragg reflector-tapered laser
  lateral beam quality of, 86–92
Tapered lasers
  fast axis, 60
  GG tapered laser, *see* Gain-guided tapered laser
  simulation examples
    comparison between devices, 76–78
    device geometries and simulation parameters, 67–68
    GG tapered laser with beam spoilers, 68–72
    GG tapered laser without beam spoilers, 73–75
    narrow index-guided tapered laser, 75–76

    simulation model
        calibration and comparison with experiments, 65–67
        physical effects, 63
        quasi-3D model, 62–65
        symmetry, 65
    slow axis, 60
Taper losses, 69, 75, 77
TAT, *see* Thermally assisted tunneling
TCAD, *see* Technology computer-aided design
TCO layer, *see* Transparent conductive oxide layer
Technology computer-aided design (TCAD), 777
TE cooling, *see* Thermoelectric cooling
TE field, *see* Transverse electric field
TEM field, *see* Transverse electromagnetic field
Temperature behavior, of solar cells, 408
Temperature effects, 7–8
Temperature range, 20, 245
Terahertz quantum cascade lasers (THz QCLs), 237, 239
Thermal black-body radiation source, 662
Thermal detuning, 161–163
Thermal lensing, 42, 47, 62
Thermally assisted tunneling (TAT), 338, 339, 450
Thermal phenomena, 263
Thermal transport, 243–244, 246–247
    device-level electrothermal simulation, 247–249
    in quantum cascade lasers, 238
    quantum cascade lasers active core, 245–246
    superlattice, 244
    twofold influence, of effective interface roughness, 244–245
Thermodynamic-based energy-transport model
    basic equations, 48–51
    energy conservation, 53–55
    spatial distributions of heat sources, 51–52
Thermodynamic equilibrium
    flux expressions with, 758
    and local charge neutrality, 736–737
Thermodynamic model, 309
Thermoelectric (TE) cooling, 542, 543
Thin-film technologies
    cadmium telluride-based cells, 499–501
    coherent and incoherent calculation, 509
    copper indium gallium selenide-based cells, 500, 501
    definition of, 497
    drift-diffusion model, *see* Drift-diffusion model
    dye-sensitized solar cell, 500–502, 527–528
    incident photocurrent efficiency, 511
    in-plane and orthogonal components, 507
    kinetic Monte Carlo method, *see* Kinetic Monte Carlo method
    morphology and grain effects, 530–531
    optical absorption, 498, 503–504
    organic heterojunction solar cell, 510
    organic photovoltaics cells, 502, 528–529
    perovskites-based cells, 503, 529–530
    propagation constant, 506
    silicon-based cells, 498–499
    transfer matrix method, 505, 508
Thomson heat, 50, 52
Three-dimensional equations, 35–37
Three-dimensional scalar approach, 273–277
Threshold current, 6–7
Threshold phonons, 250–251
THz QCLs, *see* Terahertz quantum cascade lasers
Time-bandwidth product, 184–185
Time-dependent effects, 9
Time-dependent Rabi frequencies, 703, 717
Time discretization, 763
Time-domain lumped models, 187
Time-harmonic versus time-dependent modeling, 808
Titanium dioxide (TiO_2), 501–503, 527, 528
TM field, *see* Transverse magnetic field
TMM, *see* Transfer matrix method
TPP method, *see* Treat-power-as-a-parameter method
Transfer matrix method (TMM), 312–313, 420–421, 482, 486, 676–677, 824–827
    nanoplasmonic laser, 613–616
Transistor scaling, 774–775
Transmission spectrum, 676, 678
Transparent conductive oxide (TCO) layer, 499, 500
Transport kernel, 240, 791, 799
    with nonequilibrium phonons, 241–243
Transverse electric (TE) field, 36, 614, 673–674
Transverse electromagnetic (TEM) field, 505
Transverse magnetic (TM) field, 36, 614, 673–674
Traveling-wave model (TWM), 154–155, 200, 374
    considerations, 206–210
    material gain dispersion, 158–161
    modifications of, 167–170
    nonlinear gain saturation, 165–167
    numerical efficiency of, 223–225
    numerical simulations of, 174–179
    in solitary laser, 155–158
    spatially inhomogeneous carrier density, 163–165
    thermal detuning, 161–163
Treat-power-as-a-parameter (TPP) method, 28, 29
Truncated matrix-vector, 829
Tunnel diodes, 430–431
TWM, *see* Traveling-wave model
Two-color heterojunction photodiodes, 331–334
2D Nonlinear optical waveguides
    electric-field envelope, 684
    FDTD numerical implementation, 683–684
    parallel-mirror slab waveguide, 685–686
    planar semiconductor microcavity, 683–684
    SIT soliton, 685, 687
    waveguide width, 685
2D Semiconductor microcavities, 687–691
Two-dimensional scalar approach, 270–273

Twofold influence of effective interface roughness, 244–245
Two-level quantum system, 675–676
Two-photon absorption, 32–33, 44
Two-photon interferometry, 662

U

Ultrafast resonant optical excitation, 717
Ultraviolet photodiodes, 327–331
UMFPACK, direct solvers, 760
Uniaxial perfectly matched layers (UPML), 814
Urbach tails, 514

V

Vacuum quantum fluctuations, 662
Valence band (VB) mixing model, 66
van Roosbroeck system, *see also* Discretizing van Roosbroeck system
  coupling to other model, 766
  equations, 746
  existence and uniqueness results, 752–753
  free energy and dissipation rate, 751
  initial and boundary conditions, 747
  maximum principle, 753
  non-Boltzmann statistics and generalized Einstein relation, 749–751
  recombination processes, 747–748
  thermodynamic equilibrium, 748
Varshni formula, 7
VB mixing model, *see* Valence band mixing model
VCSELs, *see* Vertical-cavity surface-emitting lasers
VECSEL, *see* Vertical external cavity surface-emitting laser
Vector maxwell equations, 671–673
VED, *see* Vertical electric dipole
Vertical-cavity surface-emitting lasers (VCSELs), 12, 563, 564, 627
  fundamental equations, 264–265
  high-contrast gratings, *see* High-contrast gratings
  for light confinement, 570–574
  light confinement in, 263–264
  one-dimensional scalar approach, 266–270
  overwiew, 261–262
  oxide aperture diameter, impact of, 277–278
  physical phenomena, 262–263
  plane-wave admittance method, 565–569
  plane-wave reflection transfer, 569–570
  structure, 265–266
  three-dimensional scalar approach, 273–277
  two-dimensional scalar approach, 270–273
Vertical electric dipole (VED), 600
Vertical external cavity surface-emitting laser (VECSEL), 202
Virtual transitions, nonlinearity due to, 44–45
Visible photodiodes, 327
Voltage-current characteristic, 18, 22, 66
Voronoi diagram, 761

W

WA-FDBPM, *see* Wide-angle finite-difference beam propagation method
Waveguide photodetector (WPD), 302
Waveguide QED approach, *see* Waveguide quantum electrodynamics approach
Waveguide quantum electrodynamics (waveguide QED) approach, 596–599
Waveguide width, 685
Wavelength selectivity, 287, 302
Wavevector, 563–564, 720, 781–782, 789–791
Wetting layer (WL) distribution, 645
Wide-angle finite-difference beam propagation method (WA-FDBPM), 62
Wigner–Weisskopf formula, 705
WL distribution, *see* Wetting layer distribution
WPD, *see* Waveguide photodetector
Wronski–Staebler effect, 499

Y

Yee grid for discretization, 674, 675

Z

Zeeman splitting, 721, 722
Zero group velocity (ZGV) points, 617–618, 621, 623
Zero-reflection, 117